K.R. Meyer
G.R. Hall

Applied
Mathematical
Sciences
90

Introduction to Hamiltonian Dynamical Systems and the N-Body Problem

Springer Science+Business Media, LLC

Applied Mathematical Sciences

Volume 90

Kenneth R. Meyer
Glen R. Hall

Introduction to Hamiltonian Dynamical Systems and the N-Body Problem

With 67 Illustrations

Springer Science+Business Media, LLC

Kenneth R. Meyer
Department of Mathematical Sciences
University of Cincinnati
Cincinnati, OH 45221-0025
USA

Glen R. Hall
Mathematics Department
Boston University
Boston, MA 02215
USA

Mathematics Subject Classifications: 70-02, 70H99, 34C99

Library of Congress Cataloging-in-Publication Data
Meyer, Kenneth R. (Kenneth Ray), 1937–
Introduction to Hamiltonian dynamical systems and the *n*-body problem / Kenneth R. Meyer, Glen R. Hall.
p. cm.—(Applied mathematical sciences; v. 90)
Includes bibliographical references and index.
ISBN 978-1-4757-4075-2 ISBN 978-1-4757-4073-8 (eBook)
DOI 10.1007/978-1-4757-4073-8
1. Hamiltonian systems. 2. Many-body problem. I. Hall, Glen R.
II. Title. III. Series: Applied mathematical sciences
v. 90.
QA614.83.M49 1991
514'.74—dc20 91-21064

Printed on acid-free paper.

Production managed by Hal Henglein; manufacturing supervised by Jacqui Ashri.
Typeset by Asco Trade Typesetting Ltd., Hong Kong.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4757-4075-2

To the memory of Charles C. Conley

Preface

The theory of Hamiltonian systems is a vast subject which can be studied from many different viewpoints. This book develops the basic theory of Hamiltonian differential equations from a dynamical systems point of view. That is, the solutions of the differential equations are thought of as curves in a phase space and it is the geometry of these curves that is the important object of study. The analytic underpinnings of the subject are developed in detail. The last chapter on twist maps has a more geometric flavor. It was written by Glen R. Hall. The main example developed in the text is the classical N-body problem, i.e., the Hamiltonian system of differential equations which describe the motion of N point masses moving under the influence of their mutual gravitational attraction. Many of the general concepts are applied to this example. But this is not a book about the N-body problem for its own sake. The N-body problem is a subject in its own right which would require a sizable volume of its own. Very few of the special results which only apply to the N-body problem are given.

This book is intended for a first course at the graduate level. It assumes a basic knowledge of linear algebra, advanced calculus, and differential equations, but does not assume the advanced topics such as Lebesgue integration, Banach spaces, or Lie algebras. Some theorems which require long technical proofs are stated without proof, but only on rare occasions. The first draft of the book was written in conjunction with a course which was attended by engineering graduate students. The interests and background of these students influenced what was included and excluded.

This book was read by many individuals who made valuable suggestions and many corrections. The first draft was read and corrected by Ricardo Moena, Alan Segerman, Charles Walker, Zhangyong Wan, and Qui Dong Wang while they were students in a seminar on Hamiltonian systems. Gregg

Buck, Konstantin Mischaikow, and Dieter Schmidt made several suggestions for improvements to early versions of the manuscript. Dieter Schmidt wrote the section on the linearization of the equation of the restricted problem at the five libration points. Robin Vandivier found copious grammatical errors by carefully reading the whole manuscript. Robin deserves a special thanks. We hope that these readers absolve us of any responsibility.

The authors were supported by grants from the National Science Foundation, Defense Advanced Research Projects Agency administered by the National Institute of Standards and Technology, the Taft Foundation, and the Sloan Foundation. Both authors were visitors at the Institute for Mathematics and its Applications and the Institute for Dynamics.

Contents

Preface vii

Chapter I. Hamiltonian Differential Equations and the *N*-Body Problem 1
A. Background and Basic Definitions 1
Notation, Hamiltonian systems, Poisson bracket
B. Examples of Hamiltonian Systems 5
The harmonic oscillator, forced nonlinear oscillator, the elliptic sine function, general Newtonian systems, a pair of harmonic oscillators, linear flow on the torus, the Kirchhoff problem
C. The *N*-Body Problem 17
The equations, the classical integrals, the Kepler problem, the restricted 3-body problem
D. Simple Solutions 21
Central configurations, the Lagrangian equilateral triangle solutions, the Euler–Moulton collinear solutions, equilibria for the restricted three-body problem
E. Further Reading 28
Problems 29

Chapter II. Linear Hamiltonian Systems 33
A. Preliminaries 33
Hamiltonian & symplectic matrices, reduction with a Lagrangian set
B. Symplectic Linear Spaces 40
Symplectic basis, Lagrangian splitting
C. The Spectra of Hamiltonian and Symplectic Operators 44
Characteristic equations, J-orthogonal, simple canonical forms

D. Nonelementary Divisors 50
General canonical forms
E. Periodic Systems and Floquet–Lyapunov Theory 53
Multipliers, monodromy matrix, Liapunov transformation
F. Parametric Stability 56
Stability, parametric stability & instability
G. The Critical Points in the Restricted Problem 59
Linear equations at Euler & Lagrange points, Routh's critical mass ratio, canonical forms
H. Further Reading 67
Appendix. Logarithm of a Symplectic Matrix 67
Problems 69

Chapter III. Exterior Algebra and Differential Forms 72
A. Exterior Algebra 72
Multilinear, covector, k-form, determinant
B. The Symplectic Form 76
Symplectic basis & form, determinant of a symplectic matrix
C. Tangent Vectors and Cotangent Vectors 77
Contra- & co-variant vectors, Einstein convention
D. Vector Fields and Differential Forms 79
Contra- & co-variant vector fields, differential forms, exterior derivative, Poincaré's lemma
E. Changing Coordinates and Darboux's Theorem 83
Symplectic structure, symplectic coordinates, Darboux's theorem
F. Integration and Stokes' Theorem 85
Problems 85

Chapter IV. Symplectic Transformations and Coordinates 87
A. Symplectic Transformations 87
Definition, remainder function, variational equations
B. Applications 91
Rotating coordinates, Jacobi coordinates. (special case $N = 2$)
C. Differential Forms and Generating Functions 95
Symplectic form, action-angle variables, d'Alembert character, generating functions, point transformations, Kepler's problem in polar coordinates, the 3-body problem in Jacobi and Jacobi-Polar coordinates
D. Symplectic Transformations with Multipliers and Scaling. 102
Gravitational constant, Equations at an equilibrium point, the restricted three body problem
E. Delaunay and Poincaré Elements 104
F. Further Reading 107
Problems 107

Chapter V. Introduction to the Geometric Theory of Hamiltonian Dynamical Systems **109**
A. Introduction to Dynamical Systems 109
Orbit, trajectory, equilibrium point, period orbit, Hamiltonian dynamical system, reparameterization
B. Discrete Dynamical Systems 113
Diffeomorphisms and symplectomorphisms, Henon map, time τ-map, period map, isotopy, billiards table, crystal model
C. The Flow Box Theorem and Local Integrals 120
Ordinary point, local integrals, Hamiltonian flow box, integrals in involution, ignorable coordinates
D. Noether's Theorem and Reduction 125
Discrete and continuous symmetry, reversible system, Noether's Theorem
E. Periodic Solutions, Fixed Points, and Cross Sections 129
Elementary equilibrium points & periodic orbits, multipliers, cross section, Poincare map, systems with integrals, the cylinder theorem
F. The Stable Manifold Theorem 136
Hyperbolic point, local and global stable and unstable manifold, transversal homoclinic point, the λ-lemma, the Poincare tangle
G. Hyperbolic Systems 141
H. Further Reading 148
Appendix. Proof of Shadowing Lemma 149
Problems 152

Chapter VI. Continuation of Periodic Solutions **154**
A. Continuation of Equilibrium Points and Periodic Solutions 155
B. Lyapunov's Center Theorem 156
Applications to the Euler and Lagrange libration points
C. Poincaré's Orbits 158
D. Hill's Orbits 159
E. Comets 161
F. Continuation from the Restricted to the Full Problem 162
G. Some Elliptic Orbits 164
H. Further Reading 166
Problems 166

Chapter VII. Perturbation Theory and Normal Forms **168**
A. The Method of Lie Transforms 168
Near identity change of coordinates, the forward Lie transform algorithm, the remainder function
B. The Perturbation Algorithm 175
Example: Duffing's equation, the general algorithm, the general perturbation theorem, Duffing's equation revisited, uniqueness of normal forms

C. Normal Form at an Equilibrium 182
The classic case, the general equilibria, example of normal forms in the non-simple case
D. Normal Form at $\mathscr{L}_4$ 189
E. Normal Forms for Periodic Systems and Diffeomorphisms 190
The reduction, general periodic case, general hyperbolic and elliptic points, higher resonance in the planar case, normal forms when multipliers are ± 1
F. Further Reading 199
Problems 199

Chapter VIII. Bifurcations of Periodic Orbits **201**
A. Bifurcations of Periodic Solutions and Points. 201
Elementary fixed points, extremal fixed points, period doubling, k-bifurcation points.
B. Duffing Revisited. 212
Duffing at 1-1 *resonance, k-bifurcation in Duffing's equation*
C. Schmidt's Bridges 216
D. Bifurcation of $\mathscr{L}_4$ 218
E. Further Reading 224
Problems 224

Chapter IX. Stability and KAM Theory **227**
A. Elementary Stability Results 227
B. The Invariant Curve Theorem 229
C. A Simple Example—Duffing's Equation Again 232
D. Applications to the Restricted Problem 233
E. Arnold's Theorem 235
F. Stability of $\mathscr{L}_4$ 238
G. Further Reading 239
Problems 239

Chapter X. Twist Maps and Invariant Curves **241**
A. Introduction 241
B. Notation and Definitions 242
C. Existence of Periodic Orbits 253
D. Monotone Orbits 255
E. Invariant Circles 266
F. Applications 274
G. Further Reading 274
Problems 277

References 279
Index 289

CHAPTER I

Hamiltonian Differential Equations and the N-Body Problem

A. Background and Basic Definitions

This chapter introduces the concept of a Hamiltonian system of ordinary differential equations, sets forth basic notation, reviews some basic facts about the solutions of differential equations, and gives several examples in detail. The primary example is the gravitational N-body problem, which is given a sizable introduction.

1. Notation

Chapters are given roman numerals, sections capital letters, and subsections arabic numerals. Thus, I.A.1 refers to this subsection. Within a chapter the beginning chapter numerals will be omitted; so within Chapter I this section will be denoted A.1. Theorems, lemmas, corollaries, etc., will be renumbered in each section; so a typical theorem reference might be Theorem II.C.2, the second theorem in Section C of Chapter II. Formulas and figures are denoted like theorems but numbered independent of theorems. Hence, Formula (II.C.2) is the second formula in Section C of Chapter II. Within chapters (or sections) the leading numerals (and letter) may be dropped when referring to formulas or theorems within that chapter (or section).

$\mathbb{R}$ will denote the field of real numbers, $\mathbb{C}$ the complex field, and $\mathbb{F}$ either $\mathbb{R}$ or $\mathbb{C}$. $\mathbb{R}^n$ ($\mathbb{C}^n$ or $\mathbb{F}^n$) will denote the space of all n-dimensional column vectors, and, unless otherwise said, all vectors will be column vectors. However, vectors will be written as row vectors within the body of the text for typographical reasons. $\mathscr{L}(\mathbb{F}^n, \mathbb{F}^m)$ will denote the set of all linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ which will sometimes be identified with the set of all $m \times n$ matrices. A

matrix A is block diagonal if it is of the form

$$A = \begin{pmatrix} A_{11} & O_{12} & O_{13} & \cdots & O_{1k} \\ O_{21} & A_{22} & O_{23} & \cdots & O_{2k} \\ O_{31} & O_{32} & A_{33} & \cdots & O_{3k} \\ & \vdots & & & \vdots \\ O_{k1} & O_{k2} & O_{k3} & \cdots & A_{kk} \end{pmatrix},$$

where the A_{ii} are square matrices, and the O_{ij} are the rectangular zero matrices. We will write $A = \operatorname{diag}(A_{11}, A_{22}, \ldots, A_{kk})$.

Functions will be real and smooth unless otherwise stated; smooth means C^∞ or real analytic. If $f(x)$ is a smooth function from an open set $\mathcal{O}$ in $\mathbb{R}^n$ into $\mathbb{R}^m$, then $\partial f/\partial x$ will denote the $m \times n$ Jacobian matrix

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}. \tag{1}$$

If A is a matrix, then A^T will denote its transpose, A^{-1} its inverse, and A^{-T} the inverse transpose if these matrices exist. If $f\colon \mathbb{R}^n \to \mathbb{R}^1$, then $\partial f/\partial x$ is a row vector; let ∇f or $\nabla_x f$ or f_x denote the column vector $\partial f^T/\partial x$. When the derivative of f is thought of as a map from $\mathcal{O}$ into $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, the space of linear operators from $\mathbb{R}^n$ ot $\mathbb{R}^m$, the derivative will be denoted by Df. The variable t will denote a real scalar variable called time and $\dot{} = d/dt$.

2. Hamiltonian Systems

Newton's second law gives raise to systems of second-order differential equations in $\mathbb{R}^n$ and so to a system of first-order equations in $\mathbb{R}^{2n}$, an even-dimensional space. If the forces are derived from a potential function, the equations of motion of the mechanical system have many special properties, most of which follow from the fact that the equations of motion can be written as a Hamiltonian system. The Hamiltonian formalism is the natural mathematical structure in which to develop the theory of conservative mechanical systems.

A Hamiltonian system is a system of $2n$ ordinary differential equations of the form

$$\dot{q} = H_p, \qquad \dot{p} = -H_q$$

$$\left(\dot{q}_i = \frac{\partial H}{\partial p_i}(t, q, p), \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}(t, q, p), \quad i = 1, \ldots, n\right), \tag{2}$$

where $H = H(t, q, p)$, called the *Hamiltonian*, is a smooth real-valued function defined for $(t, q, p) \in \mathcal{O}$, where $\mathcal{O}$ is some open set in $\mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^n$. The vectors $q = (q_1, \ldots, q_n)$ and $p = (p_1, \ldots, p_n)$ are traditionally called the position and momentum vectors, respectively, and t is called time, since that is what these variables represent in the classical examples. The variables q and p are said to be *conjugate* variables; p is conjugate to q, etc. The concept of conjugate variable will grow in importance as the theory develops. The integer n is the *number of degrees of freedom* of the system.

For the general discussion, introduce the $2n$ vector z and the $2n \times 2n$ skew symmetric matrix J and the gradient by

$$z = \begin{pmatrix} q \\ p \end{pmatrix}, \qquad J = J_n = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \qquad \nabla_z H = \nabla H = \begin{bmatrix} \dfrac{\partial H}{\partial z_1} \\ \vdots \\ \dfrac{\partial H}{\partial z_{2n}} \end{bmatrix}, \tag{3}$$

where 0 is the $n \times n$ zero matrix, and I is the $n \times n$ identity matrix. The 2×2 case is special, and so, sometimes J_2 will be denoted by K. In this notation (2) becomes

$$\dot{z} = J\nabla H(t, z). \tag{4}$$

One of the basic results from the general theory of ordinary differential equations is the existence and uniqueness theorem. This theorem states that for each $(t_0, z_0) \in \mathcal{O}$, there is a unique solution $z = \phi(t, t_0, z_0)$ of (4) defined for t near t_0 which satisfies the initial condition $\phi(t_0, t_0, z_0) = z_0$. ϕ is defined on an open neighborhood of the set $\{(t, t_0, z) \in \mathcal{O}: t = t_0\}$ into $\mathbb{R}^n$. Moreover, this solution is maximal in the sense that there are $t_- = t_-(t_0, z_0)$ and $t_+ = t_+(t_0, z_0)$, possibly $\pm\infty$, such that $\phi(t, t_0, z_0)$ is defined for $t_- < t < t_+$ and

$$\lim_{t \to t_\pm} \phi(t, t_0, z_0) = \partial\mathcal{O}, \tag{5}$$

where $\partial\mathcal{O}$ denotes the boundary of $\mathcal{O}$. That is, for any compact set $\mathcal{K} \subset \mathcal{O}$, there is an $\varepsilon > 0$ such that $\phi(t, t_0, z_0) \in \mathcal{O} \backslash \mathcal{K}$ for $t_- < t < t_- + \varepsilon$ and $t_+ - \varepsilon < t < t_+$ when t_- and t_+ are finite—a similar statement holds when one or both of t_- and t_+ are infinite. For example, if $\mathcal{O} = \mathbb{R}^1 \times \mathbb{R}^{2N}$, then either $t_+ = +\infty$ and $\phi(t, t_0, z_0)$ is defined for all $t > t_0$, or t_+ is finite and $\|\phi(t, t_0, z_0)\| \to +\infty$ as $t \to t_+$. The function $\phi(t, t_0, z_0)$ is smooth in all its displayed arguments, and so is C^∞ if the equations are C^∞, and analytic if the equations are analytic. $\phi(t, t_0, z_0)$ is called *the general solution*. See Hartman (1964) or Hale (1972) for details of the theory of ordinary differential equations.

In the special case when H is independent of t, so $H: \mathcal{O} \to \mathbb{R}^1$ where $\mathcal{O}$ is some open set in $\mathbb{R}^{2n}$, the differential equations (4) are *autonomous* and the Hamiltonian system is called *conservative*. In this case the identity

$\phi(t - t_0, 0, z_0) = \phi(t, t_0, z_0)$ holds, since both sides satisfy equation (4) and the same initial conditions. In this case usually the t_0 dependence is dropped and only $\phi(t, z_0)$ is considered, where $\phi(t, z_0)$ is the solution of (4) satisfying $\phi(0, z_0) = z_0$. In this case the solutions are pictured as parameterized curves in $\mathcal{O} \subset \mathbb{R}^{2n}$ and the set $\mathcal{O}$ is called the *phase space*. By the existence and uniqueness theorem, there is a unique curve through each point in $\mathcal{O}$, and by the uniqueness theorem, two such solution curves cannot cross in $\mathcal{O}$.

An *integral* for (4) is a smooth function $F: \mathcal{O} \to \mathbb{R}^1$ which is constant along the solutions of (4), i.e., $F(\phi(t, z_0)) = F(z_0)$ is constant. The classical conserved quantities of energy, momentum, etc., are integrals. The level surfaces $F^{-1}(c) \subset \mathbb{R}^{2n}$, c a constant, are *invariant sets*, i.e., they are sets such that if a solution starts in the set, it remains in the set. In general, the level sets are manifolds of dimension $2n - 1$, and so with an integral F, the solutions lie on the set $F^{-1}(c)$, a space of one less dimension. If you were so lucky as to find $2n - 1$ independent integrals, $F_1, \ldots, F_{2n-1}$, then holding all these integrals fixed would define a curve in $\mathbb{R}^{2n}$, the solution curve. In the classical sense, the problem has been integrated.

3. The Poisson Bracket

Many of the special properties of Hamiltonian systems are formulated in terms of the Poisson bracket operator, so this operator plays a central role in the theory to be developed. Let H, F, and G be smooth functions from $\mathcal{O} \subset \mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}^1$, and define the *Poisson bracket of* F and G by

$$\{F, G\} = \nabla F^T J \nabla G = \frac{\partial F^T}{\partial q}\frac{\partial G}{\partial p} - \frac{\partial F^T}{\partial p}\frac{\partial G}{\partial q}$$

$$\left(\{F, G\}(t, q, p) = \sum_{i=1}^{n}\left[\frac{\partial F}{\partial q_i}(t, p, q)\frac{\partial G}{\partial p_i}(t, q, p) - \frac{\partial F}{\partial p_i}(t, q, p)\frac{\partial G}{\partial q_i}(t, q, p)\right]\right). \tag{6}$$

Clearly $\{F, G\}$ is a smooth map from $\mathcal{O}$ to $\mathbb{R}^1$ as well and one can easily verify that $\{\cdot, \cdot\}$ is skew symmetric and bilinear. A little calculation verifies Jacobi's identity:

$$\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} = 0. \tag{7}$$

By a common abuse of notation, let $F(t) = F(t, \phi(t, t_0, z_0))$ where ϕ is the solution of (4) as above. By the chain rule,

$$\frac{d}{dt}F(t) = \frac{\partial F}{\partial t}(t, \phi(t, t_0, z_0)) + \{F, H\}(t, \phi(t, t_0, z_0)). \tag{8}$$

Hence

$$\frac{dH}{dt} = \frac{\partial H}{\partial t}.$$

Theorem 1. *Let F, G, and H be as above and independent of time, t. Then*

(i) *F is an integral for* (4) *if and only if* $\{F, H\} = 0$.
(ii) *H is an integral for* (4).
(iii) *If F and G are integrals for* (4), *then so is* $\{F, G\}$.
(iv) $\{F, H\}$ *is the time rate of change of F along the solutions of* (4).

PROOF. (i) follows directly from the definition of an integral, and (8). (ii) follows from (i) and the fact that the Poisson bracket is skew symmetric so $\{H, H\} = 0$. (iii) follows from the Jacobi identity (7). (iv) follows from (8). ∎

In many of the examples given below, the Hamiltonian H is the total energy of a physical system, so in this case the theorem says that energy is a conserved quantity.

In the conservative case when H is independent of t, a *critical* point of H as a function, i.e., a point where the gradient of H is zero, is an *equilibrium* (*critical*, *rest*, *stationary*) *point* of the system of differential equation (2) or (4), i.e., a constant solution. For the rest of this subsection, let H be independent of t. A equilibrium point ζ of system (4) is *stable* if for every $\varepsilon > 0$, there is a $\delta > 0$ such that $\|\zeta - \phi(t, z_0)\| < \varepsilon$ for all t whenever $\|\zeta - z_0\| < \delta$.

Theorem 2. *If ζ is a local minimum or maximum of H, then ζ is stable.*

PROOF. This is a classical theorem of Lyapunov. Without loss of generality assume that $\zeta = 0$ and $H(0) = 0$ and that 0 is a local minimum for H. Fix $\varepsilon > 0$. Since $H(0) = 0$ and 0 is a minimum for H, there is an $\eta > 0$ such that $H(z)$ is positive for $0 < \|z\| \leq \eta$. Let $\varkappa = \min(\varepsilon, \eta)$ and $M = \min\{H(z): \|z\| = \varkappa\}$. Since $H(0) = 0$ and H is continuous, there is a $\delta > 0$ such that $H(z) < M$ for $\|z\| < \delta$. If $\|z_0\| < \delta$, then $H(z_0) = H(\phi(t, z_0)\| < M$ for all t. $\|\phi(t, z_0)\| < \varkappa \leq \varepsilon$ for all t because, if not, there would be a time t' when $\|\phi(t', z_0)\| = \varkappa$, but then $H(\phi(t', z_0)) \geq M$, which is a contradiction. ∎

B. Examples of Hamiltonian Systems

1. The Harmonic Oscillator

The harmonic oscillator is the second-order, linear, autonomous, ordinary differential equation

$$\ddot{x} + \omega^2 x = 0, \tag{1}$$

where ω is a positive constant. It can be written as a system of two first-order equations by introducing the conjugate variable $u = \dot{x}/\omega$ and as a Hamiltonian system by letting $H = (\omega/2)(x^2 + y^2)$ (energy in physical problems).

The equations become

$$\begin{aligned} \dot{x} &= \omega u = \frac{\partial H}{\partial u}, \\ \dot{u} &= -\omega x = -\frac{\partial H}{\partial x}. \end{aligned} \tag{2}$$

The variable u is a scaled velocity, and so, the x, u plane is essentially the position-velocity plane, or the phase space of physics. The basic existence and uniqueness theorem of differential equations asserts that through each point (x_0, u_0) in the plane there is a unique solution which passes through this point at any particular epoch t_0. These solutions are given by the formula

$$\begin{pmatrix} x(t, t_0, x_0, u_0) \\ u(t, t_0, x_0, u_0) \end{pmatrix} = \begin{pmatrix} \cos \omega(t - t_0) & \sin \omega(t - t_0) \\ -\sin \omega(t - t_0) & \cos \omega(t - t_0) \end{pmatrix} \begin{pmatrix} x_0 \\ u_0 \end{pmatrix}. \tag{3}$$

The solution curves are parameterized circles. The reason that one introduces the scaled velocity instead of using the velocity itself, as is usually done, is so that the solution curves become circles instead of ellipses. In dynamical systems the geometry of this family of curves in the plane is of prime importance.

Since the system is time independent, it admits H as an integral by Theorem A.1 (or note $\dot{H} = \omega x\dot{x} + \omega u\dot{u} = 0$). Since a solution lies in the set where $H =$ constant, which is a circle in the x, u plane, the integral alone gives the geometry of the solution curves in the plane. See Figure B.1. The origin is a local minimum for H, and so, the origin is stable.

Introduce polar coordinates, $r^2 = x^2 + u^2$, $\theta = \tan^{-1} u/x$, so that Equations (3) become

$$\dot{r} = 0, \qquad \dot{\theta} = -\omega. \tag{4}$$

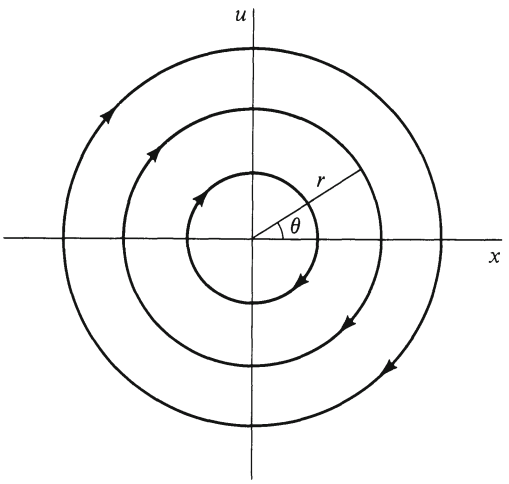

Figure B.1. Phase portrait of the harmonic oscillator.

This shows again that the solutions lie on circles about the origin since $\dot{r} = 0$, and that the circles are swept out with constant angular velocity.

2. The Forced Nonlinear Oscillator

Consider the system

$$\ddot{x} + f(x) = g(t), \tag{5}$$

where x is a scalar, and f and g are smooth real-valued functions of a scalar variable. A mechanical system which gives rise to this equation is illustrated in Figure B.2(a). Here, x is the displacement of a particle of mass 1 that is connected to a nonlinear spring whose restoring force is $-f(x)$ subject to an external force $g(t)$. One assumes that these are the only forces acting and in particular that there are no velocity-dependent forces acting like a frictional force. An electrical system which gives rise to this equation is illustrated in Figure B.2(b). In this case, x represents the charge on a nonlinear capacitor in a series circuit which contains a linear inductor and an external electromotive force $g(t)$. In this problem assume that there is no resistance in the circuit and so there are no terms in $\dot{x}$. This equation is equivalent to the system

$$\dot{x} = y = \frac{\partial H}{\partial y}, \qquad \dot{y} = -f(x) + g(t) = -\frac{\partial H}{\partial x}, \tag{6}$$

where

$$H = \tfrac{1}{2}y^2 + F(x) - xg(t), \qquad F(x) = \int_0^x f(s)ds. \tag{7}$$

Many named equations are of this form, for example: (i) the harmonic oscillator: $\ddot{x} + \omega^2 x = 0$, (ii) the pendulum equation $\ddot{\theta} + \sin\theta = 0$, (iii) the forced Duffing's equation $\ddot{x} + x + \alpha x^3 = \cos\omega t$.

In the case when the forcing term g is absent, $g \equiv 0$, H is an integral, and the solutions lie in the level curves of H. Therefore, the phase portrait is easily obtained by plotting the level curves. In fact, these equations are integrable in the classical sense that they can be solved "up to a quadrature," i.e., they are completely solved after one integration (quadrature). Let $h = H(x_0, y_0)$. Solve

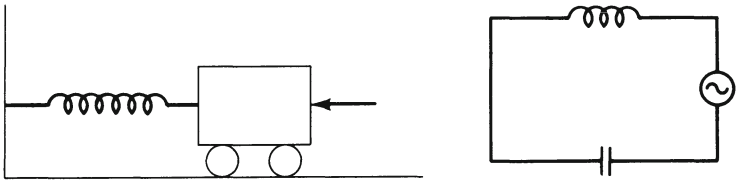

Figure B.2. Physical examples. (a) Spring-mass system; (b) nonlinear LC circuit.

$H = h$ for y and separate the variables to obtain

$$y = \frac{dx}{dt} = \pm\sqrt{2h - 2F(x)},$$
$$t - t_0 = \int_{x_0}^{x} \frac{d\tau}{\pm\sqrt{2h - 2F(\tau)}}. \tag{8}$$

Thus, the solution is obtained by performing the integration (quadrature) in (8) and then taking the inverse of the function so obtained. In general this is quite difficult, but when f is linear, the integral in (8) is elementary, and when f is quadratic or cubic, then the integral in (8) is elliptic.

3. The Elliptic Sine Function

This example is an interesting, nontrivial classical example. In an effort to extend the table of functions which were integrable, the elliptic functions were introduced in the nineteenth century. Usually the properties of these functions are developed in advanced texts on complex analysis, but much of the basic properties follow from the elementary ideas in differential equations. Here one example will be presented.

Let k be a constant $0 < k < 1$ and $\mathrm{sn}(t, k)$ the solution of

$$\ddot{x} + (1 + k^2)x - 2k^2x^3 = 0, \qquad x(0) = 0, \dot{x}(0) = 1. \tag{9}$$

The function $\mathrm{sn}(t, k)$ is called the *Jacobi elliptic sine function.* Put $y = \dot{x}$ so the Hamiltonian or integral is

$$2H = y^2 + (1 + k^2)x^2 - k^2x^4 \tag{10}$$

and on the solution curve $\mathrm{sn}(t, k)$, $2H = 1$ so

$$\dot{\mathrm{sn}}^2 = (1 - \mathrm{sn}^2)(1 - k^2\,\mathrm{sn}^2). \tag{11}$$

The phase portrait of (9) is the level line plot of H. First plot the graph of $\ell(x) = 2h - (1 + k^2)x^2 + k^2x^4 = (2h - 1) + (1 - x^2)(1 - k^2x^2)$ as shown in Figure B.3(a), and then take square roots by plotting $y^2 = \ell(x)$ to obtain the phase portrait of (9) as shown in Figure B.3(b). The solution curve of $\mathrm{sn}(t, k)$ lies in the connected component of $2H = 1$ which contains $x = 0$, $y = \dot{x} = 1$, i.e., the closed curve encircling the origin illustrated by the darker oval in Figure B.3(b). Since the solution $\mathrm{sn}(t, k)$ lies on a closed level line that does not contain an equilibrium point, it must be a periodic function. Both $\mathrm{sn}(t, k)$ and $-\mathrm{sn}(-t, k)$ satisfy (9), and, so, by the uniqueness theorem for ordinary differential equations, $\mathrm{sn}(t, k) = -\mathrm{sn}(-t, k)$, i.e., sn is odd in t.

The curve defined by sn goes through the points $x = \pm 1$, $y = 0$ also. As t increases from zero, $\mathrm{sn}(t, k)$ increases from zero until it reaches its maximum value of 1 after some time, say a time $\varkappa$. (Classically, the constant $\varkappa$ is denoted by K.) Since $\mathrm{sn}(\pm\varkappa, k) = \pm 1$ and $\dot{\mathrm{sn}}(\pm\varkappa, k) = 0$ and both $\mathrm{sn}(t + \varkappa, k)$ and

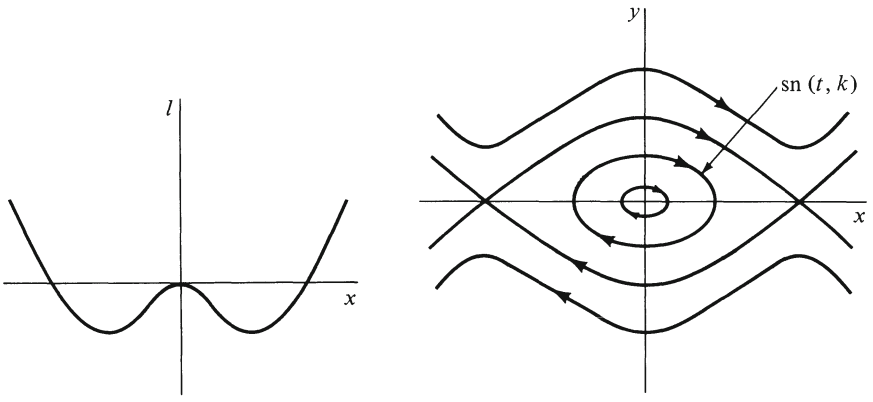

Figure B.3. Portrait of the elliptic sine function. (a) $\ell(x)$ vs x; (b) phase portrait.

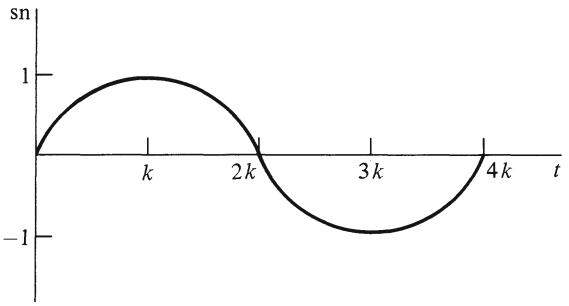

Figure B.4. Graph of sn(t, k) vs t.

$-\text{sn}(t - \varkappa, k)$ satisfy the equation in (9), by uniqueness of the solutions of differential equations it follows that $\text{sn}(t + \varkappa, k) = -\text{sn}(t - \varkappa, k)$, or that sn is $4\varkappa$ periodic and odd harmonic in t. Thus, the Fourier series expansion of sn only contains terms in $\sin(j2\pi t/4\varkappa)$, where j is an odd integer.

It is clear that sn is increasing for $-\varkappa < t < \varkappa$. Equation (9) implies $\ddot{\text{sn}} > 0$ (so sn is convex) for $-\varkappa < t < 0$, and it also implies $\ddot{\text{sn}} < 0$ (so sn is concave) for $0 < t < \varkappa$. Thus, sn has the same basic symmetry properties as the sine function. It is also clear from the equations that $\text{sn}(t, k) \to \sin t$ and $\varkappa \to \pi/2$ as $k \to 0$. The graph of sn(t, k) is indicated in Figure B.4.

The function $\varkappa(k)$ is investigated in the problems. Classical handbooks contain tables of values of the sn function, and simple numerical integration methods can be used to compute values of sn on a small computer. Thus, one knows almost as much about sn(t, k) as about $\sin t$; therefore, sn(t, k) should be added to your list of known or elementary functions. In the problems, you are asked to solve the pendulum equation with your new "elementary func-

tion." There are three other Jacobi elliptic functions which all satisfy equations similar to (9). They were introduced in order to extend the number of functions which can be integrated. In fact, with the four Jacobi elliptic functions, all equations of the form (5) with $g = 0$ and $f(x)$ a quadratic or cubic polynomial can be solved explicitly. See the classic text *Modern Analysis* by Whittaker and Watson (1958) for a complete discussion of the Jacobi elliptic functions. Many of the formulas will remind you of trigonometry.

4. General Newtonian System

The n-dimensional analog of (1) is

$$M\ddot{x} + \nabla F(x) = g(t), \tag{12}$$

where x is an n vector, M is a nonsingular, symmetric $n \times n$ matrix, F is a smooth function defined on an open domain $\mathcal{O}$ in $\mathbb{R}^n$, ∇F is the gradient of F, and g is a smooth n-vector-valued function of t, for t in some open set in $\mathbb{R}^1$. Let $p = M\dot{x}$, then (12) is equivalent to the Hamiltonian system

$$\dot{x} = \frac{\partial H}{\partial p} = M^{-1}p, \qquad \dot{p} = -\frac{\partial H}{\partial x} = -\nabla F(x) + g(t), \tag{13}$$

where the Hamiltonian is

$$H = \tfrac{1}{2}p^T M^{-1}p + F(x) - x^T g(t). \tag{14}$$

If x represents the displacement of a particle, and M is a positive scalar (the mass) times the identity, then p is the linear momentum of the particle, $\frac{1}{2}p^T M^{-1}p$ is the kinetic energy, and F is the potential energy. If $g(t) \equiv 0$, then H is an integral and is "total energy." This terminology is used in reference to nonmechanical systems of the form (12) also. In order to write (13) as a Hamiltonian system, the correct choice of the variable conjugate to x is $p = M\dot{x}$, the linear momentum, and not $\dot{x}$, the velocity.

In the special case when $g \equiv 0$, a critical point of the potential is a critical point of H and hence is an equilibrium point of the Hamiltonian system of equations (13). In many physical examples, M is positive definite. In this case, if x' is a local minimum for the potential F, then $(x', 0)$ is a local minimum for H and therefore is a stable equilibrium point by Theorem A.2.

It is tempting to think that if x' is a critical point of F which is not a minimum of the potential, then the point $(x', 0)$ is an unstable equilibrium point. Indeed, this is stated in Malkin (1952) and LaSalle and Lefschetz (1961) but the proofs given are not convincing. The problems at the end of Chapter II have some results along these lines. See Chapter IX for further discussion of stability questions.

5. A Pair of Harmonic Oscillators

Consider a pair of harmonic oscillators

$$\ddot{x} + \omega^2 x = 0, \qquad \ddot{y} + \mu^2 y = 0, \tag{15}$$

which as a system becomes the Hamiltonian system

$$\begin{aligned} \dot{x} &= \omega u = \frac{\partial H}{\partial u}, & \dot{y} &= \mu v = \frac{\partial H}{\partial v}, \\ \dot{u} &= -\omega x = -\frac{\partial H}{\partial x}, & \dot{v} &= -\mu y = -\frac{\partial H}{\partial y}, \end{aligned} \tag{16}$$

where the Hamiltonian is

$$H = (\tfrac{1}{2}\omega)(x^2 + u^2) + (\tfrac{1}{2}\mu)(y^2 + v^2). \tag{17}$$

In polar coordinates the equations become

$$\begin{aligned} \dot{r} &= 0, & \dot{\rho} &= 0, \\ \dot{\theta} &= -\omega, & \dot{\phi} &= -\mu \end{aligned} \tag{18}$$

and they admit the two integrals

$$I = (\tfrac{1}{2}\omega)(x^2 + u^2), \qquad J = (\tfrac{1}{2}\mu)(y^2 + v^2). \tag{19}$$

In many physical problems these equations are only the first approximation, and the full system does not admit the two individual integrals (energies) but does admit the sum H. Think, for example, of a pea rolling around in a bowl; the linearized system at the minimum would be of the form (16). In this case, $H^{-1}(1)$ is an invariant set for the flow and topologically a 3-sphere.

Consider the general solution through r_0, ρ_0, θ_0, and ϕ_0 at epoch $t = 0$. The solutions with $r_0 = 0$ and $\rho_0 > 0$ or $\rho_0 = 0$ and $r_0 > 0$ lie on circles and correspond to periodic solutions of period $2\pi/\mu$ and $2\pi/\omega$, respectively. These periodic solutions are special and are usually called the normal modes.

The set where $r = r_0 > 0$ and $\rho = \rho_0 > 0$ is an invariant torus for (16) or (18). Angular coordinates on this torus are θ and ϕ, and the equations are

$$\dot{\theta} = -\omega, \qquad \dot{\phi} = -\mu, \tag{20}$$

the standard linear equations on a torus.

If ω/μ is rational, then $\omega = p\tau$ and $\mu = q\tau$, where p and q are relatively prime integers. In this case the solution of (16) through θ_0, ϕ_0 at epoch $t = 0$ is $\theta(t) = \theta_0 - \omega t$, $\phi(t) = \phi_0 - \mu t$, and so if $T = 2\pi/\tau$, then $\theta(T) = \theta_0 + p2\pi$ and $\phi(T) = \phi_0 + q2\pi$. That is, the solution is periodic with period T on the torus [see Figure B.5(a)] and this corresponds to periodic solutions of (16).

If ω/μ is irrational, then none of the solutions are periodic. In fact, the solutions of (20) are dense lines on the torus [see Figure B.5(b) and Subsection

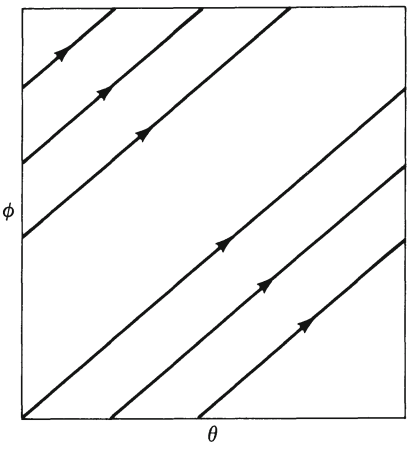

Figure B.5. Linear flow on the torus.

6], and this corresponds to the fact that the solutions of (16) are quasiperiodic but not periodic.

We can use polar coordinates to introduce coordinates on the sphere provided we are careful to observe the conventions of polar coordinates: (i) $r \geq 0$, (ii) θ is defined modulo 2π, and (iii) $r = 0$ corresponds to a point. That is, if we start with the rectilinear strip $r \geq 0$, $0 \leq \theta \leq 2\pi$ [Figure B.6(a)], then identify the $\theta = 0$ and $\theta = 2\pi$ edges to get a half-closed annulus [Figure B.6(b)], and finally if we identify the circle $r = 0$ with a point, then we have a plane [Figure B.6(c)].

Starting with the polar coordinates r, θ, ρ, ϕ for R^4, we note that on the 3-sphere, $E = r^2 + \rho^2 = 1$; so, we can discard ρ and have $0 \leq r \leq 1$. We will use r, θ, ϕ as coordinates on S^3. Now r, θ with $0 \leq r \leq 1$ are just polar coordinates for the closed unit disk. For each point of the open disk, there is a circle with coordinate ϕ (defined mod 2π), but when $r = 1$, $\rho = 0$; so, the circle collapses to a point over the boundary of the disk. The geometric model of S^3 is two solid cones with points on the boundary cones identified as shown in Figure B.7. Through each point in the open unit disk with coordinates r, θ there is a line segment (the dashed line) perpendicular to the disk. The angular coordinate ϕ is measured on this segment; $\phi = 0$ is the disk $\phi = \pi$ is the upper boundary cone, $\phi = -\pi$ is the boundary lower cone. Each point on the upper boundary cone with coordinates r, θ, $\phi = \pi$ is identified with the point on the lower boundary cone with coordinate r, θ, $\phi = -\pi$. From this model follows a series of interesting geometric facts.

For α, $0 < \alpha < 1$, the set where $r = \alpha$ is a 2-torus in the 3-sphere, and for $\alpha = 0$ or 1, the set $r = \alpha$ is a circle. Since r is an integral for the pair of oscillators, these tori and circles are invariant sets for the flow defined by the harmonic oscillators. The two circles $r = 0$, 1 are periodic solutions, called the normal modes. The two circles are linked in S^3, i.e., one of the circles intersects

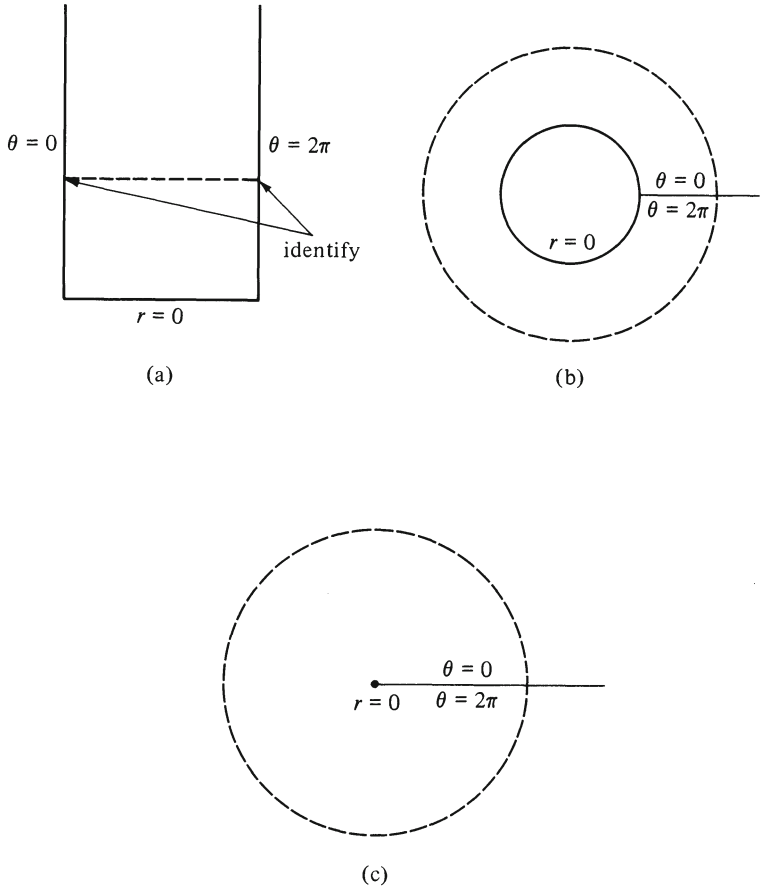

Figure B.6. The polar coordinate conventions.

a disk bounded by the other circle in an algebraically nontrivial way. The circle where $r = 1$ is the boundary of the shaded disk in Figure B.8, and the circle $r = 0$ intersects this disk once. It turns out that the number of intersections is independent of the bounding disk provided one counts the intersections algebraically.

Consider the special case when $\omega = \mu = 1$. In this case every solution is periodic, and so its orbit is a circle in the 3-sphere. Other than the two special circles, on each orbit as θ increases by 2π, so does ϕ. Thus, each such orbit hits the open disk where $\phi = 0$ (the shaded disk in Figure B.8) in one point. We can identify each such orbit with the unique point where it intersects the disk. One special orbit meets the disk at the center, and so we can identify it with the center. The other is the outer boundary circle which is a single orbit. When we identify this circle with a point, the closed disk with its outer circle identified with a point becomes a 2-sphere. Thus:

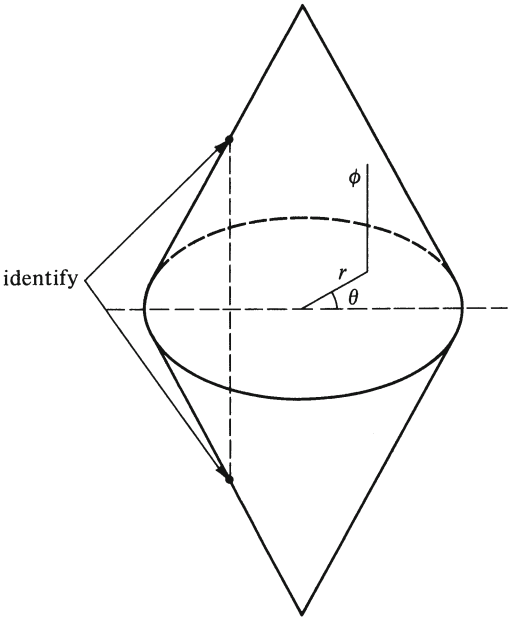

Figure B.7. A model of the 3-sphere, S^3.

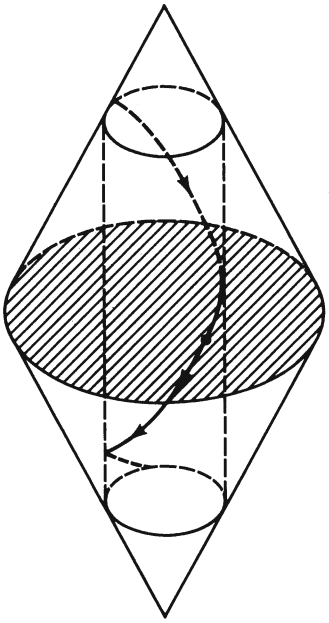

Figure B.8. Orbit on the 3-sphere S^3.

Theorem 1. *The 3-sphere, S^3, is the union of circles. Any two of these circles are linked. The quotient space obtained by identifying a circle with a point is a 2-sphere (The Hopf fibration of S^3).*

Let D be the open disk $\phi = 0$, the shaded disk in Figure B.8. The union of all the orbits which meet D is a product of a circle and a 2-disk, so each point not on the special circle $r = 1$ lies in an open set that is the product of a 2-disk and a circle. By reversing r and ρ in the discussion given above, the circle where $r = 1$ has a similar neighborhood. So locally the 3-sphere is the product of a disk and a circle, but the sphere is not the product of a two manifold and a circle. (The sphere has a trivial fundamental group, but such a product would not.)

When $\omega = p$ and $\mu = q$, p and q relatively prime integers, all solutions are periodic, and the 3-sphere is again a union of circles, but it is not locally a product near the special circles. The nonspecial circles are p-q torus knots. They link p times with one special circle and q times with the other.

The linking statements follow by a slight extension of the ideas of the previous proposition. A p-q torus knot is a closed curve which wraps around the standard torus in R^3 in the longitudinal direction p times and in the meridional direction q times. If p and q are different from 1, the knot is nontrivial. The details are too lengthy for here, but Figure B.9 shows that the 3-2 torus knot is the classical trefoil or clover-leaf knot. Figure B.9(a) is the standard model of a torus—a square with opposite sides identified. The line with

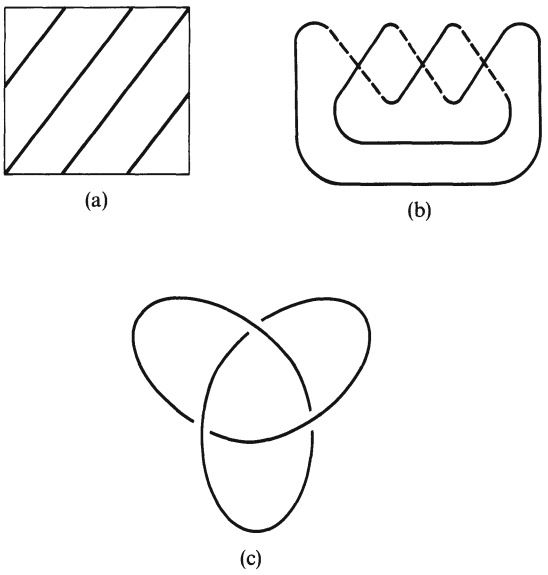

Figure B.9. Toral knot. (a) Orbit on torus; (b) unwinding the orbit; (c) orbit as trefoil.

slope 3/2 is shown warpping three times around one way and twice around the other. Think of folding the top half of the square back and around and then gluing the top edge to the bottom to form a cylinder. Add two extra segments of curves to connect the right and left ends of the curve to get Figure B.9(b). Smoothly deform Figure B.9(b) to Figure B.9(c), the standard presentation of the trefoil. See Rolfsen (1976) for more information on knots.

6. Linear Flow on the Torus

In order to show that the solutions of (20) on the torus are dense when ω/μ is irrational, the following simple lemmas from number theory are needed.

Lemma 2. *Let δ be any irrational number. Then for every $\varepsilon > 0$, there exist integers q and p such that*

$$|q\delta - p| < \varepsilon. \tag{21}$$

PROOF. Case 1: $0 < \delta < 1$. Let $N \geq 2$ be an integer and $S_N = \{s\delta - r : 1 \leq s, r \leq N\}$. For each element of this set we have $|s\delta - r| < N$. Since δ is irrational, there are N^2 distinct members in the set S; so at least one pair is at least $4/N$ apart. [If not, the total length would be greater than $(N^2 - 1)4/N > 2N$.] Call this pair $s\delta - r$ and $s'\delta - r'$. So

$$0 < |(s - s')\delta - (r - r')| < \frac{2}{N} < \frac{2}{|s - s'|}. \tag{22}$$

Take $N > 2/\varepsilon$, $q = s - s'$, and $p = r' - r$ to finish this case. The other cases follow from the above. If $-1 < \delta < 0$, then apply the above to $-\delta$; and if $|\delta| > 1$, apply the above to $1/\delta$. ∎

Lemma 3. *Let δ be any irrational number and ξ any real number. Then for every $\varepsilon > 0$ there exist integers p and q such that*

$$|q\delta - p - \xi| < \varepsilon. \tag{23}$$

PROOF. Let p' and q' be as given in Lemma 1, so $\eta = q'\delta - p'$ satisfies $0 < |\eta| < \varepsilon$. There is an integer m such that $|m\eta - \xi| < \varepsilon$. The lemma follows by taking $q = mq'$ and $p = mp'$. ∎

Theorem 4. *Let ω/μ be irrational. Then the solution curves defined by Equations* (20) *are dense on the torus.*

PROOF. Measure the angles in revolutions instead of radians so that the angles θ and ϕ are defined modulo 1 instead of 2π. The solution of Equations (20) through $\theta = \phi = 0$ at $t = 0$ is $\theta(t) = \omega t$, $\phi(t) = \mu t$. Let $\varepsilon > 0$ and ξ be given. Then $\theta \equiv \xi$ and $\phi \equiv 0 \bmod 1$ is an arbitrary point on the circle $\phi \equiv 0 \bmod 1$

on the torus. Let $\delta = \omega/\mu$ and p, q be as given in Lemma 2. Let $\tau = q/\mu$, so $\theta(\tau) = \delta q$, $\phi(\tau) = q$. Thus, $|\theta(\tau) - p - \xi| < \varepsilon$, but since p is an integer, this means that $\theta(\tau)$ is within ε of ξ; and so, the solution through the origin is dense on the circle $\phi \equiv 0 \bmod 1$. The remainder of the proof follows by translation.

■

7. The Kirchhoff Problem

Mechanical problems are not the only way Hamiltonian systems arise. Kirchhoff (1897) derived the equations of motion of N vortices of an incompressible fluid moving in the plane under their mutual interaction. Let η_i be the position vector of the ith vortex whose circulation is $\varkappa_i$; then the equations of motion are

$$\varkappa_j \dot{\eta}_j = K \frac{\partial U}{\partial \eta_j}, \qquad j = 1, \ldots, N, K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{24}$$

where

$$U = \sum_{1 \le i < j \le N} \varkappa_i \varkappa_j \log \|\eta_i - \eta_j\|. \tag{25}$$

If we set $\eta_i = (q_i, p_i)$, $p = (p_1, \ldots, p_N)$, and $q = (q_1, \ldots, q_N)$, the equations become

$$\dot{q} = \frac{\partial U}{\partial p}, \quad \dot{p} = -\frac{\partial U}{\partial q}. \tag{26}$$

Sometimes these equations can be treated like the N-body problem to be introduced in the Section C. We shall develope some of the more basic facts about these equations in the problems.

C. The N-Body Problem

1. The Equations

The beginning of science as we know it today started with Newton's formulations of the three laws of motion, the universal law of gravity, and his solution of the two-body problem. With a few simple principles and some mathematics, he could explain the three empirical laws of Kepler on the motion of Mars and the other planets. For the planets, the sun and the planet can be considered as a 2-body problem in the first approximation. Newton turned his attention to the motion of the moon which required three bodies: the sun, Earth, and moon, in the first approximation. Unable to solve the 3-body

problem, he wrote that the theory of the moon gave him headaches. The 3-body problem thus became the most celebrated problem in mathematics.

Consider N point masses moving in a Newtonian reference system, $\mathbb{R}^3$, with the only forces acting on them being their mutual gravitational attraction. Let the ith particle have position vector q_i and mass $m_i > 0$; then applying Newton's second law and law of gravity yields the equations of motion

$$m_i \ddot{q}_i = \sum_{j=1}^{N}{}' \frac{G m_i m_j (q_j - q_i)}{\|q_i - q_j\|^3} = \frac{\partial U}{\partial q_i}, \tag{1}$$

where

$$U = \sum_{1 \le i < j \le N} \frac{G m_i m_j}{\|q_i - q_j\|}. \tag{2}$$

In the above, G is the universal gravitational constant, $G = 6.6732 \times 10^{-11}$ m^3/s^2 kg, U is the self-potential or the negative of the potential, and the prime on the summation sign indicates that you do not divide by zero; and so, the term when $i = j$ is omitted. Let $q = (q_1, q_2, \ldots, q_N) \in \mathbb{R}^{3N}$ and $M = \operatorname{diag}(m_1, m_1, m_1, \ldots, m_N, m_N, m_N)$; so, equations (1) are of the form

$$M\ddot{q} - \frac{\partial U}{\partial q} = 0. \tag{3}$$

As in the above, define $p = (p_1, \ldots, p_N) \in \mathbb{R}^{3N}$ by $p = M\dot{q}$ so $p_i = m_i \dot{q}_i$ is the momentum of the ith particle. The equations of motion become

$$\dot{q}_i = \frac{p_i}{m_i} = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = \sum_{i=1}^{N}{}' \frac{G m_i m_j (q_j - q_i)}{\|q_i - q_j\|^3} = \frac{\partial H}{\partial q_i}, \tag{4}$$

where the Hamiltonian is

$$H = \sum_{i=1}^{N} \frac{\|p_i\|^2}{2m_i} - U. \tag{5}$$

Here again the correct conjugate of position q is momentum p.

2. The Classical Integrals

The N-body problem is a system of $6N$ first-order equations; so, a complete solution would require $6N$-1 time-independent integrals plus one time-dependent integral. It is now fairly clear that for $N > 2$, there are not that many global integrals. However, there are 10 integrals for the system.

Let $L = p_1 + \cdots + p_N$ be total linear momentum. From (4) it follows that $\dot{L} \equiv 0$ since each term in the sum appears twice with opposite sign. This gives $\ddot{C} = 0$, where $C = \sum m_i q_i$ is the center of mass of the system since $\dot{C} = L$. So the total linear momentum is constant, and the center of mass of the system moves with uniform rectilinear motion. Integrating the center of mass equation gives $C = Lt + C_0$, where L and C_0 are constants of integration. L and

C_0 are functions of the initial conditions; so are integrals of the motion. Thus, we have six constants of motion or integrals, namely, the three components of L and the three components of C_0.

Let $A = \sum_1^n q_i \times p_i$ be the total angular momentum of the system. Then

$$\frac{dA}{dt} = \sum_1^n (\dot{q}_i \times p_i + q_i \times \dot{p}_i)$$

$$= \sum_1^n q_i \times m_i q_i + \sum_1^n \sum_1^{n\prime} \frac{G m_i m_j q_i \times (q_j - q_i)}{\|q_i - q_j\|^3} = 0.$$

The first sum above is zero because $q_i \times q_i = 0$. In the second sum use $q_i \times (q_j - q_i) = q_i \times q_j$ and then observe that each term in the remaining sum appears twice with opposite sign. Thus, the three components of angular momentum are constants of the motion or integrals also. Remembering that energy, H, is also an integral we have the classical 10 integrals of the N-body problem.

3. The Kepler Problem

A special case of the 2-body problem is when one body is assumed to be fixed at the origin—say, for example, it is so massive, like the sun, that to the first approximation it does not move. In this case the equations describe the motion of the other body and have the form

$$\ddot{q} = -\frac{\mu q}{\|q\|^3}, \tag{6}$$

where $q \in \mathbb{R}^3$ is the position vector of the other body and μ is the constant Gm where G is the universal gravitational constant and m is the mass of the body fixed at the origin. In this case, by defining $p = \dot{q}$, this equation becomes Hamiltonian with Hamiltonian

$$H = \frac{\|p\|^2}{2} - \frac{\mu}{\|q\|}. \tag{7}$$

As before $A = q \times p$, angular momentum is constant along the solutions; and so, the three components of A are integrals. If $A = 0$, then

$$\frac{d}{dt}\left(\frac{q}{\|q\|}\right) = \frac{(q \times \dot{q}) \times q}{\|q\|^3} = \frac{A \times q}{\|q\|^3} = 0. \tag{8}$$

The first equality above is a vector identity. So, if angular momentum is zero, the motion is collinear. Letting the line of motion be one of the coordinate axes makes the problem a one degree of freedom problem and so solvable by formulas (B.4). In this case the integrals are elementary and one obtains simple formulas (see the Problem Section).

If $A \neq 0$, then both q and $p = \dot{q}$ are orthogonal to A; and so, the motion takes place in the plane orthogonal to A. In this case take one coordinate axis, say the last, to point along A; so, the motion is in a coordinate plane. The equations of motion in this coordinate plane have the same form as (6), but now $q \in \mathbb{R}^2$. In the planar problem only the component of angular momentum perpendicular to the plane is nonzero; so the problem is reduced to a two degree of freedom with one integral. Such a problem is solvable "up to quadrature." It turns out that the problem is solvable (well, almost) in terms of elementary functions. There are many ways to solve the Kepler problem. One way is outlined in the Problem Section at the end of this chapter, and another will be given in Chapter 4 once transformation theory for Hamiltonian systems is discussed.

4. The Restriced 3-Body Problem

A special case of the 3-body problem comes about as a limiting case when one lets one of the masses tend to zero. A careful derivation of this problem will be given in Chapter 3 after transformation theory is developed. In the traditional derivation of the restricted 3-body problem, one is asked to consider the motion of an infinitesimally small particle moving in the plane under the influence of the gravitational attraction of two finite particles which revolve around each other in a circular orbit with uniform velocity. Although this description is picturesque, it is hard to see the relationship this problem has to the full 3-body problem. For now we shall simply give the Hamiltonian. Let the two finite particles, called the primaries, have mass $\mu > 0$ and $1 - \mu > 0$. Let $x \in \mathbb{R}^2$ be the coordinate of the infinitesimal particle in a uniformly rotating coordinate system and $y \in \mathbb{R}^2$ the momentum conjugate to x. The rotating coordinate system is so chosen that the particle of mass μ is always at $(1 - \mu, 0)$ and the particle of mass $1 - \mu$ is at $(-\mu, 0)$. See Figure C.1.

The Hamiltonian governing the motion of the third (infinitesimal) particle in these coordinates is

$$H = \frac{\|y\|^2}{2} - x^T K y - U, \tag{9}$$

where $x, y \in \mathbb{R}^2$ are conjugate, K is J_2, and U is the self-potential

$$U = \frac{\mu}{d_1} + \frac{1 - \mu}{d_2}, \tag{10}$$

and d_i is the distance from the infinitesimal body to the ith primary or

$$d_1^2 = (x_1 - 1 + \mu)^2 + x_2^2, \qquad d_2^2 = (x_1 + \mu)^2 + x_2^2. \tag{11}$$

The equations of motion are

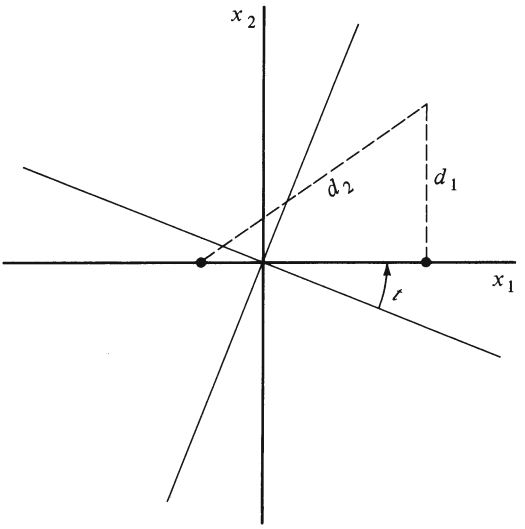

Figure C.1. The rotating coordinates for the restricted problem.

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial y} = y + Kx, \\ \dot{y} &= -\frac{\partial H}{\partial x} = Kx + \frac{\partial H}{\partial x}. \end{aligned} \tag{12}$$

The term $x^T K y$ in the Hamiltonian H is due to the fact that the coordinate system is not a Newtonian system, but a rotating coordinate system. It gives rise to the so called Coriolis forces in the equations of motion (12). The line joining the masses is known as the line of syzygy.

The proper definition of the restricted 3-body problem is the system of differential equations (12) defined by the Hamiltonian in (9). It is a two degree of freedom problem that seems simple but has defied integration. It has given rise to an extensive body of research. We shall return to this problem often in the subsequent chapters.

D. Simple Solutions

The N-body problem for $N > 2$ has resisted all attempts to be solved; indeed it is generally believed that the problem cannot be integrated in the classical sense. Over the years many special types of solutions have been found by using various mathematical techniques. In this section we shall find some solutions by the time honored method of guess and test. The simplest type of

solution one might look for are equilibrium or rest solutions. From (C.1) or (C.3) an equilibrium solution would have to satisfy

$$\frac{\partial U}{\partial q_i} = 0, \qquad \text{for } i = 1, \ldots, N. \tag{1}$$

However, U is homogeneous of degree -1; and so, by Euler's theorem on homogeneous polynomials

$$\sum q_i \frac{\partial U}{\partial q_i} = -U. \tag{2}$$

Since U is the sum of positive terms, it is positive, but (1) would make the right-hand side of (2) zero, which is a contradiction. Thus, there are no equilibrium solutions of the N-body problem.

1. Central Configurations

To seek collinear solutions of (C.1), try $q_i(t) = \phi(t)a_i$, where the a_i's are constant vectors and $\phi(t)$ is a scalar-valued function. Substituting into (C.1) and rearranging, yields

$$|\phi|^3 \phi^{-1} \ddot{\phi} m_i a_i = \sum_{j=1}^{N}{}' \frac{G m_i m_j (a_j - a_i)}{\|a_j - a_i\|^3}. \tag{3}$$

Since the right-hand side is constant, the left hand side must be also; therefore, (3) has a solution if there is a scalar function $\phi(t)$, a constant λ, and constant vectors a_i such that

$$\ddot{\phi} = -\frac{\lambda \phi}{|\phi|^3}, \tag{4}$$

$$-\lambda m_i a_i = \sum_{j=1}^{N}{}' \frac{G m_i m_j (a_j - a_i)}{\|a_j - a_i\|^3}, \quad i = 1, \ldots, N. \tag{5}$$

Equation (4) is a simple ordinary differential equation (the one-dimensional Kepler problem!); and so has many solutions. For example, one solution is $\alpha t^{2/3}$, where $\alpha^3 = 9\lambda/2$. This is a solution which goes from zero to infinity as t goes from zero to infinity. The complete analysis of (4) is left to the problems. Equation (5) is a nontrivial system of nonlinear algebraic equations. The complete solution is known only for $N = 2, 3$, but there are many special solutions known.

Now consider the planar N-body problem, so all the vectors lie in $\mathbb{R}^2$. Identify $\mathbb{R}^2$ with the complex plane $\mathbb{C}$ by considering the q_i, p_i, etc., as complex numbers. Seek a homographic solution of (C.1) by letting $q_i(t) = \phi(t)a_i$, where the a_i's are constant complex numbers and $\phi(t)$ is a time-dependent

complex-valued function. Geometrically, multiplication by a complex number is a rotation followed by a dilation or expansion, i.e., a homography. Thus, we seek a solution such that the configuration of the particles are always homographically equivalent to a fixed configuration. Substituting this guess into (C.1) and rearranging gives the same Equation (3), and the same argument gives Equations (4), which are now the two-dimensional Kepler problem, and Equation (5). That is, if you have a solution of (5) where the a_i's are planar, then there is a solution of the N-body problem of the form $q_i = \phi(t)a_i$, where $\phi(t)$ is any solution of the planar Kepler problem, e.g., circular, elliptic, etc.

A geometric configuration of the N particles given by constant vectors $a_1, \ldots, a_N$ which satisfy (5) for some λ is called a *central configuration* (or c.c. for short). In the special case when the a_i's are coplanar, a central configuration is also called a *relative equilibrium* because, as we shall see, they become equilibrium solutions in a rotating coordinate system. Central configurations are important in the study of the total collapse of the system because it can be shown that the limiting configuration of a system as it tends to a total collapse is a central configuration.

Note that any uniform scaling of a c.c. is also a c.c. In order to measure the size of the system, we define the moment of inertia of the system as

$$I = \frac{1}{2}\sum_{i=1}^{N} m_i \|q_i\|^2. \tag{6}$$

Then (5) can be rewritten as

$$\frac{\partial U}{\partial q}(a) + \lambda \frac{\partial I}{\partial q}(a) = 0, \tag{7}$$

where $q = (q_1, \ldots, q_N)$ and $a = (a_1, \ldots, a_N)$. The constant λ can be considered as a Lagrange multiplier; and so, a central configuration is a critical point of the self-potential U restricted to a constant moment of inertia manifold, $I = I_0 =$ constant. Fixing I_0 fixes the scale.

Let a be a central configuration. U is homogeneous of degree -1, and I is homogeneous of degree 2. By taking the dot product of the vector a with the equation in (7) and applying Euler's theorem on homogeneous polynomials we find that $-U + 2\lambda I = 0$ or

$$\lambda = \frac{U(a)}{2I(a)} > 0. \tag{8}$$

Summing (5) on i gives $\sum m_i a_i = 0$, so the center of mass is at the origin. If A is an orthogonal matrix, either 3×3 in general or 2×2 in the planar case, then clearly $Aa = (Aa_1, \ldots, Aa_N)$ is a c.c. also with the same λ. If $\tau \neq 0$, then $(\tau a_1, \tau a_2, \ldots, \tau a_N)$ is a c.c. also with λ replaced by λ/τ^3. Thus, any configuration similar to a c.c. is a c.c. When counting c.c., one only counts similarity classes.

2. The Lagrangian Equilateral Triangle Solutions

Consider the c.c. formula (5) for the planar 3-body problem. Then we seek six unknowns, two components each for a_1, a_2, a_3. If we hold the center of mass at the origin, we can eliminate two variables; if we fix the moment of inertia, I, we can reduce the dimension by one; and if we identify two configurations which differ by a rotation only, we can reduce the dimension by one again. Thus, in theory you can reduce the problem by four dimensions, so that you have a problem of finding critical points of a function on a two-dimensional manifold. This reduction is difficult in general, but there is a trick that works well for the planar 3-body problem.

Let $\rho_{ij} = \|q_i - q_j\|$ denote the distance between the ith and jth particles. Once the center of mass is fixed at the origin and two configurations are identified which are rotationally equivalent, then the three variables $\rho_{12}, \rho_{23}, \rho_{31}$ are local coordinates near a noncollinear configuration. That is, by specifying the angle between a fixed line and say $q_1 - q_2$, the location of the center of mass, and the three variables $\rho_{12}, \rho_{23}, \rho_{31}$, then the configuration of the masses is uniquely specified. The function U is already written in terms of these variables since

$$U = G\left(\frac{m_1 m_2}{\rho_{12}} + \frac{m_2 m_3}{\rho_{23}} + \frac{m_3 m_1}{\rho_{31}}\right). \tag{9}$$

Let M be the total mass, i.e., $M = \sum m_i$, and assume the center of mass is at the origin; then

$$\begin{aligned}\sum_i \sum_j m_i m_j \rho_{ij}^2 &= \sum_i \sum_j m_i m_j \|q_i - q_j\|^2 \\ &= \sum_i \sum_j m_i m_j \|q_i\|^2 - 2\sum_i \sum_j m_i m_j (q_i, q_j) + \sum_i \sum_j m_i m_j \|q_j\|^2 \\ &= 2MI - 2\sum_i m_i \Big(q_i, \sum m_j q_j\Big) + 2MI \\ &= 4MI.\end{aligned}$$

Thus, if the center of mass is fixed at the origin, then

$$I = \frac{1}{4M} \sum_i \sum_j m_i m_j \rho_{ij}^2. \tag{10}$$

So, I can be written in terms of the mutual distances also. Holding I fixed is the same as holding $J = \frac{1}{2}(m_{12}\rho_{12}^2 + m_{23}\rho_{23}^2 + m_{31}\rho_{31}^2)$ fixed. Thus, the conditions for U to have a critical point on the set $J =$ constant in these coordinates is

$$-G\frac{m_i m_j}{\rho_{ij}^2} + \lambda m_i m_j \rho_{ij} = 0 \qquad (i, j) = (1, 2), (2, 3), (3, 1), \tag{11}$$

which clearly has as its only solution $\rho_{12} = \rho_{23} = \rho_{31} = (G/\lambda)^{-1/3}$. This solution is an equilateral triangle, and λ is a scale parameter. These solutions are attributed to Lagrange.

Theorem 1. *For any values of the masses, there is one and only one noncollinear central configuration for the 3-body problem, namely, the three particles are at the vertices of an equilateral triangle.*

It is trivial to see in these coordinates that the equilateral triangle c.c. is a nondegenerate minimum of the self-potential U. The above argument would also show that for any values of the masses there is one and only one noncoplanar c.c. for the 4-body problem, namely, the regular tetrahedron configuration.

3. The Euler–Moulton Collinear Solutions

Consider the collinear N-body problem, so $q = (q_1, \ldots, q_N) \in \mathbb{R}^N$. Set $S' = \{q: I(q) = 1\}$, an ellipsoid or topological sphere, of dimension $N - 1$ in $\mathbb{R}^N$; set $G = \{C(q) = \sum m_i q_i = 0\}$, a plane of dimension $N - 1$ in $\mathbb{R}^N$; and $S = S' \cap G$, a sphere of dimension $N - 2$ in the plane G. See Figure D.1 where $N = 3$, S' is a 2-sphere, G is a plane, and S is a great circle. Let $\Delta'_{ij} = \{q: q_i = q_j\}$ and $\Delta' = \bigcup \Delta'_{ij}$; so, U is defined and smooth on $\mathbb{R}^N \backslash \Delta'$. Δ' is a union of planes through the origin and, so, intersects S in spheres of dimension $N - 3$, denoted by Δ.

Let $\mathcal{U}$ be the restriction of U to $S \backslash \Delta$, and so, a critical point of $\mathcal{U}$ is a central configuration. Note that, $S \backslash \Delta$ has $N!$ connected components. This is because a component of $S \backslash \Delta$ corresponds to a particular ordering of the q_i's. That is, to each connected component there is an ordering $q_{i_1} < q_{i_2} < \cdots < q_{i_N}$, where $(i_1, i_2, \ldots, i_N)$ is a permutation of $1, 2, \ldots, N$. There are $N!$ such permutations. Since $\mathcal{U} \to \infty$ as $q \to \Delta$, the function $\mathcal{U}$ has at least one minimum per connected component. Thus, there are at least $N!$ critical points.

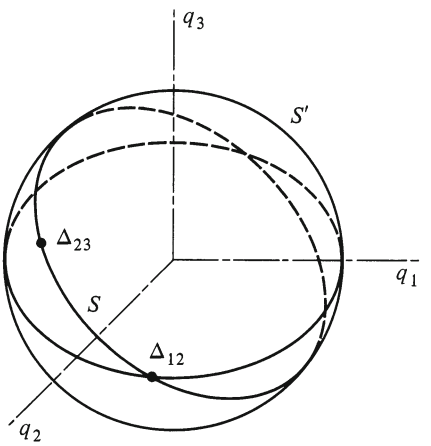

Figure D.1. The spaces S' and S for $N = 3$.

Let a be a critical point of $\mathcal{U}$; so that a satisfies (5) and $\lambda = U(a)/2I(a)$. The derivative of $\mathcal{U}$ at a in the direction $v = (v_1, \ldots, v_N) \in T_a S$ is

$$D\mathcal{U}(a)(v) = -\sum \frac{Gm_i m_j (v_j - v_i)}{|a_j - a_i|} + \lambda \sum m_i a_i v_i, \tag{12}$$

and the second derivative is

$$D^2\mathcal{U}(a)(v, w) = 2\sum \frac{Gm_i m_j}{|a_j - a_i|^3}[(w_j - w_i)(v_j - v_i)] + \lambda \sum m_i w_i v_i. \tag{13}$$

From the above, $D^2\mathcal{U}(a)(v, v) > 0$ when $v \neq 0$; so, the Hessian is positive definite at a critical point and each such critical point is a minimum of $\mathcal{U}$. Thus, there can only be one critical point of $\mathcal{U}$ on each connected component, or there are $N!$ critical points.

In counting the critical points above, we have not removed the symmetry from the problem. The only one-dimensional orthogonal transformation is a reflection in the origin. Thus, we have counted a c.c. and its reflection; we have counted each c.c. twice. Thus, we have

Theorem 2 (Euler–Moulton). *There are exactly $N!/2$ collinear central configurations in the N-body problem, one for each ordering of the masses on the line.*

These c.c. are minimum of $\mathcal{U}$ only on the line. It can be shown that they are saddle points in the planar problem.

4. Equilibria for the Restricted 3-Body Problem

The full 3-body problem has no equilibrium points, but as we have seen there are solutions of the planar problem where the particles move on uniformly rotating solutions. In particular, there are the solutions where the particles move along the equilateral triangular solutions of Lagrange, and there are the collinear solutions of Euler. These solutions would be rest solutions in a rotating coordinates system. Since the restricted 3-body problem is a limiting case in rotating coordinates, we expect to see vestiges of these solutions as equilibria.

From (C.11) an equilibrium solution for the restricted problem would satisfy

$$0 = y + Kx, \qquad 0 = Kx + \frac{\partial U}{\partial x}, \tag{14}$$

which implies

$$0 = x + \frac{\partial U}{\partial x} = \frac{\partial V}{\partial x}, \tag{15}$$

where V is the amended potential

$$V = \tfrac{1}{2}\|x\|^2 + U. \tag{16}$$

Thus, an equilibrium solution is a critical point of the amended potential. First, seek solutions that do not lie on the line joining the primaries. As in the discussion of the Lagrange point, use the distances d_1, d_2 given in (C.11) as coordinates. From (C.11) we obtain the identity

$$x_1^2 + x_2^2 = \mu d_1^2 + (1 - \mu)d_2^2 - \mu(1 - \mu); \tag{17}$$

so, V can be written in terms of the distances d_1 and d_2. The equation $\partial V/\partial x = 0$ becomes in these variables

$$\mu d_1 - \frac{\mu}{d_1^2} = 0, \qquad (1 - \mu)d_2 - \frac{(1 - \mu)}{d_2^2} = 0, \tag{18}$$

which clearly has the only solution $d_1 = d_2 = 1$. This solution lies at the vertex of an equilateral triangle whose base is the line segment joining the two primaries. Since there are two orientations, there are two such equilibria solutions; one in the lower half-plane denoted by $\mathscr{L}_4$, and one in the upper half-plane denoted by $\mathscr{L}_5$. These are attributed to Lagrange also.

Lagrange thought that these solutions had no astronomical significance, but in the twentieth century about 15 asteroids, the Trojans, were found at the $\mathscr{L}_4$ position and about 15 asteroids, the Greeks, were found at the $\mathscr{L}_5$ position in the sun Jupiter system. That is, one group of asteroids, the sun, and Jupiter form an equilateral triangle, approximately, and so does the other group with the sun and Jupiter.

Now consider equilibria along the line of the primaries where $x_2 = 0$. In this case the amended potential is a function of x_2 which we shall denote by x for the present and has the form

$$V = \frac{1}{2}x^2 \pm \frac{\mu}{(x - 1 + \mu)} \pm \frac{(1 - \mu)}{(x + \mu)}. \tag{19}$$

In the above, one takes the signs so that each term in the above is positive. There are three cases: (i) $x < -\mu$ where the signs are $-$ and $-$; (ii) $-\mu < x < 1 - \mu$ where the signs are $-$ and $+$; and (iii) $1 - \mu < x$ where the signs are $+$ and $+$. Clearly $V \to \infty$ as $x \to \pm\infty$, as $x \to -\mu$, or as $x \to 1 - \mu$. So V has at least one critical point on each of these three intervals. Also

$$\frac{d^2V}{dx^2} = 1 \pm \frac{\mu}{(x - 1 + \mu)^3} \pm \frac{(1 - \mu)}{(x + \mu)^3}, \tag{20}$$

where the signs are again taken so that each term is positive; so, V is a convex function. Therefore, V has precisely one critical point in each of these intervals, or three critical points. A sketch of the graph of V is given in Figure D.2. These three collinear equilibria are attributed to Euler also and are denoted by $\mathscr{L}_1$, $\mathscr{L}_2$, $\mathscr{L}_3$ as shown in Figure D.3. In classical celestial mechanics literature, these equilibrium points are called *libration points*; hence, the use of the symbol $\mathscr{L}$.

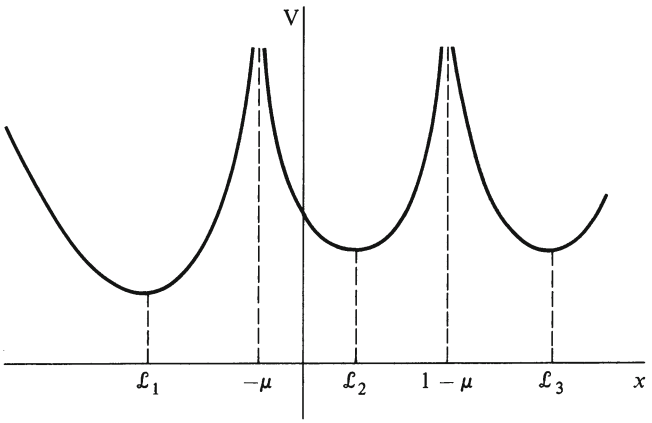

Figure D.2. The amended potential.

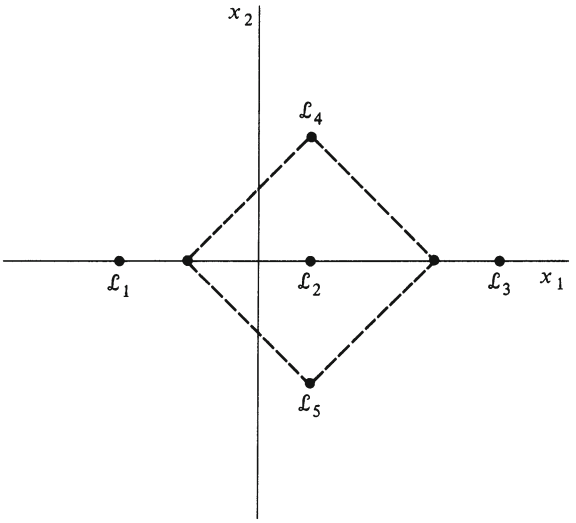

Figure D.3. The five equilibria of the restricted problem.

E. Further Reading

This chapter, and the book in general, assumes some knowledge of basic differential equations as found, for example, in the introductory texts: Brauer and Nohel (1969), Hurewicz (1958), or Sánchez (1968). They are all readable, short introductions to the geometric theory of differential equations and any one of them should give sufficient background. References to special advanced topics will be given as needed.

Pollard (1966) gives a clean and complete description of the solution to the 2-body problem, an introduction to Hamiltonian equations, and a brief treatment of the restricted problem. This short book is an ideal starting point for the study of Hamiltonian systems and celestial mechanics. A more elementary and classical introduction is found in Moulton (1914).

At the same or higher level of difficulty are the following: Abraham and Marsden (1978), an austere development of symplectic geometry which omits most of the details in its later chapters; Arnold (1983), an intuitive book which introduces many topics but lacks proofs at times; and Siegel and Moser (1971), a clearly written book with complete proofs. Of the three, Siegal and Moser is the book to read.

Problems

1. Let x, y, z be the usual coordinates in $\mathbb{R}^3$, $r = xi + yj + zk$, $X = \dot{x}$, $Y = \dot{y}$, $Z = \dot{z}$, $R = \dot{r} = Xi + Yj + Zk$.
   - **a.** Compute the three components of angular momentum $mr \times R$.
   - **b.** Compute the Poisson brakcet of any two of the components of angular momentum and show that it is plus/minus the third component of angular momentum.
   - **c.** Show that if a system admits two components of angular momentum as integrals, then the system admits all three components of angular momentum as integrals.

2. A *Lie algebra* $\mathscr{A}$ is a vector space with a product $*: \mathscr{A} \times \mathscr{A} \to \mathscr{A}$ which satisfies $a * b = -b * a$ (anticommutative), $a * (b + c) = a * b + a * c$ (distributive), $(\alpha a) * b = \alpha(a * b)$ (scalar associative), $a * (b * c) + b * (c * a) + c * (a * b) = 0$ (Jacobi's identity), where $a, b, c \in \mathscr{A}$ and $\alpha \in \mathbb{R}$ or $\mathbb{C}$.
   - **a.** Show that vectors in $\mathbb{R}^3$ form a Lie algebra where the product $*$ is the cross product.
   - **b.** Show that smooth functions on an open set in $\mathbb{R}^{2n}$ form a Lie algebra, where $f * g = \{f, g\}$, the Poisson bracket.
   - **c.** Show that the set of all $n \times n$ matrices, $gl(n, \mathbb{R})$, is a Lie algebra, where $A * B = AB - BA$, the Lie product.

3. The pendulum equation is $\ddot{\theta} + \sin\theta = 0$.
   - **a.** Show that $2I = \frac{1}{2}\dot{\theta}^2 + (1 - \cos\theta) = \frac{1}{2}\dot{\theta}^2 + 2\sin^2(\theta/2)$ is an integral.
   - **b.** Sketch the phase portrait.
   - **c.** Make the substitution $y = \sin(\theta/2)$ to get $\dot{y}^2 = (1 - y^2)(I - y^2)$. Show that when $0 < I < 1$, $y = k\,\mathrm{sn}(t, k)$ solves this equation when $k^2 = I$. Solve the pendulum equation in terms of the known function sn.

4. Using the definitions introduced in the section on Jacobi sine function:
   - **a.** Show

$$\varkappa = \int_0^1 \frac{d\tau}{\{(1 - \tau^2)(1 - k^2\tau^2)\}^{1/2}}.$$

b. In the integral above make the substitution $\tau = \sin u$ to get

$$\varkappa = \int_0^{\pi/2} \frac{du}{\{1 - k^2 \sin^2 u\}^{1/2}}.$$

c. Use the binomial series to expand the denominator in the above integrand in a series in $k^2 \sin^2 u$. Use Wallis' formula,

$$\frac{2}{\pi}\int_0^{\pi/2} \sin^{2n} u \, du = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)},$$

to integrate term by term to get

$$\varkappa = \frac{\pi}{2}\left\{1 + \left(\frac{1}{2}\right)^2 k^2 + \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 k^4 + \cdots\right\}.$$

5. Continue the notation of the previous problem.

a. Show that $\varkappa(k) \to \pi/2$ as $k \to 0$, and $\varkappa(k) \to \infty$ as $k \to 1$.

b. Show $\varkappa(k)$ is increasing in k.

c. Skecth a plot of $\varkappa$ vs k.

6. Show that the Kirchhoff problem can have equilibrium solutions. Discuss central configurations for the Kirchhoff problem. Find all central configurations for the Kirchhoff problem with three vortices.

7. Draw the complete phase portrait of the collinear Kepler problem. Using the formulas (I.B.8), solve the collinear Kepler problem.

8. Use the notation of Subsection I.C.3, the Kepler problem, so $A = q \times p$.

a. Use the identity in (I.C.8) to show that $\mu d(q/\|q\|) = \dot{p} \times A$. Integrate this to get $\mu(e + q/\|q\|\} = p \times A$, where e is a vector constant of integration. The vector e is a vector of integrals for the Kepler problem.

b. Show that if $A = 0$, then $e = -q/\|q\|$ or e points along the collinear motion.

c. Show that $e \cdot A = 0$, so e lies in the plane of motion and there is a relation among the integrals A and e. Henceforth, let the motion be in the plane.

d. Dot the expression for e with q to get $e \cdot q + \|q\| = \|A\|^2/\mu$.

e. Show that if $e = 0$, then the motion is circular, and the motion is with uniform speed.

f. Let $e \neq 0$, and let (r, θ) be the position of the particle in polar coordinates. Let $e = \varepsilon(\cos \omega, \sin \omega)$; so, $\varepsilon = \|e\|$, and ω is the argument of e. Let $f = \theta - \omega$; so, f is the angle of the particle measured from e. Show that $e \cdot q = \varepsilon r \cos f$ and so $r = (\|A\|^2/\mu)/(1 + \varepsilon \cos f)$. This is the equation of a conic section in polar coordinates with one focus at the origin. ε is the *eccentricity* of the orbit, $0 < \varepsilon < 1$ is an ellipse; $\varepsilon = 1$ is a parabola;, and $\varepsilon > 1$ is a hyperbola. e points to the point of closest approach of the particle to the origin (the *perihelion* if the sun is at the origin or the *perigee* if the earth is at the origin). The angle f is called the *true anomaly* and ω the *argument of the perigee.*

9. Let

$$K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix};$$

so,

$$\exp(Kt) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

Find a circular solution of the two-dimensional Kepler problem of the form $q = \exp(Kt)a$, where a is a constant vector.

10. Let $H\colon \mathbb{R}^{2n} \to \mathbb{R}$ be a globally defined conservative Hamiltonian, and assume that $H(z) \to \infty$ as $\|z\| \to \infty$. Show that all solutions of $\dot{z} = J\nabla H(z)$ are bounded.

11. Let $L = L(q, s, t)$ be a smooth real valued function of $q \in \mathbb{R}^m$, $s \in \mathbb{R}^m$ and $t \in \mathbb{R}$. The system of differential equations for $q(t)$ given by

$$\frac{d}{dt}\left(\frac{\partial L}{\partial s}\right)_{s=\dot{q}(t)} - \frac{\partial L}{\partial q} = 0$$

is called *Euler's equation.* It is often written

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0$$

for short. If $L = Q(q, \dot{q}) - P(q)$ where q is the generalized coordinates of a mechanical system, Q is the kinetic energy, and P is the potential energy, then L is called the *Lagrangian of the system,* and the equations are called *Lagrange's equations.*

a. Show that if $\partial^2 L/\partial s^2$ is nonsingular, then Euler's equation is a second-order differential equation in q. If $Q(q, \dot{q}) = \dot{q}^T M \dot{q}$, where M is a nonsingular, symmetric matrix, show that Lagrange's equations are Newton's equations.

b. Again assume that $\partial^2 L/\partial s^2$ is nonsingular. Show that the change of variables, known as *the Legendre transformation,* given by $(q, s) = (q, \dot{q}) \to (q, p)$ where $p = \partial L/\partial s$ transforms Euler's equations to the Hamiltonian system $\dot{q} = H_p$, $\dot{p} = -H_q$ where $H = p^T s - L(q, s, t)$. Hence, if $\partial^2 L/\partial s^2$ is nonsingular, then a Lagrangian system can be transformed to a Hamiltonian system.

12. If a particle is constrained to move on a surface in $\mathbb{R}^3$ without friction, then the force of constraint acts normal to the surface. If there are no external forces, then the particle is said to be a *free particle* on the surface and the only force acting on the particle is the force of constraint. In the free case, the acceleration is normal to the surface. In differential geometry a curve on a surface which minimizes distance (at least locally) is called a *geodesic of the surface,* and it can be shown that geodesics are characterized by the fact that their acceleration is normal to the surface. Thus, a free particle moves on a geodesic.

a. Consider a free particle on the 2-sphere $S^2 = \{x \in \mathbb{R}^3\colon \|x\| = 1\}$ so it moves to satisfy an equation of the form $\ddot{x} = \lambda x$, where λ is the scalar of proportionality. Show that $\lambda = -\|\dot{x}\|^2$, $x^T\dot{x} = 0$, and λ is constant along a solution (i.e., $d\lambda/dt = 0$).

b. Show that if the initial velocity is nonzero, then the solutions are great circles.

c. Show that the set of unit tangent vectors to S^2, called the *unit tangent bundle of* S^2 and denoted by T_1S^2, is an invariant set and is given by $\{(x, y) \in \mathbb{R}^3 \times \mathbb{R}^3\colon \|x\| = 1, \|y\| = 1, x^T y = 0\}$.

d. Show that the unit tangent bundle of the two sphere is the same as $SO(3, \mathbb{R})$, the *special orthogonal group.* $SO(3, \mathbb{R})$ is the set of all 3×3 orthogonal matrices

with determinant equal $+1$, or the set of all 3×3 rotation matrices. Hint: Think of the orthonormal frame consisting of the unit tangent, normal, and binormal vectors.

13. If the metric on a surface is given in local coordinates, $x = (x^1, x^2)$, by $ds^2 = \sum_{i,j=1}^{2} g_{ij}(x)\, dx^i\, dx^j$, where $\{g_{ij}(x)\}$ is a smooth 2×2 positive definite matrix, then define the Lagrangian by $L(x, \dot{x}) = (1/2)\sum_{i,j=1}^{2} g_{ij}(x)\dot{x}^i\dot{x}^j$. Show that the Euler–Lagrange equations are of the form

$$\ddot{x}^k + \sum_{i,j=1}^{2} \Gamma_{ij}^k(x)\dot{x}^i\dot{x}^j = 0,$$

where Γ_{ij}^k are the *Christoffel symbols*

$$\Gamma_{ij}^k = \sum_{s=1}^{2} g^{ks}\left\{\frac{\partial g_{si}}{\partial x_j} + \frac{\partial g_{sj}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_s}\right\},$$

and $\{g^{ij}(x)\}$ is the inverse of $\{g_{ij}(x)\}$.

14. Hill's lunar problem is defined by the Hamiltonian

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{1}{\|x\|} + (3x_1^2 - \|x\|^2),$$

where $x,\ y \in \mathbb{R}^2$. Write the equations of motion. Show that it has equilibrium points on the x_2 axis.

CHAPTER II

Linear Hamiltonian Systems

A. Preliminaries

In this chapter we study Hamiltonian systems which are linear differential equations. Many of the basic facts about Hamiltonian systems and symplectic geometry are easy to understand in this simple context. The basic linear algebra introduced in this chapter is the cornerstone of many of the later results on nonlinear systems. Some of the more advanced results which require a knowledge of multilinear algebra or the theory of analytic functions of a matrix are relegated to the appendices or references to the literature. These results are not important for the main development.

We assume a familiarity with the basic theory of linear algebra and linear differential equations. Let $gl(m, \mathbb{F})$ denote the set of all $m \times m$ matrices with entries in the field $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) and $Gl(m, \mathbb{F})$ the set of all nonsingular $m \times m$ matrices with entries in $\mathbb{F}$. $Gl(m, \mathbb{F})$ is a group under matrix multiplication and so is called the *general linear group*. (See the Problem Section for the definition and some basic facts about groups.) $I = I_m$ and $0 = 0_m$ will denote the $m \times m$ identity and zero matrices, respectively. In general, the subscript will be clear from the context.

In this theory a special role is played by the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{1}$$

Note that J is orthogonal and skew symmetric, i.e.,

$$J^{-1} = J^T = -J. \tag{2}$$

Let z be a coordinate vector in $\mathbb{R}^{2n}$, $\mathbb{I}$ an interval in $\mathbb{R}$, and $S: \mathbb{I} \to gl(2n, \mathbb{R})$

be continuous and symmetric. A *linear Hamiltonian system* is the system of $2n$ ordinary differential equations

$$\dot{z} = J\frac{\partial H}{\partial z} = JS(t)z = A(t)z, \tag{3}$$

where

$$H = H(t, z) = \tfrac{1}{2}z^T S(t)z \tag{4}$$

and $A(t) = JS(t)$. H, the *Hamiltonian*, is a quadratic form in the z's with coefficients which are continuous in $t \in \mathbb{I} \subset \mathbb{R}$. In the case when S, and hence H, is independent of t, then H is an integral for (3) by Theorem I.A.1.

Let $t_0 \in \mathbb{I}$ be fixed. From the theory of differential equations, for each $z_0 \in \mathbb{R}^{2n}$, there exists a unique solution $\phi(t, t_0, z_0)$ of (3) for all $t \in \mathbb{I}$ which satisfies the initial condition $\phi(t_0, t_0, z_0) = z_0$. Let $Z(t, t_0)$ be the $2n \times 2n$ fundamental matrix solution of (3) which satisfies $Z(t_0, t_0) = I$. Then $\phi(t, t_0, z_0) = Z(t, t_0)z_0$.

In case S and A are constants, we take $t_0 = 0$ and

$$Z(t) = e^{At} = \exp(At) = \sum_{i=1}^{\infty} \frac{A^n t^n}{n!}. \tag{5}$$

A matrix $A \in gl(2n, \mathbb{F})$ is called *Hamiltonian* (or *infinitesimally symplectic*) if

$$A^T J + JA = 0. \tag{6}$$

The set of all $2n \times 2n$ Hamiltonian matrices is denoted by $sp(n, \mathbb{R})$.

Theorem 1. *The following are equivalent*: (i) *A is Hamiltonian*, (ii) *$A = JR$ where R is symmetric*, (iii) *JA is symmetric. If A and B are Hamiltonian, then so are A^T, αA ($\alpha \in \mathbb{F}$), $A \pm B$, $[A, B] \equiv AB - AB$.*

PROOF. $A = J(-JA)$ and (6) is equivalent to $(-JA)^T = (-JA)$; thus, (i) and (ii) are equivalent. Since $J^2 = -I$, (ii) and (iii) are equivalent. Thus, the coefficient matrix $A(t)$ of the linear Hamiltonian system (3) is a Hamiltonian matrix. The first three parts of the next statement are easy. Let $A = JR$ and $B = JS$, where R and S are symmetric. Then $[A, B] = J(RJS - SJR)$ and $(RJS - SJR)^T = S^T J^T R^T - R^T J^T S^T = -SJR + RJS$ so $[A, B]$ is Hamiltonian. ∎

In the 2×2 case,

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \tag{7}$$

and so,

$$A^T J + JA = \begin{pmatrix} 0 & \alpha + \delta \\ -\alpha - \delta & 0 \end{pmatrix}.$$

So, a 2×2 matrix is Hamiltonian if and only if its trace, $\alpha + \delta$, is zero. If you write a second-order equation $\ddot{x} + p(t)\dot{x} + q(t)x = 0$ as a system in the usual

way with $\dot{x} = y$, $\dot{y} = -q(t)x - p(t)y$, then it is a linear Hamiltonian system when and only when $p(t) \equiv 0$. The $p(t)\dot{x}$ is usually considered the friction term.

Now let A be a $2n \times 2n$ matrix and write it in block form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{8}$$

and so

$$A^T J + JA = \begin{pmatrix} c - c^T & a^T + d \\ -a - d^T & b - b^T \end{pmatrix}.$$

So, A is Hamiltonian if and only if $a^T + d = 0$ and b and c are symmetric. In higher dimensions, being Hamiltonian is more restrictive than just having trace zero.

The function $[\cdot, \cdot]: gl(m, \mathbb{F}) \times gl(m, \mathbb{F}) \to gl(m, \mathbb{F})$ of Theorem 1 is called the *Lie product*. The second part of this theorem implies that the set of all $2n \times 2n$ Hamiltonian matrices, $sp(n, \mathbb{R})$, is a Lie algebra. We shall develop some interesting facts about Lie algebras of matrices in the problems, but they are not important to our main development.

A $2n \times 2n$ matrix T is called *symplectic with multiplier* μ if

$$T^T J T = \mu J, \tag{9}$$

where μ is a nonzero constant. If $\mu = +1$, then T is simply *symplectic*. The set of all $2n \times 2n$ symplectic matrices is denoted by $\mathrm{Sp}(n, \mathbb{R})$.

Theorem 2. *If T is symplectic with multiplier μ, then T is nonsingular and*

$$T^{-1} = -\mu^{-1} J T^T J. \tag{10}$$

If T and R are symplectic with multiplier μ and ν, respectively, then T^T, T^{-1}, and TR are symplectic with multiplier μ, μ^{-1}, and $\mu\nu$, respectively.

PROOF. Since the right-hand side, J, of (9) is nonsingular, T must be also. Formula (10) follows at once from (9). If T is symplectic, then from (10) one gets $T^T = -\mu J T^{-1} J$; so, $TJT^T = TJ(-\mu J T^{-1} J) = \mu J$. Thus, T^T is symplectic with multiplier μ. The remaining facts are proved in a similar manner. ∎

This theorem implies that $\mathrm{Sp}(n, \mathbb{R})$ is a group, a subgroup of $Gl(2n, \mathbb{R})$. Weyl says that originally he advocated the name "complex group" for $\mathrm{Sp}(n, \mathbb{R})$, but it became an embarrassment due to the collisions with the word "complex" in the connotation of complex number. "I therefore proposed to replace it by the corresponding greek adjective 'symplectic'" (Weyl, 1946, p. 165).

In the 2×2 case

$$T = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \tag{11}$$

and so

$$T^TJT = \begin{pmatrix} 0 & \alpha\delta - \beta\gamma \\ -\alpha\delta + \beta\gamma & 0 \end{pmatrix}.$$

So a 2×2 matrix is symplectic (with multiplier μ) if and only if it has determinant $+1$ (respectively μ). Thus, a 2×2 symplectic matrix defines a linear transformation which is orientation and area preserving.

Now let T be a $2n \times 2n$ matrix and write it in block form

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{12}$$

and so

$$T^TJT = \begin{pmatrix} a^Tc - c^Ta & a^Td - c^Tb \\ b^Tc - d^Ta & b^Td - d^Tb \end{pmatrix}.$$

So T is symplectic with multiplier μ if and only if $a^Td - c^Tb = \mu I$ and a^Tc and b^Td are symmetric. Being symplectic is more restrictive in higher dimensions. Formula (10) gives

$$T^{-1} = \mu^{-1}\begin{pmatrix} d^T & -b^T \\ -c^T & a^T \end{pmatrix}. \tag{13}$$

This reminds one of the formula for the inverse of a 2×2 matrix!

Theorem 3. *The fundamental matrix solution $Z(t, t_0)$ of a linear Hamiltonian system is symplectic for all t, $t_0 \in \mathbb{I}$. Conversely, if $Z(t, t_0)$ is a continuously differentiable function of symplectic matrices, then Z is a matrix solution of a linear Hamiltonian system.*

PROOF. Let $U(t) = Z(t, t_0)^TJZ(t, t_0)$. Since $Z(t_0, t_0) = I$, it follows that $U(t_0) = J$. $\dot{U}(t) = \dot{Z}^TJZ + Z^TJ\dot{Z} = Z^T(A^TJ + JA)Z = 0$; so, $U(t) \equiv J$.

If $Z^TJZ = J$ for $t \in \mathbb{I}$, then $\dot{Z}^TJZ + Z^TJ\dot{Z} = 0$; so, $(\dot{Z}Z^{-1})^TJ + J(\dot{Z}Z^{-1}) = 0$. This shows that $A = \dot{Z}Z^{-1}$ is Hamiltonian and $\dot{Z} = AZ$. ∎

Corollary 4. *The (constant) matrix A is Hamiltonian if and only if e^{At} is symplectic for all t.*

Change variables in the system (3) by $u = U(t)z$ where U is nonsingular and let $T(t)$ be the inverse of $U(t)$; so, $z = T(t)u$. Equation (3) becomes

$$\dot{u} = (T^{-1}AT - T^{-1}\dot{T})u. \tag{14}$$

In general this equation will not be Hamiltonian, however:

Theorem 5. *If U is symplectic with multiplier μ, then* (14) *is a Hamiltonian system with Hamiltonian $H(t, u) = \frac{1}{2}u^T[\mu T^TS(t)T + R(t)]u$, where $R(t) =$*

$-JT^{-1}\dot{T}$. *Conversely, if* (14) *is Hamiltonian for every Hamiltonian system* (3), *then* U *is symplectic with constant multiplier* μ.

PROOF. Let $U(t)$ be symplectic with multiplier μ; so, $T(t)$ is symplectic with multiplier μ^{-1}. Since $TJT^T = \mu^{-1}J$ for all t, $\dot{T}JT^T + TJ\dot{T}^T = 0$ or $(T^{-1}\dot{T})J + J(T^{-1}\dot{T})^T = 0$; so, $T^{-1}\dot{T}$ is Hamiltonian. Also $T^{-1}J = \mu JT^T$; so, $T^{-1}AT = T^{-1}JST = \mu JT^TST$, and so, $T^{-1}AT = J(\mu T^TST)$ is Hamiltonian also.

Now let (14) always be Hamiltonian. By taking $A \equiv 0$ we have that $T^{-1}\dot{T} = B(t)$ is Hamiltonian and T is a matrix solution of the Hamiltonian system

$$\dot{v} = vB(t). \tag{15}$$

So, $T(t) = KV(t, t_0)$, where $V(t, t_0)$ is the fundamental matrix solution of (15), and $K = T(t_0)$ is a constant matrix. By Theorem 3, V is symplectic.

Consider the change of variables $z = T(t)u = V(t, t_0)Ku$ as a two-stage change of variables: first $z = V(t, t_0)w$ and second $w = Ku$. The first transformation from z to w is symplectic, and so, by the first part of this theorem, preserves the Hamiltonian character of the equations. Since the first transformation is reversible, it would transform the set of all linear Hamiltonian systems onto the set of all linear Hamiltonian systems. Thus, the second transformation from w to u must always take a Hamiltonian system to a Hamiltonian system.

If $w = Ku$ transforms all Hamiltonian systems $\dot{w} = JCw$, C constant and symmetric, to a Hamiltonian system $\dot{u} = Dw$, then $D = K^{-1}JCK$ is Hamiltonian, and $JK^{-1}JCK$ is symmetric for all symmetric C. Thus,

$$\begin{gathered} JK^{-1}JCK = (JK^{-1}JCK)^T = K^TCJK^{-T}J, \\ C(KJK^TJ) = (JKJK^T)C, \\ CR = R^TC, \end{gathered} \tag{16}$$

where $R = KJK^TJ$. Fix i, $1 \leq i \leq 2n$ and take C to be the symmetric matrix which has $+1$ at the i, i position and zero elsewhere. Then the only nonzero row of CR is the ith, which is the ith row of R and the only nonzero column of R^TC is the ith, which is the ith column of R^T. Since these must be equal, the only nonzero entry in R or R^T must be on the diagonal. So R and R^T are diagonal matrices. Thus, $R = R^T = \operatorname{diag}(r_1, \ldots, r_{2n})$, and $RC - CR = 0$ for all symmetric matrices C. But $RC - CR = [(r_i - r_j)c_{ij}] = (0)$. Since c_{ij}, $i < j$, is arbitrary, $r_i = r_j$, or $R = -\mu I$ for some constant μ. $R = KJK^TJ = -\mu I$ implies $KJK^T = \mu J$. ∎

This is an example of a change of variables which preserves the Hamiltonian character of the system of equations. The general problem of which changes of variables preserve the Hamiltonian character is discussed in detail in the next chapter.

The fact that the fundamental matrix of (3) is symplectic means that the fundamental matrix must satisfy the identity (9). There are many functional relations in (9); so, there are functional relations between the solutions. The theorem given below is just one example of how these relations can be used. See Meyer and Schmidt (1982b) for some other examples.

Let $z_1, z_2: \mathbb{I} \to \mathbb{R}^{2n}$ be two smooth functions; we define the *Poisson bracket* of z_1 and z_2 to be

$$\{z_1, z_2\}(t) = z_1^T(t) J z_2(t); \tag{17}$$

so $\{z_1, z_2\}: \mathbb{I} \to \mathbb{R}^{2n}$ is smooth. The Poisson bracket is bilinear and skew symmetric. Two functions z_1 and z_2 are said to be in *involution* if $\{z_1, z_2\} \equiv 0$. A set of n linearly independent functions and pairwise in involution functions $z_1, \ldots, z_n$ are said to be a *Lagrangian set*. In general, the complete solution of a $2n$-dimensional system requires $2n$ linearly independent solutions, but for a Hamiltonian system a Lagrangian set of solutions suffices.

Theorem 6. *If a Lagrangian set of solutions of* (3) *is known, then a complete set of* $2n$ *linearly independent solutions can be found by quadrature. (See formula* (20).)

PROOF. Let $C = C(t)$ be the $2n \times n$ matrix whose columns are the n linearly independent solutions. Since the columns are solutions, $\dot{C} = AC$; since they are in involution, $C^T J C = 0$; and since they are independent, $C^T C$ is an $n \times n$ nonsingular matrix. Define the $2n \times n$ matrix by $D = JC(C^T C)^{-1}$. Then $D^T J D = 0$ and $C^T J D = -I$, and so $P = (D, C)$ is a symplectic matrix. Therefore,

$$P^{-1} = \begin{pmatrix} -C^T J \\ D^T J \end{pmatrix}.$$

Change coordinates by $z = P\zeta$ so that

$$\dot{\zeta} = P^{-1}(AP - \dot{P})\zeta = \begin{pmatrix} C^T S D + C^T J \dot{D} & 0 \\ -D^T S D - D^T J \dot{D} & 0 \end{pmatrix} \zeta. \tag{18}$$

All the submatrices in (18) are $n \times n$. The one in the upper left-hand corner is also zero, which can be seen by differentiating $C^T J D = -I$ to get $\dot{C}^T J D + C^T J \dot{D} = (AC)^T J D + C^T J \dot{D} = C^T S D + C^T J \dot{D} = 0$. Therefore,

$$\begin{aligned} \dot{u} &= 0, \\ \dot{v} &= -D^T(SD + J\dot{D})u, \end{aligned} \qquad \text{where } \zeta = \begin{pmatrix} u \\ v \end{pmatrix}, \tag{19}$$

which has a general solution $u = u_0, v = v_0 - V u_0$, where

$$V = \int_{t_0}^{t} D^T(SD + J\dot{D})\, dt. \tag{20}$$

A symplectic fundamental matrix solution of (3) is $Z = (D - CV, C)$. Thus, the complete set of solutions is obtained by performing the integration or quadrature in formula (20). ∎

This result is closely related to the general result given in a later chapter which says that k integrals in involution for a general Hamiltonian system can be used to reduce the number of degrees of freedom by k and, hence, the dimension by $2k$.

Recall that a nonsingular matrix T has two polar decompositions, $T = PO = O'P'$, where P and P' are positive definite matrices and O and O' are orthogonal matrices. These representations are unique. P is the unique positive definite square root of TT^T; P' is the unique positive definite square root of T^TT, $O = (TT^T)^{-1/2}T$, and $O' = T(T^TT)^{-1/2}$.

Theorem 7. *If T is symplectic, then the P, O, P', O' of the polar decomposition given above are all symplectic too.*

PROOF. The formula for T^{-1} in (10) is an equivalent condition for T to be symplectic. Let $T = PO$. Since $T^{-1} = -JT^TJ$, $O^{-1}P^{-1} = -JO^TP^TJ = (J^1O^TJ)(J^TP^TJ)$. In this last equation, the left-hand side is the product of an orthogonal matrix O^{-1} and a positive definite matrix P^{-1}, as is the right-hand side a product of an orthogonal matrix $J^{-1}OJ$ and a positive definite matrix J^TPJ. By the uniqueness of the polar representation, $O^{-1} = J^{-1}O^TJ = -JO^TJ$ and $P^{-1} = J^TPJ = -JP^TJ$. By (10) these last relations imply that P and O are symplectic. A similar argument gives that P' and O' are symplectic. ∎

Theorem 8. *The determinant of a symplectic matrix is $+1$.*

PROOF. Depending on how much linear algebra you know, this theorem is either easy or difficult. In Section D and Chapter 3 we give alternate proofs. Let T be symplectic. Formula (9) gives $\det(T^TJT) = \det T^T \det J \det T = (\det T)^2 = \det J = 1$ so $\det T = \pm 1$. The problem is to show that $\det T = +1$.

The determinant of a positive definite matrix is positive; so, by the polar decomposition theorem it is enough to show that an orthogonal, symplectic matrix has a positive determinant. So let T be orthogonal also.

Using the block representation in (12) for T, formula (13) for T^{-1}, and the fact that T is orthogonal, $T^{-1} = T^T$, one has that T is of the form

$$T = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \tag{21}$$

Define P as in (22) and compute

$$P = \frac{1}{\sqrt{2}}\begin{pmatrix} I & iI \\ I & -iI \end{pmatrix}, \qquad P^{-1} = \frac{1}{\sqrt{2}}\begin{pmatrix} I & I \\ -iI & iI \end{pmatrix},$$
$$PTP^{-1} = \begin{pmatrix} a-bi & 0 \\ 0 & a+bi \end{pmatrix}, \quad \det T = \det PTP^{-1} = \det(a-bi)\det(a+bi) > 0. \tag{22}$$

The last inequality follows because the left-hand side is a product of a complex number and its conjugate. ∎

B. Symplectic Linear Spaces

What is the matrix J? There are many different answers to this question depending on the context in which the question is asked. In this section we will answer this question from the point of view of abstract linear algebra. We shall present other answers later on, but certainly not all.

Let $\mathbb{V}$ be an m-dimensional vector space over the field $\mathbb{F}$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A *bilinear* form is a mapping $B: \mathbb{V} \times \mathbb{V} \to \mathbb{F}$ which is linear in both variables. A bilinear form is *skew symmetric* or *alternating* if $B(u, v) = -B(v, u)$ for all $u, v \in \mathbb{V}$. A bilinear form B is *nondegenerate* if $B(u, v) = 0$ for all $v \in \mathbb{V}$ implies $u = 0$. An example of an alternating bilinear form on $\mathbb{F}^m$ is $B(u, v) = u^T S v$, where S is any skew symmetric matrix.

Let B be a bilinear form and $e_1, \ldots, e_m$ a basis for $\mathbb{V}$. Given any vector $v \in \mathbb{V}$, we write $v = \sum \alpha_i e_i$ and define an isomorphism $\Phi: \mathbb{V} \to \mathbb{F}^m : v \to a = (\alpha_1, \ldots, \alpha_m)$. Define $s_{ij} = B(e_i, e_j)$ and S to be the $m \times m$ matrix $S = (s_{ij})$, the *matrix of B in the basis (e)*. Let $\Phi(u) = b = (\beta_1, \ldots, \beta_m)$, then $B(u, v) = \sum\sum \alpha_i \beta_j B(e_i, e_j) = a^T S b$. So in the coordinates defined by the basis (e_i), the bilinear form is just $a^T S b$, where S is the matrix $(B(e_i, e_j))$. If B is alternating, then S is skew symmetric, and if B is nondegenerate, then S is nonsingular and conversely. If you change basis by $e_i = \sum q_{ij} f_j$ and Q is the matrix $Q = (q_{ij})$, then the bilinear form B has the matrix R in the basis (f), where $R = QSQ^T$. One says that R and S are congruent (by Q). If Q is any elementary matrix so that premultiplication of S by Q is an elementary row operation, then post-multiplication of S by Q^T is the corresponding column operation. Thus, R is obtained from S by performing a sequence of row operations and the same sequence of column operations.

Theorem 1. *Let S be any skew symmetric matrix; then there exists a nonsingular matrix Q such that $R = QSQ^T = \operatorname{diag}(K, K, \ldots, K, 0, 0, \ldots, 0)$, where*

$$K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Or given an alternating form B where is a basis for $\mathbb{V}$ such that the matrix of B in this basis is R.

PROOF. If $S = 0$, we are finished. Otherwise, there is a nonzero entry which can be transferred to the (2, 1) position by interchanging rows. By scaling the second row, the (2, 1) position can be made 1. Performing the corresponding column operations yields a skew symmetric matrix with the 2×2 matrix K in the upper left-hand corner. Using row operations we can eliminate all the nonzero elements in the first two columns below the first two rows. Performing corresponding column operation yields a matrix of the form $\operatorname{diag}(K, S')$, where S' is a $(m-2) \times (m-2)$ skew symmetric matrix. Repeat the above argument on S'. ∎

Note that the rank of a skew symmetric matrix is always even; so, a nondegenerate, alternating bilinear form is defined on an even-dimensional space.

A *symplectic linear space*, or just a *symplectic space*, is a pair, $(\mathbb{V}, \omega)$ where $\mathbb{V}$ is a $2n$-dimensional vector space over the field $\mathbb{F}$, $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and ω is a nondegenerate alternating bilinear form on $\mathbb{V}$. The form ω is called the *symplectic form* or the *symplectic inner product*. Throughout the rest of this section we shall assume that $\mathbb{V}$ is a symplectic space with symplectic form ω. The standard example is $\mathbb{F}^{2n}$ and $\omega(x, y) = x^T J y$. In this example we shall write $\{x, y\} = x^T J y$ and call the space $(\mathbb{F}^{2n}, J)$ or simply $\mathbb{F}^{2n}$, if no confusion can arise.

A *symplectic basis* for $\mathbb{V}$ is a basis $v_1, \ldots, v_{2n}$ for $\mathbb{V}$ such that $\omega(v_i, v_j) = J_{ij}$, the i,jth entry of J. A symplectic basis is a basis so that the matrix of ω is just J. The standard basis $e_1, \ldots, e_{2n}$, where e_i is 1 in the ith position and zero elsewhere, is a symplectic basis for $(\mathbb{F}^{2n}, J)$. Given two symplectic spaces $(\mathbb{V}_i, \omega_i)$, $i = 1, 2$, a *symplectic isomorphism* or an *isomorphism* is a linear isomorphism $L: \mathbb{V}_1 \to \mathbb{V}_2$ such that $\omega_2(L(x), L(y)) = \omega_1(x, y)$ for all $x, y \in \mathbb{V}_1$—that is, L preserves the symplectic form. In this case we say that the two spaces are *symplectically isomorphic* or *symplectomorphic*.

Corollary 2. *Let $(\mathbb{V}, \omega)$ be a symplectic space of dimension $2n$. Then $\mathbb{V}$ has a symplectic basis. $(\mathbb{V}, \omega)$ is symplectically isomorphic to $(\mathbb{F}^{2n}, J)$, or all symplectic spaces of dimension $2n$ are isomorphic.*

PROOF. By Theorem 1 there is a basis for $\mathbb{V}$ such that the matrix of ω is $\operatorname{diag}(K, \ldots, K)$. Interchanging rows $2i$ and $n + 2i - 1$ and the corresponding columns brings the matrix to J. The basis for which the matrix of ω is J is a symplectic basis; so, a symplectic basis exists.

Let $v_1, \ldots, v_{2n}$ be a symplectic basis for $\mathbb{V}$ and $u \in \mathbb{V}$. There exist constants α_i such that $u = \sum \alpha_i v_i$. The linear map $L: \mathbb{V} \to \mathbb{F}^{2n} : u \to (\alpha_1, \ldots, \alpha_{2n})$ is the desired symplectic isomorphism. ∎

The study of symplectic linear spaces is really the study of one canonical example, e.g., $(\mathbb{F}^{2n}, J)$. Or put another way, J is just the coefficient matrix of the symplectic form in a symplectic basis. This is one answer to the question "What is J?."

If $\mathbb{V}$ is a vector space over $\mathbb{F}$, then f is a *linear functional* if, $f: \mathbb{V} \to \mathbb{F}$ is linear, $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$ for all $u, v \in \mathbb{V}$, and $\alpha, \beta \in \mathbb{F}$. Linear functionals are sometimes called *1-forms* or *covectors*. If $\mathbb{E}$ is the vector space of displacements of a particle in Euclidean space, then the work done by a force on a particle is a linear functional on $\mathbb{E}$. The usual geometric representation for a vector in $\mathbb{E}$ is a directed line segment. Represent a linear functional by showing its level planes as shown in Figure B.1. The value of the linear functional f on a vector v is represented by the number of level planes the vector crosses. The more level planes the vector crosses, the larger the value of f on v.

The set of all linear functionals on a space $\mathbb{V}$ is itself a vector space when addition and scalar multiplication is just the usual addition and scalar multi-

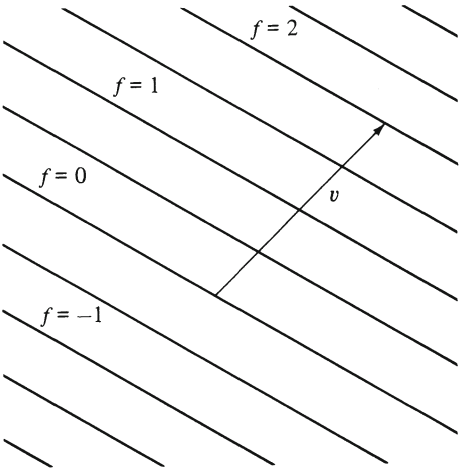

Figure B.1. A vector and a functional.

plication of functions. That is, if f and f' are linear functionals on $\mathbb{V}$ and $\alpha \in \mathbb{F}$, then define the linear functionals $f + f'$ and αf by the formulas $(f + f')(v) = f(v) + f'(v)$ and $(\alpha f)(v) = \alpha f(v)$. The space of all linear functionals is called the *dual space* (*to* $\mathbb{V}$) and will be denoted by $\mathbb{V}^*$. In the case when $\mathbb{V}$ is finite dimensional, $\mathbb{V}^*$ is finite dimensional with the same dimension. Let $u_1, \ldots, u_m$ be a basis for $\mathbb{V}$; then for any $v \in \mathbb{V}$, there are scalars $f^1, \ldots, f^m$ such that $v = f^1 u_1 + \cdots + f^m u_m$. The f^i are functions of v; so we will write $f^i(v)$, and they are linear. It is not too hard to show that $f^1, \ldots, f^m$ forms a basis for $\mathbb{V}^*$; this basis is called the *dual basis* (*dual to* $u_1, \ldots, u_m$). The defining property of this basis is $f^i(u_j) = \delta^i_j$ (the Kronecker delta function, defined by $\delta^i_j = 1$ if $i = j$ and zero otherwise.) If $\mathbb{W}$ is a subspace of $\mathbb{V}$ of dimension r, then define $\mathbb{W}^0 = \{f \in \mathbb{V}^* : f(e) = 0 \text{ for all } e \in \mathbb{W}\}$. $\mathbb{W}^0$ is called the *annihilators* of $\mathbb{W}$ and is easily shown to be a subspace of $\mathbb{V}^*$ of dimension $m - r$. Likewise, if $\mathbb{W}$ is a subspace of $\mathbb{V}^*$ of dimension r, then $\mathbb{W}^0 = \{e \in \mathbb{V} : f(e) = 0 \text{ for all } f \in \mathbb{W}^*\}$ is a subspace of $\mathbb{V}$ of dimension $m - r$. Also $\mathbb{W}^{00} = \mathbb{W}$. See any book on vector space theory for a complete discussion of dual spaces with proofs. See, for example, Halmos (1958).

Since ω is a bilinear form for each $v \in \mathbb{V}$, the function $\omega(v, \cdot)$: $\mathbb{V} \to \mathbb{R}$ is a linear functional and so is in the dual space $\mathbb{V}^*$. Since ω is nondegenerate, the map $\flat$: $\mathbb{V} \to \mathbb{V}^* : v \to \omega(v, \cdot) = v^\flat$ is an isomorphism. Let $\sharp$: $\mathbb{V}^* \to \mathbb{V} : v \to v^\sharp$ be the inverse of $\flat$. Sharp $\sharp$ and flat $\flat$ are musical symbols for raising and lowering notes and are used here because these isomorphisms are index raising and lowering operations in the classical tensor notation.

Let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Define $\mathbb{U}^\perp = \{v \in \mathbb{V} : \omega(v, \mathbb{U}) = 0\}$. Clearly, $\mathbb{U}^\perp$ is a subspace, $\{\mathbb{U}, \mathbb{U}^\perp\} = 0$ and $\mathbb{U} = \mathbb{U}^{\perp\perp}$.

Lemma 3. $\mathbb{U}^\perp = \mathbb{U}^{0\sharp}$; $\dim \mathbb{U} + \dim \mathbb{U}^\perp = \dim \mathbb{V} = 2n$.

PROOF.

$$\begin{aligned}\mathbb{U}^{\perp} &= \{x \in \mathbb{V} : \omega(x, y) = 0 \text{ for all } y \in \mathbb{U}\} \\ &= \{x \in \mathbb{V} : x^{\flat}(y) = 0 \text{ for all } y \in \mathbb{U}\} \\ &= \{x \in \mathbb{V} : x^{\flat} \in \mathbb{U}^0\} = \mathbb{U}^{0\sharp}.\end{aligned}$$

The second statement follows from $\dim \mathbb{U} + \dim \mathbb{U}^0 = \dim \mathbb{V}$ and the fact that $\sharp$ is an isomorphism. ∎

A *symplectic subspace* $\mathbb{U}$ of $\mathbb{V}$ is a subspace such that ω restricted to this subspace is nondegenerate. By necessity, $\mathbb{U}$ must be of even dimension, and so, $(\mathbb{U}, \omega)$ is a symplectic space.

Proposition 4. *If $\mathbb{U}$ is symplectic, then so is $\mathbb{U}^{\perp}$, and $\mathbb{V} = \mathbb{U} \oplus \mathbb{U}^{\perp}$. Conversely, if $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ and $\omega(\mathbb{U}, \mathbb{W}) = 0$, then $\mathbb{U}$ and $\mathbb{W}$ are symplectic.*

PROOF. Let $x \in \mathbb{U} \cap \mathbb{U}^{\perp}$; so, $\omega(x, y) = 0$ for all $y \in \mathbb{U}$, but $\mathbb{U}$ is symplectic so $x = 0$. Thus, $\mathbb{U} \cap \mathbb{U}^{\perp} = 0$. This, with Lemma 3, implies $\mathbb{V} = \mathbb{U} \oplus \mathbb{U}^{\perp}$. Now let $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ and $\omega(\mathbb{U}, \mathbb{W}) = 0$. If ω is degenerate on $\mathbb{U}$, then there is an $x \in \mathbb{U}$, $x \neq 0$, with $\omega(x, \mathbb{U}) = 0$. Since $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ and $\omega(\mathbb{U}, \mathbb{W}) = 0$, this implies $\omega(x, \mathbb{V}) = 0$ or that ω is degenerate on all of $\mathbb{V}$. This contradiction yields the second statement. ∎

A *Lagrangian space* $\mathbb{U}$ is a subspace of $\mathbb{V}$ of dimension n such that ω is zero on $\mathbb{U}$, i.e., $\omega(u, w) = 0$ for all $u, w \in \mathbb{U}$. A direct sum decomposition $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$, where $\mathbb{U}$ and $\mathbb{W}$ are Lagrangian spaces, is called a *Lagrangian splitting*, and $\mathbb{W}$ is called the *Lagrangian complement* of $\mathbb{U}$. In $\mathbb{R}^2$ any line through the origin is Lagrangian, and any other line through the origin is a Lagrangian complement.

Lemma 5. *Let $\mathbb{U}$ be a Lagrangian subspace of $\mathbb{V}$, then there exists a Lagrangian complement of $\mathbb{U}$.*

PROOF. The example above shows the complement is nonunique. Let $\mathbb{V} = \mathbb{F}^{2n}$ and $\mathbb{U} \subset \mathbb{F}^{2n}$. Then $\mathbb{W} = J\mathbb{U}$ is a Lagrangian complement to $\mathbb{U}$. ∎

Lemma 6. *Let $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ be a Lagrange splitting and $x_1, \ldots, x_n$ any basis for $\mathbb{U}$. Then there exists a unique basis $y_1, \ldots, y_n$ of $\mathbb{W}$ such that $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ is a symplectic basis for $\mathbb{V}$.*

PROOF. Define $\phi_i \in \mathbb{W}^0$ by $\phi_i(w) = \omega(x_i, w)$ for $w \in \mathbb{W}$. If $\sum \alpha_i \phi_i = 0$, then $\omega(\sum \alpha_i x_i, w) = 0$ for all $w \in \mathbb{W}$ or $\omega(\sum \alpha_i x_i, \mathbb{W}) = 0$. But since $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ and $\omega(\mathbb{U}, \mathbb{U}) = 0$, if follows that $\omega(\sum \alpha_i x_i, \mathbb{V}) = 0$. This implies $\sum \alpha_i x_i = 0$, since ω is nondegenerate, and this implies $\alpha_i = 0$, since the x_is' are independent. Thus, $\phi_1, \ldots, \phi_n$ are independent, and so, they form a basis for $\mathbb{W}^0$. Let $y_1, \ldots, y_n$ be the dual basis in $\mathbb{W}$; so, $\omega(x_i, y_j) = \phi_i(y_j) = \delta_{ij}$. ∎

A linear operator $L: \mathbb{V} \to \mathbb{V}$ is called *Hamiltonian* if

$$\omega(Lx, y) + \omega(x, Ly) = 0 \tag{1}$$

for all $x, y \in \mathbb{V}$. A linear operator $L: \mathbb{V} \to \mathbb{V}$ is called *symplectic* if

$$\omega(Lx, Ly) = \omega(x, y) \tag{2}$$

for all $x, y \in \mathbb{V}$. If $\mathbb{V}$ is the standard symplectic space $(\mathbb{F}^{2n}, J)$ and L is a matrix, then (1) means $x^T(L^TJ + JL)y = 0$ for all x and y. But this implies that L is a Hamiltonian matrix. On the other hand, if L satisfies (2) then $x^TL^TJLy = x^TJy$ for all x and y. But this implies L is a symplectic matrix. The matrix representation of a Hamiltonian (respectively symplectic) linear operator in a symplectic coordinate system is a Hamiltonian (respectively symplectic) matrix.

Lemma 7. *Let $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ be a Lagrangian splitting and $A: \mathbb{V} \to \mathbb{V}$ a Hamiltonian (respectively symplectic) linear operator which respects the splitting, i.e., $A: \mathbb{U} \to \mathbb{U}$ and $A: \mathbb{W} \to \mathbb{W}$. Choose any basis of the form given in Lemma 6; the matrix representation of A in these symplectic coordinates is of the form*

$$\begin{pmatrix} B^T & 0 \\ 0 & -B \end{pmatrix} \quad \left[\textit{respectively} \begin{pmatrix} B^T & 0 \\ 0 & B^{-1} \end{pmatrix} \right]. \tag{3}$$

PROOF. Since A respects the splitting, and the basis for $\mathbb{V}$ is the union of the bases for $\mathbb{U}$ and $\mathbb{W}$, the matrix representation for A must be in block-diagonal form. A Hamiltonian or symplectic matrix which is in block-diagonal form must be of the form given in (3)—see A.8 and A.12. ∎

C. The Spectra of Hamiltonian and Symplectic Operators

In this section we shall obtain some canonical forms for Hamiltonian and symplectic matrices in some simple cases. The complete picture is very detailed and would lead us too far afield to develop fully. We shall start with only real matrices, but sometimes we will need to go into the complex domain to finish the arguments. We simply assume that all our real spaces are imbedded in a complex space of the same dimension.

If A is Hamiltonian and T is symplectic, then $T^{-1}AT$ is Hamiltonian also. Thus, if we start with a linear constant coefficient Hamiltonian system $\dot{z} = Az$ and make the change of variables $z = Tu$, then in the new coordinates the equations become $\dot{u} = (T^{-1}AT)u$, which is again Hamiltonian. If $B = T^{-1}AT$, where T is symplectic, then we say that A and B are *symplectically similar*. We seek canonical forms for Hamiltonian and symplectic matrices under symplectic similarity. Since it is a form of similarity transformation, the eigenvalue structure plays an important role in the following discussion.

Since symplectic similarity is more restrictive than ordinary similarity, one should expect more canonical forms than the usual Jordan canonical forms. Consider, for example, the two Hamiltonian matrices

$$A_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{1}$$

both of which could be the coefficient matrix of a harmonic oscillator. In fact, they are both the real Jordan forms for the harmonic oscillator. The reflection $T = \text{diag}(1, -1)$ defines a similarity between these two, i.e., $T^{-1}A_1T = A_2$. Since the determinant of T is not $+1$, T is not symplectic. In fact, A_1 and A_2 are not symplectically equivalent. If $T^{-1}A_1T = A_2$, then $T^{-1}\exp(A_1t)T = \exp(A_2t)$, and T would take the clockwise rotation $\exp(A_1t)$ to the counterclockwise rotation $\exp(A_2t)$. But, if T were symplectic, its determinant would be $+1$ and thus would be orientation preserving. Therefore, T cannot be symplectic. Another way to see that the two Hamiltonian matrices in (1) are not symplectically equivalent is to note that $A_1 = JI$ and $A_2 = J(-I)$. So the symmetric matrix corresponding to A_1 is I, the identity, and to A_2 is $-I$. I is positive definite, whereas $-I$ is negative definite. If A_1 and A_2 where symplectically equivalent, then I and $-I$ would be congruent, which is clearly false.

A polynomial $p(\lambda) = a_m\lambda^m + a_{m-1}\lambda^{m-1} + \cdots + a_0$ is *even* if $p(-\lambda) = p(\lambda)$, which is the same as $a_k = 0$ for all odd k. If λ_0 is a zero of an even polynomial, then so is $-\lambda_0$; therefore, the zeros of a real, even polynomial are symmetric about the real and imaginary axes. The polynomial $p(\lambda)$ is a *reciprocal* polynomial if $p(\lambda) = \lambda^m p(\lambda^{-1})$, which is the same as $a_k = a_{m-k}$ for all k. If λ_0 is a zero of a reciprocal polynomial, then so is λ_0^{-1}; therefore, the zeros of a real, reciprocal polynomial are symmetric about the real axis and the unit circle (in the sense of inversion).

Proposition 1. *The characteristic polynomial of a real Hamiltonian matrix is an even polynomial. Thus, if λ is an eigenvalue of a Hamiltonian matrix, then so are $-\lambda$, $\bar{\lambda}$, $-\bar{\lambda}$. The characteristic polynomial of a real symplectic matrix is a reciprocal polynomial. Thus, if λ is a eigenvalue of a symplectic matrix, then so are λ^{-1}, $\bar{\lambda}$, $\bar{\lambda}^{-1}$.*

PROOF. Recall that $\det J = 1$. Let A be a Hamiltonian matrix; then $p(\lambda) = \det(A - \lambda I) = \det(JA^TJ - \lambda I) = \det(JA^TJ + \lambda JJ) = \det J \det(A + \lambda I) \det J = \det(A + \lambda I) = p(-\lambda)$.

Let T be a symplectic matrix; by Theorem A.8, $\det T = +1$. $p(\lambda) = \det(T - \lambda I) = \det(T^T - \lambda I) = \det(-JT^{-1}J - \lambda I) = \det(-JT^{-1}J + \lambda JJ) = \det(-T^{-1} + \lambda I) = \det T^{-1} \det(-I + \lambda T) = \lambda^{2n} \det(-\lambda^{-1}I + T) = \lambda^{2n}p(\lambda^{-1})$. ∎

Actually we can prove much more. By A.6, a Hamiltonian matrix A satisfies $A = J^{-1}(-A^T)J$; so, A and $-A^T$ are similar, and the multiplicity of

the eigenvalues λ_0 and $-\lambda_0$ are the same. In fact, the whole Jordan block structures will be the same for λ_0 and $-\lambda_0$. By A.10, a symplectic matrix T satisfies $T^{-1} = J^{-1}T^TJ$; so, T^{-1} and T^T are similar, and the multiplicity of the eigenvalues λ_0 and λ_0^{-1} are the same. The whole Jordan block structures will be the same for λ_0 and λ_0^{-1}.

Consider the linear constant coefficient Hamiltonian system of differential equations

$$\dot{x} = Ax, \tag{2}$$

where A is a Hamiltonian matrix and $Z(t) = e^{At}$ is the fundamental matrix solution. By the above it is impossible for all the eigenvalues of A to be in the left-hand plane, and, therefore, it is impossible for all the solutions to be exponentially decaying. Thus, the origin cannot be asymptotically stable.

Henceforth, let A be a real Hamiltonian matrix and T a real symplectic matrix. First we will develop the theory for Hamiltonian matrices and then the theory of symplectic matrices. Since eigenvalues are sometimes complex, it will be necessary to consider complex matrices at times, but we will always be concerned with the real answers in the end.

First consider the Hamiltonian case. Let λ be an eigenvalue of A, and define subspaces of $\mathbb{C}^{2n}$ by $\eta_k(\lambda) = \text{kernel}(A - \lambda I)^k$, $\eta^\dagger(\lambda) = \bigcup_{k=1}^{2n} \eta_k(\lambda)$. The eigenspace of A corresponding to the eigenvalue λ is $\eta(\lambda) = \eta_1(\lambda)$, and the generalized eigenspace is $\eta^\dagger(\lambda)$. If $\{x, y\} = x^TJy = 0$, then x and y are J-orthogonal.

Lemma 2. *Let λ and μ be eigenvalues of A with $\lambda + \mu \neq 0$, then $\{\eta(\lambda), \eta(\mu)\} = 0$. That is, the eigenvectors corresponding to λ and μ are J-orthogonal.*

PROOF. Let $Ax = \lambda x$, and $Ay = \mu y$, where $x, y \neq 0$. $\lambda\{x, y\} = \{Ax, y\} = x^TA^TJy = -x^TJAy = -\mu\{x, y\}$; and so, $(\lambda + \mu)\{x, y\} = 0$. ∎

Corollary 3. *Let A be a $2n \times 2n$ Hamiltonian matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n, -\lambda_1, \ldots, -\lambda_n$; then there exists a symplectic matrix S (possibly complex) such that $S^{-1}AS = \text{diag}(\lambda_1, \ldots, \lambda_n, -\lambda_1, \ldots, -\lambda_n)$.*

PROOF. Let $\mathbb{U} = \eta_1(\lambda_1) \cup \cdots \cup \eta_1(\lambda_n)$ and $\mathbb{W} = \eta_1(-\lambda_1) \cup \cdots \cup \eta_1(-\lambda_n)$; by the above, $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ is a Lagrange splitting, and A respects this splitting. Choose a symplectic basis for $\mathbb{V}$ by Lemma B.6. Changing to that basis is affected by a symplectic matrix G, i.e., $G^{-1}AG = \text{diag}(B^T, -B)$, where B has eigenvalues $\lambda_1, \ldots, \lambda_n$. Let C be such that $C^{T-1}B^TC^T = \text{diag}(\lambda_1, \ldots, \lambda_n)$, and define a symplectic matrix by $Q = \text{diag}(C^T, C^{-1})$. The required symplectic matrix is $S = GQ$. ∎

If complex transformations are allowed, then the two matrices in (1) can both be brought to $\text{diag}(i, -i)$ by a symplectic similarity, and to B; thus, one is symplectically similar to the other. However, they are not similar by a real symplectic similarity. Let us investigate the real case in detail.

A subspace $\mathbb{U}$ of $\mathbb{C}^n$ is called a *complexification (of a real subspace)* if $\mathbb{U}$ has a real basis. If $\mathbb{U}$ is a complexification, then there is a real basis $x_1, \ldots, x_k$ for $\mathbb{U}$, and for any $u \in \mathbb{U}$, there are complex numbers $\alpha_1, \ldots, \alpha_k$ such that $u = \alpha_1 x_1 + \cdots + \alpha_n x_n$. But then $\bar{u} = \bar{\alpha}_1 x_1 + \cdots + \bar{\alpha}_n x_n \in \mathbb{U}$ also. Conversely, if $\mathbb{U}$ is a subspace such that $u \in \mathbb{U}$ implies $\bar{u} \in \mathbb{U}$, then $\mathbb{U}$ is a complexification. Because if $x_1, \ldots, x_k$ is a complex basis with $x_j = u_j + v_j i$, then $u_j = (x_j + \bar{x}_j)/2$ and $v_j = (x_j - \bar{x}_j)/2i$ are in $\mathbb{U}$, and the totality of $u_i, \ldots, u_k, v_1, \ldots, v_k$ span $\mathbb{U}$. From this real spanning set, one can extract a real basis. Thus, $\mathbb{U}$ is a complexification if and only if $\mathbb{U} = \overline{\mathbb{U}}$ (i.e., $u \in \mathbb{U}$ implies $\bar{u} \in \mathbb{U}$).

Until otherwise said let A be a real Hamiltonian matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n, -\lambda_1, \ldots, -\lambda_n$, so 0 is not an eigenvalue. The eigenvalues of A fall into three groups: (1) the real eigenvalues $\pm\alpha_1, \ldots, \pm\alpha_s$, (2) the pure imaginary $\pm\beta_1 i, \ldots, \pm\beta_r i$, and (3) the truly complex $\pm\gamma_1 \pm \delta_1 i, \ldots, \pm\gamma_t \pm \delta_t i$. This defines a direct sum decomposition

$$\mathbb{V} = \left(\bigoplus_j \mathbb{U}_j\right) \oplus \left(\bigoplus_j \mathbb{W}_j\right) \oplus \left(\bigoplus_j \mathbb{Z}_j\right), \tag{3}$$

where

$$\begin{aligned} \mathbb{U}_j &= \eta(\alpha_j) \oplus \eta(-\alpha_j), \\ \mathbb{W}_j &= \eta(\beta_j i) \oplus \eta(-\beta_j i), \\ \mathbb{Z}_j &= \{\eta(\gamma_j + \delta_j i) \oplus \eta(\gamma_j - \delta_j i)\} \oplus \{\eta(-\gamma_j - \delta_j i) \oplus \eta(-\gamma_j + \delta_j i)\}. \end{aligned} \tag{4}$$

Each of the summands in (4) is an invariant subspace for A. By Lemma 2, each space is J-orthogonal to every other, and so by Proposition B.4 each space must be a symplectic subspace. Because each subspace is invariant under complex conjugation, each is the complexification of a real space. Thus, we can choose symplectic coordinates for each of the spaces, and A in these coordinates would be block diagonal. Therefore, the next task is to consider each space separately.

Lemma 4. *Let A be a 2×2 Hamiltonian matrix with eigenvalues $\pm\alpha$, α real, $\alpha \neq 0$. Then there exists a real 2×2 symplectic matrix S such that*

$$S^{-1}AS = \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix}. \tag{5}$$

PROOF. Let $Ax = \alpha x$, and $Ay = -\alpha y$, where x and y are nonzero. Since x and y are eigenvectors corresponding to different eigenvalues, they are independent. Thus, $\{x, y\} \neq 0$. Let $u = \{x, y\}^{-1} y$: so, x, u is a real symplectic basis, $S = (x, u)$ is a real symplectic matrix, and S is the matrix of the lemma. ∎

Lemma 5. *Let A be a real 2×2 Hamiltonian matrix with eigenvalues $\pm\beta i$, $\beta \neq 0$. Then there exists a real 2×2 symplectic matrix S such that*

$$S^{-1}AS = \begin{pmatrix} 0 & \beta \\ -\beta & 0 \end{pmatrix} \quad \text{or} \quad S^{-1}AS = \begin{pmatrix} 0 & -\beta \\ \beta & 0 \end{pmatrix}. \tag{6}$$

PROOF. Let $Ax = i\beta x$, $x = u + vi \neq 0$. So $Au = -\beta v$ and $Av = \beta u$. Since $u + iv$ and $u - iv$ are independent, u and v are independent. Thus, $\{u, v\} = \delta \neq 0$. If $\delta = -\gamma^2 < 0$, then define $S = (\gamma^{-1}u, \gamma^{-1}v)$ to get the first option in (6); or if $\delta = \gamma^2 > 0$, then define $S = (\gamma^{-1}v, \gamma^{-1}u)$ to get the second option in (6). ∎

Sometimes it is more advantageous to have a diagonal matrix than to have a real one; yet you want to keep track of the real origin of the problem. This is usually accomplished by *reality conditions* as defined in the next lemma.

Lemma 6. *Let A be a real 2×2 Hamiltonian matrix with eigenvalues $\pm\beta i$, $\beta \neq 0$. Then there exists a 2×2 matrix S and a matrix R such that*

$$S^{-1}AS = \begin{pmatrix} i\beta & 0 \\ 0 & -i\beta \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad S^TJS = \pm 2iJ, \qquad \bar{S} = SR. \quad (7)$$

PROOF. Let $Ax = i\beta x$, where $x \neq 0$. Let $x = u + iv$ as in the above lemma. Compute $\{x, \bar{x}\} = 2i\{v, u\} \neq 0$. Let $\gamma = 1/|\{v, u\}|$ and $S = (\gamma x, \gamma\bar{x})$. ∎

If S satifies (7), then S is said to satisfy *reality conditions with respect to* R. The matrix S is no longer a symplectic matrix but is what is called a symplectic matrix with multiplier $\pm 2i$. We shall discuss these types of matrices later. The matrix R is used to keep track of the fact that the columns of S are complex conjugates. We could require $S^TJS = +2iJ$ by allowing an interchange of the signs in (7).

Lemma 7. *Let A be a 4×4 Hamiltonian matrix with eigenvalue $\pm\gamma \pm \delta i$, $\gamma \neq 0, \delta \neq 0$. Then there exists a real 4×4 symplectic matrix S such that*

$$S^{-1}AS = \begin{pmatrix} B^T & 0 \\ 0 & -B \end{pmatrix},$$

where B is a real 2×2 matrix with eigenvalues $+\gamma \pm \delta i$.

PROOF. $\mathbb{U} = \eta(\gamma_j + \delta_j i) \oplus \eta(\gamma_j - \delta_j i)$ is the complexification of a real subspace and by Lemma 2 is Lagrangian. A restricted to this subspace has eigenvalues $+\gamma \pm \delta i$. A Lagrangian complement to $\mathbb{U}$ is $\mathbb{W} = \eta(-\gamma_j + \delta_j i) \oplus \eta(-\gamma_j - \delta_j i)$. Choose any real basis for $\mathbb{U}$ and complete it by Lemma B.6. The result follows from Lemma B.7. ∎

In particular you can choose coordinates so that B is in real Jordan form; so,

$$B = \begin{pmatrix} \gamma & \delta \\ -\delta & \gamma \end{pmatrix}.$$

This completes the case when A has distinct eigenvalues.

Next consider the symplectic case. Let λ be an eigenvalue of T, and define subspaces of $\mathbb{C}^{2n}$ by $\eta_k(\lambda) = \text{kernel}(T - \lambda I)^k$, $\eta^\dagger(\lambda) = \bigcup_1^{2n} \eta_k(\lambda)$. The eigenspace of T corresponding to the eigenvalue λ is $\eta(\lambda) = \eta_1(\lambda)$, and the generalized eigenspace is $\eta^\dagger(\lambda)$. Since the proof of the next set of lemmas is similar to those given just before, the proofs will be left as problems.

Lemma 8. *Let λ and μ be eigenvalues of T with $\lambda\mu \neq 1$; then $\{\eta(\lambda), \eta(\mu)\} = 0$. That is, the eigenvectors corresponding to λ and μ are J-orthogonal.*

Corollary 9. *Let T be a $2n \times 2n$ symplectic matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n, \lambda_1^{-1}, \ldots, \lambda_n^{-1}$; then there exists a symplectic matrix S (possibly complex) such that $S^{-1}TS = \text{diag}(\lambda_1, \ldots, \lambda_n, \lambda_1^{-1}, \ldots, \lambda_n^{-1})$.*

If complex transformations are allowed, then the two matrices,

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}, \quad \alpha^2 + \beta^2 = 1, \tag{8}$$

can both be brought to $\text{diag}(\alpha + \beta i, \alpha - \beta i)$ by a symplectic similarity, and thus, one is symplectically similar to the other. However, they are not similar by a real symplectic similarity. Let us investigate the real case in detail.

Until otherwise said, let T be a real symplectic matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n, \lambda_1^{-1}, \ldots, \lambda_n^{-1}$, so 1 is not an eigenvalue. The eigenvalues of T fall into three groups: (1) the real eigenvalues, $\mu_1^{\pm 1}, \ldots, \mu_s^{\pm 1}$, (2) the eigenvalues of unit modulus, $\alpha_1 \pm \beta_1 i, \ldots, \alpha_1 \pm \beta_r i$, and (3) the complex eigenvalues of modulus different from one, $\pm\gamma_1 \pm \delta_1 i, \ldots, \pm\gamma_t \pm \delta_t i$. This defines a direct sum decomposition

$$\mathbb{V} = \left(\bigoplus_j \mathbb{U}_j\right) \oplus \left(\bigoplus_j \mathbb{W}_j\right) \oplus \left(\bigoplus_j \mathbb{Z}_j\right), \tag{9}$$

where

$$\begin{aligned} \mathbb{U}_j &= \eta(\mu_j) \oplus \eta(\mu_j^{-1}), \\ \mathbb{W}_j &= \eta(\alpha_j + \beta_j i) \oplus \eta(\alpha_j - \beta_j i), \\ \mathbb{Z}_j &= \{\eta(\gamma_j + \delta_j i) \oplus \eta(\gamma_j - \delta_j i)\} \oplus \{\eta(\gamma_j - \delta_j i)^{-1} \oplus \eta(\gamma_j + \delta_j i)^{-1}\}. \end{aligned} \tag{10}$$

Each of the summands in (9) is an invariant subspace for T. By Lemma 8 each space is J-orthogonal to every other, and so each space must be a symplectic subspace. Because each subspace is invariant under complex conjugation, each is the complexification of a real space. Thus, we can choose symplectic coordinates for each of the spaces, and T in these coordinates would be block diagonal. Therefore, the next task is to consider each space separately.

Lemma 10. *Let T be a 2×2 symplectic matrix with eigenvalues $\mu^{\pm 1}$, μ real, $\mu \neq 1$. Then there exists a real 2×2 symplectic matrix S such that*

$$S^{-1}TS = \begin{pmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{pmatrix}. \tag{11}$$

Lemma 11. *Let T be a real 2×2 symplectic matrix with eigenvalues $\alpha \pm \beta i$, $\alpha^2 + \beta^2 = 1$, $\beta \neq 0$. Then there exists a real 2×2 symplectic matrix S such that*

$$S^{-1}TS = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \quad \text{or} \quad S^{-1}TS = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}. \tag{12}$$

Sometimes it is more advantageous to have a diagonal matrix than to have a real one; yet you want to keep track of the real origin of the problem. This is usually accomplished by *reality conditions* as defined in the next lemma.

Lemma 12. *Let T be a real 2×2 symplectic matrix with eigenvalues $\alpha \pm \beta i$, $\alpha^2 + \beta^2 = 1$, $\beta \neq 0$. Then there exists a 2×2 matrix S and a matrix R such that*

$$S^{-1}TS = \begin{pmatrix} \alpha + i\beta & 0 \\ 0 & \alpha - i\beta \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{13}$$

$S^T J S = \pm 2iJ$, *and* $\bar{S} = SR$.

Lemma 13. *Let T be a 4×4 symplectic matrix with eigenvalues $(\gamma \pm \delta i)^{-1}$, $\gamma^2 + \delta^2 \neq 1$, $\delta \neq 0$. Then there exists a real 4×4 symplectic matrix S such that*

$$S^{-1}TS = \begin{pmatrix} B^T & 0 \\ 0 & B^{-1} \end{pmatrix},$$

where B is a real 2×2 matrix with eigenvalues $+\gamma \pm \delta i$.

In particular you can choose coordinates so that B is in real Jordan form; so,

$$B = \begin{pmatrix} \gamma & \delta \\ -\delta & \gamma \end{pmatrix}.$$

This completes the case when T has distinct eigenvalues.

D. Nonelementary Divisors

In this section we shall indicate some extensions of the basic theory as given in Sections A–C. We will give references to the literature for further information.

Lemma 1. *Let A (respectively T) be any $2n \times 2n$ real Hamiltonian (respectively symplectic) matrix with eigenvalues λ and μ, $\lambda + \mu \neq 0$ (respectively $\lambda\mu \neq 1$). Then $\{\eta^\dagger(\lambda), \eta^\dagger(\mu)\} = 0$, i.e., the generalized eigenspaces are J-orthogonal.*

PROOF. Let A be Hamiltonian with eigenvalues λ and μ, $\lambda + \mu \neq 0$. By Lemma C.2, $\{\eta_1(\lambda), \eta_1(\mu)\} = 0$; so, make the induction assumption $\{\eta_k(\lambda), \eta_k(\mu)\} = 0$. We will first show that $\{\eta_{k+1}(\lambda), \eta_k(\mu)\} = 0$. Recall that $\{Ax, y\} = \{x, Ay\}$ for all $x, y \in \mathbb{V}$. Let $u \in \eta_{k+1}(\lambda)$ and $v \in \eta_k(\mu)$; so, $(A - \lambda I)^{k+1}u = 0$ and $(A - \mu I)^k v = 0$. Then

$$\begin{aligned} 0 = \{u, (A - \mu I)^k v\} &= \{u, (A + \lambda I + [-\lambda - \mu]I)^k v\} \\ &= \sum_{i=0}^{k} \binom{k}{i} (-\lambda - \mu)^{k-i} \{u, (A + \lambda I)^i v\} \\ &= \sum_{i=0}^{k} \binom{k}{i} (-\lambda - \mu)^{k-i} \{(-A + \lambda I)^i u, v\}. \end{aligned} \tag{1}$$

Since $u \in \eta_{k+1}(\lambda)$, $(-A + \lambda I)^i u \in \eta_k(\lambda)$ for $i = 1, 2, \ldots$; so, all the terms in the last sum in (1) are zero except the term when $i = 0$, and so, $(-\lambda - \mu)^k \{u, v\} = 0$. This proves $\{\eta_{k+1}(\lambda), \eta_k(\mu)\} = 0$. A similar argument shows that $\{\eta_{k+1}(\lambda), \eta_k(\mu)\} = 0$ implies $\{\eta_{k+1}(\lambda), \eta_{k+1}(\mu)\} = 0$; so, by induction the lemma holds. A similar argument holds for the symplectic matric T. See Laub and Meyer (1974). ∎

Let A be a real Hamiltonian matrix. The eigenvalues of A fall into four groups: (1) the eigenvalue 0, (2) the real eigenvalues $\pm\alpha_1, \ldots, \pm\alpha_s$, (3) the pure imaginary $\pm\beta_1 i, \ldots, \pm\beta_r i$, and (4) the truly complex $\pm\gamma_1 \pm \delta_1 i, \ldots, \pm\gamma_t \pm \delta_t i$. This defines a direct sum decomposition

$$\mathbb{V} = \mathbb{X} \oplus \left(\bigoplus_j \mathbb{U}_j\right) \oplus \left(\bigoplus_j \mathbb{W}_j\right) \oplus \left(\bigoplus_j \mathbb{Z}_j\right), \tag{2}$$

where

$$\begin{aligned} \mathbb{X} &= \eta^\dagger(0), \\ \mathbb{U}_j &= \eta^\dagger(\alpha_j) \oplus \eta^\dagger(-\alpha_j), \\ \mathbb{W}_j &= \eta^\dagger(\beta_j i) \oplus \eta^\dagger(-\beta_j i), \\ \mathbb{Z}_j &= \{\eta^\dagger(\gamma_j + \delta_j i) \oplus \eta^\dagger(\gamma_j - \delta_j i)\} \oplus \{\eta^\dagger(-\gamma_j - \delta_j i) \oplus \eta^\dagger(-\gamma_j + \delta_j i)\}. \end{aligned} \tag{3}$$

Each of the summands in (3) is an invariant subspace for A. By Lemma 1 each space is J-orthogonal to every other, and so by Proposition B.4, each space must be a symplectic subspace. Because each subspace is invariant under complex conjugation, each is the complexification of a real space. Thus, we can choose symplectic coordinates for each of the spaces where A in these coordinates would be block diagonal. Therefore, the next task would be to consider each space separately, but there are too many cases for a complete treatment.

We can group some of the subspaces together. Consider the decomposition

$$\begin{aligned}
\mathbb{V} &= \mathbb{I} \oplus \mathbb{N} \oplus \mathbb{P}, \\
\mathbb{I} &= \mathbb{X} \oplus \left(\bigoplus_j \mathbb{W}_j \right), \\
\mathbb{N} &= \left\{ \bigoplus_i \eta^\dagger(-\alpha_i) \right\} \oplus \left\{ \bigoplus_j [\eta^\dagger(-\gamma_j - \delta_j i) \oplus \eta^\dagger(-\gamma_j + \delta_j i)] \right\}, \\
\mathbb{P} &= \left\{ \bigoplus_i \eta^\dagger(+\alpha_i) \right\} \oplus \left\{ \bigoplus_j [\eta^\dagger(+\gamma_j - \delta_j i) \oplus \eta^\dagger(+\gamma_j + \delta_j i)] \right\}.
\end{aligned} \tag{4}$$

$\mathbb{I}$ is the symplectic, A-invariant subspace such that the restriction of A to $\mathbb{I}$ has only imaginary eigenvalues. $\mathbb{N} \oplus \mathbb{P}$ is the symplectic, A-invariant subspace such that the restriction of A to $\mathbb{N} \oplus \mathbb{P}$ has eigenvalues with nonzero real parts. $\mathbb{N}$ and $\mathbb{P}$ are Lagrangian subspaces of $\mathbb{N} \oplus \mathbb{P}$; so, the following lemma is a direct result of Lemma B.7.

Lemma 2. *Let A be a real, $2n \times 2n$, Hamiltonian matrix all of whose eigenvalues have nonzero real parts. Then there exists a real $2n \times 2n$ symplectic matrix P such that $P^{-1}AP = \operatorname{diag}(B^T, -B)$, where B is a real $n \times n$ matrix, all of whose eigenvalues have negative real parts. In particular, B could be taken in real Jordan form.*

There are many cases when A has eigenvalues with zero real parts, i.e., zero or pure imaginary. See Laub and Meyer (1974) for a complete discussion. In the case the eigenvalue zero is of multiplicity 2 or 4 the canonical forms are the 2×2 and 4×4 zero matrices and

$$\begin{pmatrix} 0 & \pm 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \pm 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}. \tag{5}$$

The corresponding Hamiltonians are

$$\pm \eta_1^2/2, \ \xi_2 \eta_1, \ \xi_2 \eta_1 \pm \eta_2^2/2. \tag{6}$$

In the case of double eigenvalue $\pm \alpha i$, $\alpha \neq 0$, the canonical forms in the 4×4 case are

$$\begin{bmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & \pm \alpha \\ -\alpha & 0 & 0 & 0 \\ 0 & \mp \alpha & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \alpha & 0 & 0 \\ -\alpha & 0 & 0 & 0 \\ \pm 1 & 0 & 0 & \alpha \\ 0 & \pm 1 & -\alpha & 0 \end{bmatrix}. \tag{7}$$

The corresponding Hamiltonians are

$$(\alpha/2)(\xi_1^2 + \eta_1^2) \pm (\alpha/2)(\xi_2^2 + \eta_2^2), \qquad \alpha(\xi_2\eta_1 - \xi_1\eta_2) \mp (\xi_1^2 + \xi_2^2)/2. \tag{8}$$

Let T be a real symplectic matrix. The eigenvalues of T fall into five groups: (1) $+1$, (2) -1, (3) the real eigenvalues $\mu_1^{\pm 1}, \ldots, \mu_s^{\pm 1}$, (4) the eigenvalues of unit modulus $\alpha \pm \beta_i i, \ldots, \alpha \pm \beta_r i$, and (5) the complex eigenvalues of modulus different from zero $(\gamma_1 \pm \delta_1 i)^{-1}, \ldots, (\gamma_t \pm \delta_t i)^{-1}$. This defines a direct sum decomposition

$$\mathbb{V} = \mathbb{X} \oplus \mathbb{Y} \oplus \left(\bigoplus_j \mathbb{U}_j\right) \oplus \left(\bigoplus_j \mathbb{W}_j\right) \oplus \left(\bigoplus_j \mathbb{Z}_j\right), \tag{9}$$

where

$$\begin{aligned}
&\mathbb{X} = \eta^\dagger(+1), \qquad Y = \eta^\dagger(-1),\\
&\mathbb{U}_j = \eta^\dagger(\mu_j) \oplus \eta^\dagger(\mu_j^{-1}),\\
&\mathbb{W}_j = \eta^\dagger(\alpha_j + \beta_j i) \oplus \eta^\dagger(\alpha_j - \beta_j i),\\
&\mathbb{Z}_j = \{\eta^\dagger(\gamma_j + \delta_j i) \oplus \eta^\dagger(\gamma_j - \delta_j i)\} \oplus \{\eta^\dagger(\gamma_j - \delta_j i)^{-1} \oplus \eta^\dagger(\gamma_j + \delta_j i)^{-1}\}.
\end{aligned} \tag{10}$$

Each of the summands in (10) is an invariant subspace for T. By Lemma 1, each space is J-orthogonal to every other, and so each space must be a symplectic subspace. Because each subspace is invariant under complex conjugation, each is the complexification of a real space. Thus, we can choose symplectic coordinates for each of the spaces where T in these coordinates would be block diagonal. Again, considering all of the cases would take us too far afield. Each of the spaces in (10) is symplectic and so, in particular, even dimensional. Thus, the multiplicity of the eigenvalue -1 is even; so, the restriction of T to this space has determinant $+1$. This gives an alternate proof of Theorem A.8. That is:

Corollary 3. *The determinant of a symplectic matrix is* $+1$.

Canonical forms for symplectic matrices are discussed in Laub and Meyer (1974). Also see Williamson (1936, 1937, 1939).

E. Periodic Systems and Floquet–Lyapunov Theory

In this section we shall introduce some of the vast theory of periodic Hamiltonian systems. A detailed discussion of periodic systems can be found in the two volume set by Yakubovich and Starzhinksii (1975).

Consider a periodic, linear Hamiltonian system

$$\dot{z} = J\frac{\partial H}{\partial z} = JS(t)z = A(t)z, \tag{1}$$

where

$$H = H(t, z) = \tfrac{1}{2}z^T S(t)z \tag{2}$$

and $A(t) = JS(t)$. Assume that A and S are continuous and T-periodic, i.e.,

$$A(t + T) = A(t), \qquad S(t + T) = S(t) \text{ for all } t \in \mathbb{R} \tag{3}$$

for some fixed $T > 0$. H, the *Hamiltonian*, is a quadratic form in the z's with coefficients which are continuous and T-periodic in $t \in \mathbb{R}$. Let $Z(t)$ be the fundamental matrix solution of (1) which satisfies $Z(0) = I$.

Lemma 1. $Z(t + T) = Z(t)Z(T)$ for all $t \in \mathbb{R}$.

PROOF. Let $X(t) = Z(t + T)$ and $Y(t) = Z(t)Z(T)$. $\dot{X}(t) = \dot{Z}(t + T) = A(t + T)Z(t + T) = A(t)X(t)$; so, $X(t)$ satisfies (1) and $X(0) = Z(T)$. $Y(t)$ also satisfies (1) and $Y(0) = Z(T)$. By the uniqueness theorem for differential equations, $X(t) \equiv Y(t)$. ∎

The above lemma only requires (1) to be periodic, not necessarily Hamiltonian. Even though the equations are periodic, the fundamental matrix need not be so, and the matrix $Z(T)$ is the measure of the nonperodicity of the solutions. $Z(T)$ is called the *monodromy matrix* of (1), and the eigenvalues of $Z(T)$ are called the (*characteristic*) *multipliers* of (1). The multipliers measure how much solutions are expanded, contracted, or rotated after a period. The monodromy matrix is symplectic by Theorem A.3, and so, the multipliers are symmetric with respect to the real axis and the unit circle by Theorem C.1. Thus, the origin cannot be asymptotically stable.

In order to understand periodic systems we need some information of logarithms of matrices. The complete proof requires a knowledge of the theory of functions of a matrix; so, the proof has been relegated to the Appendix. See Yakubovich and Stazhinskii (1975) or Sibuya (1960). Here we shall prove the result in the case when the matrices are diagonalizable.

A matrix R has a *logarithm* if there is a matrix Q such that $R = \exp Q$, and we write $Q = \ln R$. The logarithm is not unique in general, even in the real case, since $I = \exp O = \exp(2\pi J)$. If R has a logarithm, $R = \exp Q$, then R is nonsingular and has a square root $R^{1/2} = \exp(Q/2)$. The matrix

$$R = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

has no real square root and hence no real logarithm (see the Problem Section).

Theorem 2. *Let R be a nonsingular matrix; then there exists a matrix Q such that $R = \exp Q$. If R is real and has a square root, then Q may be taken as real. If R is symplectic, then Q may be taken as Hamiltonian.*

PROOF. We shall only prove this result in the case when R is symplectic and has distinct eigenvalues because in this case we only need consider the canonical forms of Section C. The complete proof, given in the Appendix to this

chapter, requires some of the theory of functions of a matrix. In particular,

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

is a real logarithm when $\mu > 0$ and complex when $\mu < 0$. A direct computation shows that

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

has no real square root when $\mu < 0$. If α and β satisfy $\alpha^2 + \beta^2 = 1$, then let θ be the solution of $\alpha = \cos\theta$ and $\beta = \sin\theta$ so that

$$\ln \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} = \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix},$$

$\ln \operatorname{diag}(B^T, B^{-1}) = \operatorname{diag}(\ln B^T, -\ln B)$, where

$$B = \begin{pmatrix} \gamma & \delta \\ -\delta & \gamma \end{pmatrix},$$

and

$$\ln B = \ln \rho \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}$$

is real where $\rho = \sqrt{(\gamma^2 + \delta^2)}$, and $\gamma = \rho \cos\theta$ and $\delta = \rho \sin\theta$. ∎

The monodromy matrix $Z(T)$ is nonsingular and symplectic, so there exists a Hamiltonian matrix K such that $Z(T) = \exp(KT)$. Define $X(t)$ by $X(t) = Z(t)\exp(-tK)$ and compute $X(t+T) = Z(t+T)\exp K(-t-T) = Z(t)Z(T)\exp(-KT)\exp(-Kt) = Z(t)\exp(-Kt) = X(t)$. Therefore, $X(t)$ is T-periodic. Since $X(t)$ is the product of two symplectic matrices, it is symplectic. In general, X and K are complex even if A and Z are real. To ensure a real decomposition, note that by Lemma 1, $Z(2T) = Z(T)Z(T)$; so, $Z(2T)$ has a real square root. Define K as the real solution of $Z(2T) = \exp(2KT)$ and $X(t) = Z(t)\exp(-Kt)$. Then X is $2T$ periodic.

Theorem 3 (Floquet–Lyapunov). *The fundamental matrix solution $Z(t)$ of the Hamiltonian* (1) *which satisfies $Z(0) = I$ is of the form $Z(t) = X(t)\exp(Kt)$, where $X(t)$ is symplectic and T-periodic and K is Hamiltonian. Real $X(t)$ and K can be found by taking $X(t)$ to be $2T$-periodic.*

Let Z, X and K be as above. In Equation (1) make the symplectic, periodic change of variables $z = X(t)w$; so,

$$\begin{aligned} \dot{z} &= \dot{X}w + X\dot{w} = (\dot{Z}e^{-Kt} - Ze^{-Kt}K)w + Ze^{-Kt}\dot{w} \\ &= AZe^{-Kt}w - Ze^{-Kt}Kw + Ze^{-Kt}\dot{w} = Az = AXw = AZe^{-Kt}w \end{aligned}$$

and hence

$$-Ze^{-Kt}Kw + Ze^{-Kt}\dot{w} = 0$$

or

$$\dot{w} = Kw.$$

Corollary 4. *The symplectic periodic change of variables* $z = X(t)w$ *transforms the periodic Hamiltonian system* (1) *to the constant Hamiltonian system* (4). *If* X *is* $2T$*-periodic, then* $X(t)$ *and* K *can be taken as real.*

The eigenvalues of K are called the (*characteristic*) *exponents* of (1) where K is taken as $\ln[X(T)/T]$ even in the real case. The exponents are the logarithms of the multipliers and so are defined modulo $2\pi i$.

F. Parametric Stability

Stability questions for Hamiltonian systems have been studied since the time of Newton. Is the solar system stable? This is an easy question to ask with obvious consequences, but is difficult to answer. We have seen some simple results—some positive and some negative. A satisfactory stability theory for Hamiltonian systems exists for linear autonomous and periodic systems only. The richness of this theory in this simplest of all cases foreshadows the complexity of the nonlinear problem. A detailed discussion of periodic systems can be found in the two volume set by Yakubovich and Stazhinskii (1975).

By the Floquet–Lyapunov theory, a periodic Hamiltonian system can be transformed by a linear, symplectic, periodic change of variables to an autonomous (constant coefficient) Hamiltonian system. In theory, at least, the stability question for a periodic system can be reduced to an autonomous system; so, we shall only discuss this case.

Consider the general linear system

$$\dot{x} = Ax, \tag{1}$$

where A is a constant square matrix. Solutions of such an equation are linear combinations of the basic solutions of the form $t^k \exp(\lambda t)a$, where k is a nonnegative integer, a is a constant vector, and λ is an eigenvalue of A. All solutions of (1) will tend to 0 as $t \to \infty$ (the origin is asymptotically stable) if and only if all the eigenvalues of A have negative real parts. By Theorem C.1 this never happens for a Hamiltonian system. All solutions of (1) are bounded for $t > 0$ if and only if (i) all the eigenvalues of (1) have nonpositive real parts and (ii) if λ is an eigenvalue of A with a zero real part (pure imaginary), then the k in the basic solutions, $t^k \exp(\lambda t)a$, is zero. This last condition, (ii), is equivalent to the condition that the Jordan blocks for all the pure imaginary eigenvalues of A in the Jordan canonical form for A are diagonal. That is, there are

no off-diagonal terms in the Jordan blocks for pure imaginary eigenvalues of A. For Hamiltonian systems by Theorem C.1, again, if all the eigenvalues have nonpositive real parts, then they must be pure imaginary. Thus, if a Hamiltonian system has all solutions bounded for $t > 0$, then all solutions are bounded for all time. So for linear Hamiltonian systems, the meaningful concept of stability is that all solutions are bounded for all $t \in \mathbb{R}$, or the origin is positively and negatively stable.

Henceforth, let A in (1) be a $2n \times 2n$ constant Hamiltonian matrix. The Hamiltonian system (1) is said to be *stable* if all solutions of (1) are bounded for all $t \in \mathbb{R}$. By the above discussion we have the following.

Theorem 1. *A linear constant Hamiltonian system* (1) *is stable if and only if all the eigenvalues of A are pure imaginary, and A is diagonalizable (over the complex numbers).*

Corollary 2. *A linear periodic Hamiltonian system is stable if and only if the monodromy matrix has eigenvalues with unit modulus and is diagonalizable.*

PROOF. The change of variables given in Corollary D.4, which reduced the periodic system (D.1) to the constant system (D.4), is periodic and continuous and so takes bounded solutions to bounded solutions. The matrix K of the constant system (D.4) is the logarithm of the monodromy matrix or its square, and, so, the eigenvalues of K are logarithms of the eigenvalues of the monodromy matrix or its square. Thus, the eigenvalues of the monodromy matrix are of unit modulus if and only if the eigenvalues of K are pure imaginary. A matrix is diagonalizable if and only if its logarithm is. ∎

The question of stability of a single linear autonomous or periodic Hamiltonian system is completely answered in theory. The problem of performing the calculations is another matter. If Equations (1) are the mathematical model of a physical problem, then the coefficients in the equation, i.e., the matrix A, may not be known exactly. Is the question of stability sensitive to small changes in the Hamiltonian matrix A? This question gives rise to the following concept. The constant linear Hamiltonian system (1) is said to be *parametrically stable* if it and all sufficiently small constant linear Hamiltonian perturbations of it are stable. That is, (1) is parametrically stable if there is an $\varepsilon > 0$ such that $\dot{x} = Bx$ is stable, where B is any constant Hamiltonian matrix with $\|A - B\| < \varepsilon$.

Lemma 3. *If the Hamiltonian H is positive (or negative) definite, then A is parametrically stable.*

PROOF. Let $H = x^T Sx/2$ be the Hamiltonian of (1); so, $A = JS$. Since H is positive (respectively negative) definite, the level set $H = h$, where h is a positive (respectively negative) constant, is an ellipsoid in $\mathbb{R}^{2n}$ and hence a bounded

set. Since H is an integral, any solution which starts on $H = h$ remains on $H = h$ and so is bounded. So H being positive definite implies (1) is stable. Any sufficiently small perturbation of a positive definite matrix is positive definite, and so any sufficiently small perturbation of (1) is stable also. ∎

Lemma 4. *If* (1) *is parametrically stable, then the eigenvalues of A must be pure imaginary, and A must be diagonalizable.*

PROOF. If (1) is parametrically stable, then it is stable. ∎

Lemma 5. *If* (1) *is parametrically stable, then zero is not any eigenvalue of A.*

PROOF. Assume not; so, assume that $\eta^\dagger(0)$ is not trivial. By the discussion in Section E, the subspace $\eta^\dagger(0)$ is an A-invariant symplectic subspace; so, A restricted to this subspace, denoted by A', is Hamiltonian. Since A is diagonalizable so is A'. So A' is a diagonalizable matrix all of whose eigenvalues are zero or A' is the zero matrix. Let B be a Hamiltonian matrix with real eigenvalues ± 1; then εB is a small perturbation of $A' = 0$ for small ε and has eigenvalues $\pm\varepsilon$. Thus, by perturbing A along the subspace $\eta^\dagger(0)$ by εB and leaving A fixed on the other subspaces we give a small Hamiltonian perturbation which is not stable. ∎

Let A have distinct eigenvalues $\pm\beta_1 i, \ldots, \pm\beta_s i$. The space $\eta^\dagger(+\beta_j i) \oplus \eta^\dagger(-\beta_j i)$ is the complexification of a real space $\mathbb{V}_j$ and A restricted to $\mathbb{V}_j$ will be denoted by A_j. $\mathbb{V}_j$ is a symplectic linear space, and A_j is invariant. A_j is Hamiltonian with eigenvalues $\pm\beta i$.

Theorem 6. *Using the notation given above, system* (1) *is parametrically stable if and only if the Hamiltonian A_j is positive or negative definite for each j.*

Thus, the systems defined by the Hamiltonians $H = 4(x_1^2 + y_1^2) - (x_2^2 + y_2^2)$ and $H = (x_1^2 + y_1^2) + (x_2^2 + y_2^2)$ are parametrically stable, whereas the system defined by the Hamiltonian $H = (x_1^2 + y_1^2) - (x_2^2 + y_2^2)$ is not parametrically stable.

Idea of Proof

We cannot give a complete proof here, but the ideas are simple. First, the if part. Given A and the decomposition of $\mathbb{V}$ into the invariant symplectic subspaces $\mathbb{V}_1, \ldots, \mathbb{V}_s$, there is an ε so small that if B is any Hamiltonian ε-perturbation of A, then there are B-invariant symplectic spaces $\mathbb{W}_1, \ldots, \mathbb{W}_s$. Moreover, $\dim \mathbb{V}_j = \dim \mathbb{W}_j$, $\mathbb{V}_j$ and $\mathbb{W}_j$ are closed and the eigenvalues of B restricted to $\mathbb{W}_j$ are close to $\pm\beta_j i$. Given these facts, it is clear that the Hamiltonian for the B system is positive or negative definite on each $\mathbb{W}_j$, and so all the solutions are bounded.

Second, the only if part. What we need to show is that if the Hamiltonian is not definite on one of the spaces A_j, then some perturbation will be unstable. We will show how the perturbation is constructed in one typical case.

Consider the Hamiltonian of two harmonic oscillators with equal frequencies; namely,

$$2H = (x_1^2 + y_1^2) \pm (x_2^2 + y_2^2). \tag{2}$$

There are two choices of the sign in (2). If the plus sign is taken, then the Hamiltonian is positive definite, and, in this case, (2) is parametrically stable. If the minus sign is taken, then the perturbation

$$2H = (x_1^2 + y_1^2) \pm (x_2^2 + y_2^2) + \varepsilon y_1 y_2 \tag{3}$$

is unstable for small ε because the characteristic equation of the system with Hamiltonian (3) is $(\lambda^2 + 1)^2 + \varepsilon^2$; and so, the eigenvalues are $\pm\sqrt{(-1 \pm \varepsilon i)}$, which has a real part nonzero for $\varepsilon \neq 0$.

G. The Critical Points in the Restricted Problem

In Section I.D.4 it was shown that the restricted problem of three bodies has five equilibrium points. They are the three collinear points $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}_3$, and the two triangular points $\mathscr{L}_4$ and $\mathscr{L}_5$. We will use the methods developed in this chapter to investigate the behavior of solutions near these equilibria. Only if the corresponding linearized system has periodic solutions can we hope to find solutions of the full nonlinear system which will liberate near one of these equilibrium points.

The Hamiltonian function of the restricted problem of three bodies is

$$H = \tfrac{1}{2}(y_1^2 + y_2^2) + x_2 y_1 - x_1 y_2 - U, \tag{1}$$

where U is the self-potential given by

$$U = \frac{1-\mu}{d_1} + \frac{\mu}{d_2} \tag{2}$$

with

$$d_1^2 = (x_1 + \mu)^2 + x_2^2 \quad \text{and} \quad d_2^2 = (x_1 + \mu - 1)^2 + x_2^2. \tag{3}$$

If x_1, x_2 is a critical point of the amended potential,

$$V = \tfrac{1}{2}(x_1^2 + x_2^2) + U(x_1, x_2), \tag{4}$$

then $x_1, x_2, y_1 = -x_2, y_2 = x_1$ is an equilibrium point. Let ξ_1 and ξ_2 be one of the five critical points of (4). In order to study the motion near this equilibrium point, we translate to new coordinates by

$$\begin{aligned} u_1 &= x_1 - \xi_1, & v_1 &= y_1 + \xi_2, \\ u_2 &= x_2 + \xi_2, & v_2 &= y_2 - \xi_1. \end{aligned} \tag{5}$$

This translation to the new coordinates (u_1, u_2, v_1, v_2) is obviously symplectic; so, we can perform this change of coordinates in the Hamiltonian (1) and preserve its structure. Expanding through second-order terms in the new variables, we obtain

$$H = \tfrac{1}{2}(v_1^2 + v_2^2) + u_2 v_1 - u_1 v_2 - \tfrac{1}{2}(U_{x_1x_1}u_1^2 + 2U_{x_1x_2}u_1u_2 + U_{x_2x_2}u_2^2) + \cdots. \tag{6}$$

There are no linear terms because the expansion is performed near an equilibrium and the constant term has been omitted because it contributes nothing in forming the corresponding system of differential equations. The above quadratic Hamiltonian function gives rise to the following Hamiltonian matrix:

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ U_{x_1x_1} & U_{x_1x_2} & 0 & 1 \\ U_{x_1x_2} & U_{x_2x_2} & -1 & 0 \end{bmatrix}. \tag{7}$$

The eigenvalues of this matrix determine the behavior of the linearized system of (1). With the help of (4), we obtain for the characteristic equation of (7)

$$\lambda^4 + (4 - V_{x_1x_1} - V_{x_2x_2})\lambda^2 + V_{x_1x_1}V_{x_2x_2} - V_{x_1x_2}^2 = 0. \tag{8}$$

The partial derivatives are

$$\begin{aligned} V_{x_1x_1} &= 1 + (1-\mu)\frac{3(x_1+\mu)^2 - d_1^2}{d_1^5} + \mu\frac{3(x_1+\mu-1)^2 - d_2^2}{d_2^5}, \\ V_{x_1x_2} &= 3x_1x_2\left(\frac{1-\mu}{d_1^5} + \frac{\mu}{d_2^5}\right), \\ V_{x_2x_2} &= 1 + (1-\mu)\frac{3x_2^2 - d_1^2}{d_1^5} + \mu\frac{3x_2^2 - d_2^2}{d_2^5}. \end{aligned} \tag{9}$$

They have to be evaluated at the critical points. Thus, we have to consider the collinear points and the triangular points separately.

Lemma 1. *At the collinear points, the matrix* (7) *has two real eigenvalues and two purely imaginary eigenvalues.*

PROOF. By direct computation one finds that for the collinear points $V_{x_1x_1} = 1 + 2(1-\mu)d_1^{-3} + 2\mu d_2^{-3} > 0$, $V_{x_1x_2} = 0$, and $V_{x_2x_2} = 1 - (1-\mu)d_1^{-3} - \mu d_2^{-3} < 0$. Only the last statement requires some additional work. We will present it for $\mathscr{L}_1$ and leave the other cases as exercises.

If $(\xi_1, 0)$ are the coordinates of the Lagrangian point $\mathscr{L}_1$, then $d_1 = \xi_1 + \mu$, $d_2 = \xi_1 - 1 + \mu$, and ξ_1 is the real solution of $V_{x_1} = 0$, that is, of a quintic polynomial

$$\xi_1 - (1-\mu)d_1^{-2} - \mu d_2^{-2} = 0. \tag{10}$$

We will use this relationship in the form

$$(1-\mu)d_1^{-2} = d_1 - \mu d_2^{-2} - \mu \tag{11}$$

when we evaluate the second derivative of V at $(\xi_1, 0)$, that is, we get

$$\begin{aligned} V_{x_2x_2} &= 1 - \frac{1}{d_1}(d_1 - \mu d_2^{-2} - \mu) - \mu d_2^{-3} \\ &= \frac{\mu}{d_1}(1 + d_2^{-2} - d_1 d_2^{-3}) \\ &= \frac{\mu}{d_1}(1 - d_2^{-3}) < 0. \end{aligned} \tag{12}$$

The last equality follows from $d_1 = 1 + d_2$ and the inequality follows then from the fact that $0 < d_2 < 1$.

Setting $A = 2 - \frac{1}{2}(V_{x_1x_1} + V_{x_2x_2})$ and $B = -V_{x_1x_1}V_{x_2x_2}$, the characteristic equation for the collinear points takes on the form

$$\lambda^4 + 2A\lambda^2 - B = 0 \tag{13}$$

with the solutions

$$\lambda^2 = -A \pm \sqrt{A^2 + B}. \tag{14}$$

Since $B > 0$ the statement of the lemma follows. It also means that the collinear points of Euler are unstable. Therefore, some solutions which start near to the Euler points will tend away from these points as time tends to infinity. ∎

Lemma 2. *At the triangular equilibrium points, the matrix* (8) *has purely imaginary eigenvalues for values of the mass ratio* μ *in the interval* $0 < \mu < \mu_1$, *where* $\mu_1 = \frac{1}{2}(1 - \sqrt{69/9})$ *is called* Routh's *critical mass ratio. For* $\mu = \mu_1$ *the matrix has the repeated eigenvalues* $+i\sqrt{2}/2$ *and* $-i\sqrt{2}/2$ *with nonelementary divisors. For* $\mu_1 < \mu \leq \frac{1}{2}$, *the eigenvalues are off the imaginary axis.*

PROOF. Since the coordinates for the Lagrangian point $\mathscr{L}_4$ have been found to be $\xi_1 = \frac{1}{2} - \mu$ and $\xi_2 = \frac{1}{2}\sqrt{3}$, the second derivatives of V can be computed explicitly. They are

$$V_{x_1x_1} = \frac{3}{4}, \qquad V_{x_1x_2} = -\frac{3\sqrt{3}}{4}(1 - 2\mu), \qquad V_{x_2x_2} = \frac{9}{4}. \tag{15}$$

The characteristic equation for (8) is then

$$\lambda^4 + \lambda^2 + \frac{27}{4}\mu(1 - \mu) = 0. \tag{16}$$

It has the roots

$$\lambda^2 = \tfrac{1}{2}[-1 \pm \sqrt{1 - 27\mu(1 - \mu)}]. \tag{17}$$

When the above square root is zero, we have the double eigenvalues $\pm i\sqrt{2}/2$. This occurs for $\mu = \mu_1 = \frac{1}{2}(1 - \sqrt{69}/9)$, that is, for Routh's critical mass ratio (and due to symmetry also for $1 - \mu_1$). It can be seen that the matrix (7) has nonsimple elementary divisors, which means it is not diagonalizable at μ_1. We will return to this case later on.

For $\mu_1 < \mu < 1 - \mu_1$, the square root in (17) produces imaginary values, and so λ will be complex with a nonzero real part. The eigenvalues of (7) lie off the imaginary axis, and the triangular Lagrangian points cannot be stable.

This leaves the interval $0 < \mu < \mu_1$ (and $1 - \mu_1 < \mu < 1$) where the matrix (8) has purely imaginary eigenvalues of the form $\pm i\omega_1$ and $\pm i\omega_2$. We will adopt the convention that ω_1 will be the larger of the two values so that ω_1 and ω_2 are uniquely defined by the conditions which follow from (17):

$$0 < \omega_2 < \sqrt{2}/2 < \omega_1, \tag{18a}$$

$$\omega_1^2 + \omega_2^2 = 1, \tag{18b}$$

$$\omega_1^2\omega_2^2 = 27\mu(1 - \mu)/4. \tag{18c}$$

■

We restrict now our attention to the case when the mass ratio μ is smaller than Routh's critical value μ_1. The quadratic part of the Hamiltonian function near $\mathscr{L}_4$ is

$$Q = \frac{1}{2}(v_1^2 + v_2^2) + u_2v_1 - u_1v_2 + \frac{1}{8}u_1^2 - \frac{3\sqrt{3}}{4}\mu(1 - \mu)u_1u_2 - \frac{5}{8}u_2^2. \tag{19}$$

We will construct the symplectic linear transformation which brings this Hamiltonian function into its normal form. In terms of complex coordinates, this normal form will turn out to be

$$K = -i\omega_1 z_1\bar{z}_1 + i\omega_2 z_2\bar{z}_2. \tag{20}$$

It is the Hamiltonian function for two harmonic oscillators with frequencies ω_1 and ω_2. Since the original Hamiltonian was indefinite, the two terms do not have the same sign.

When we perform these calculations, it is not so convenient to work with the parameter μ. It hides the symmetry of the problem with respect to $\mu = \frac{1}{2}$. The calculations are simpler if we use $1 - 2\mu$ as a parameter instead of μ. At the same time we can simplify the calculations further by absorbing the factor $3\sqrt{3}$ into this parameter. We, thus, use

$$\gamma = 3\sqrt{3}(1 - 2\mu). \tag{21}$$

The other difficulty in performing the calculations by hand and even more so by machine has to do with the fact that the expressions for ω_1 and ω_2 are rather lengthy and it is easier to express everything in terms of these variables instead of μ. But ω_1 and ω_2 are not independent as (18) shows. Indeed, in order to simplify intermediate results, one has to use the relationships (18b) and (18c). One strategy is to replace ω_2^2 by $1 - \omega_1^2$ whenever it occurs and,

thus, restrict the exponents of ω_2 to 0 and 1. But most expressions are shorter if the symmetry between the frequencies ω_1 and ω_2 is preserved within the formulas.

Our approach will reduce the problem so that it has the minimum number of essential parameters. We will divide the Hamiltonian function by ω_1, and we will set $\omega = \omega_2/\omega_1$. Due to our previous convention for ω_1 and ω_2 one sees that ω lies in $0 < \omega < 1$. We use (18b) to express terms containing ω_1 and ω_2 as a function of ω. The relationship (18c) then reads

$$\frac{16\omega^2}{(1+\omega^2)^2} = 27 - \gamma^2 \tag{22}$$

or

$$\gamma^2 = \frac{27 + 38\omega^2 + 27\omega^4}{(1+\omega^2)^2}. \tag{23}$$

The last form will be used to limit the exponent of γ to 0 and 1 in all intermediate expressions.

The Hamiltonian matrix derived from (7) is

$$A = \frac{1}{\omega_1}\begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ -1/4 & \gamma/4 & 0 & 1 \\ \gamma/4 & 5/4 & -1 & 0 \end{bmatrix}. \tag{24}$$

Its eigenvalues are $\pm i$ and $\pm i\omega$. The eigenvectors belonging to $+i$ and to $+i\omega$ are denoted by α_1 and α_2, respectively. They are given by

$$\alpha_1 = \begin{bmatrix} 1 \\ \dfrac{-(\omega^2+1)\gamma + 8i\sqrt{\omega^2+1}}{9\omega^2+13} \\ \dfrac{(\omega^2+1)\gamma + i(\omega^2+5)/\sqrt{\omega^2+1}}{9\omega^2+13} \\ \dfrac{9\omega^2+5 - i\gamma\sqrt{\omega^2+1}}{9\omega^2+13} \end{bmatrix}, \tag{25a}$$

$$\alpha_2 = \begin{bmatrix} 1 \\ \dfrac{-(\omega^2+1)\gamma + 8i\omega\sqrt{\omega^2+1}}{13\omega^2+9} \\ \dfrac{(\omega^2+1)\gamma + i\omega(5\omega^2+1)/\sqrt{\omega^2+1}}{13\omega^2+9} \\ \dfrac{5\omega^2+9 - i\gamma\omega\sqrt{\omega^2+1}}{13\omega^2+9} \end{bmatrix}. \tag{25b}$$

Since $\alpha_1^T J\bar{\alpha}_1 = -ir_1^2/2$ and $\alpha_2^T J\bar{\alpha}_2 = ir_2^2/2$, where r_1 and r_2 are the positive real roots of

$$r_1^2 = \frac{16(1-\omega^2)}{\sqrt{\omega^2+1}(9\omega^2+13)} \quad \text{and} \quad r_2^2 = \frac{16(1-\omega^2)}{\sqrt{\omega^2+1}(13\omega^2+9)}, \tag{26}$$

respectively, we create the transformation matrix T to the new set of complex valued variables $(z_1, z_2, \bar{z}_1, \bar{z}_2)$ by

$$T = (\bar{\alpha}_1/r_1, \alpha_2/r_2, \alpha_1/r_1, \bar{\alpha}_2/r_2). \tag{27}$$

Since we have $T^T JT = (1/-2i)J$, the transformation is symplectic with multiplier $-1/2i$. The old and new Hamiltonians are related by

$$K(z_1, z_2, \bar{z}_1, \bar{z}_2) = -2iQ(u_1, u_2, v_1, v_2), \tag{28}$$

and it leads to

$$K = -iz_1\bar{z}_1 + i\omega z_2\bar{z}_2. \tag{29}$$

We remark in passing that T is not the only symplectic matrix which accomplishes the transformation to this complex normal form. The matrix $(\bar{\alpha}_1/r_1^2, \alpha_2/r_2^2, \alpha_1, \bar{\alpha}_2)$ would do the same and at the same time has a simpler form than T. On the other hand, the reality conditions for it are more complicated. The advantage of a simpler form is lost when we want to go back to real coordinates.

Therefore, we stay with the above form for T and introduce a new set of real variables $(\xi_1, \xi_2, \eta_1, \eta_2)$ by $z_j = \xi_j + i\eta_j, j = 1, 2$. It is a symplectic transformation with multiplier $-2i$, and the transformed Hamiltonian becomes

$$\mathcal{K} = \frac{1}{2}(\xi_1^2 + \eta_1^2) - \frac{\omega}{2}(\xi_2^2 + \eta_2^2). \tag{30}$$

The transformation from the original coordinates to these new coordinates is then given by

$$\begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \frac{1}{2\sqrt{1-\omega^2}} RS \begin{bmatrix} \xi_1 \\ \xi_2 \\ \eta_1 \\ \eta_2 \end{bmatrix}, \tag{31}$$

where R is the matrix

$$R = \begin{bmatrix} 9\omega^2+13 & 13\omega^2+9 & 0 & 0 \\ -\gamma(\omega^2+1) & -\gamma(\omega^2+1) & 8(\omega^2+1) & -8(\omega^2+1) \\ \gamma(\omega^2+1) & \gamma(\omega^2+1) & \omega^2+5 & -5\omega^2-1 \\ 9\omega^2+5 & 5\omega^2+9 & -\gamma(\omega^2+1) & \gamma(\omega^2+1) \end{bmatrix} \tag{32}$$

and S the diagonal matrix

$$S = \operatorname{diag}\left(\frac{\sqrt[4]{\omega^2+1}}{\sqrt{9\omega^2+13}}, \frac{\sqrt[4]{\omega^2+1}}{\sqrt{\omega(13\omega^2+9)}}, \frac{1}{\sqrt[4]{\omega^2+1}\sqrt{9\omega^2+13}}, \frac{\sqrt{\omega}}{\sqrt[4]{\omega^2+1}\sqrt{13\omega^2+9}}\right). \tag{33}$$

The matrix A in (24) and the subsequent Hamiltonian $\mathcal{K}$ in (30) have been scaled. The true matrix of the restricted problem at $\mathcal{L}_4$ is $\omega_1 A$. The transformations given above will diagonalize $\omega_1 A$ also. In fact, K in (29) becomes $K = -i\omega_1 z_1 \bar{z}_1 + i\omega_2 z_2 \bar{z}_2$, and $\mathcal{K}$ in (30) becomes $\mathcal{K} = (\omega_1/2)(\xi_1^2 + \eta_1^2) - (\omega_2/2)(\xi_2^2 + \eta_2^2)$.

The above transformation becomes singular when $\omega = 1$. This is due to the fact that the Hamiltonian matrix (24) is not diagonalizable when $\gamma = \sqrt{23}$ or when $\mu = \mu_1$. We will construct the linear transformation which brings

$$A = \sqrt{2}\begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ -1/4 & \sqrt{23}/4 & 0 & 1 \\ \sqrt{23}/4 & 5/4 & -1 & 0 \end{bmatrix} \tag{34}$$

into its complex normal form

$$C = \begin{bmatrix} -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & -1 & i & 0 \\ -1 & 0 & 0 & -i \end{bmatrix}, \tag{35}$$

and afterwards we will convert it into the corresponding real normal form

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \end{bmatrix}. \tag{36}$$

For the eigenvalue $+i$ of the matrix A, we calculate the eigenvector α and the generalized eigenvector β. They are given by

$$\alpha = r\begin{bmatrix} 2\sqrt{23} + 8i\sqrt{2} \\ -10 \\ 2 + i\sqrt{46} \\ 2\sqrt{23} + 3i\sqrt{2} \end{bmatrix}, \qquad \beta = s\alpha + \frac{r}{5}\begin{bmatrix} -8\sqrt{2} - 8i\sqrt{23} \\ 0 \\ -\sqrt{46} - 48i \\ 17\sqrt{2} - 8i\sqrt{23} \end{bmatrix}, \tag{37}$$

where r and s are complex-valued constants which have to be determined such that the transformation is symplectic. Due to the form of C the transformation matrix T from real coordinates to the new complex coordinates has to be $T = (\bar{\beta}, \beta, \alpha, \bar{\alpha})$.

The only terms which are nonzero in $T^T J T$ are $\beta^T J \bar{\beta}$, $\beta^T J \bar{\alpha}$, and those directly related to them. We compute $\beta^T J \bar{\alpha} = (80\sqrt{2})\, r\bar{r}$ and $\beta^T J \bar{\beta} = i16\sqrt{2}[10\ \mathrm{Im}(r\bar{s}) - r\bar{r}]$.

In order to get a symplectic transformation to the new complex coordinates (z_1, z_2, z_3, z_4), we set $r = 1/\sqrt{80\sqrt{2}}$ and $s = -ir/10$. From the form of the matrix C, it also follows that the reality conditions have to be $z_1 = \bar{z}_2$ and $z_3 = \bar{z}_4$. It requires that the transformation to real position coordinates ξ_1, ξ_2 and their conjugate momenta η_1, η_2 has to be set up in the following special way:

$$\begin{aligned} z_1 &= \xi_1 + i\xi_2, \\ z_2 &= \xi_1 - i\xi_2, \\ z_3 &= \eta_1 - i\eta_2, \\ z_4 &= \eta_1 + i\eta_2. \end{aligned} \tag{38}$$

This form is forced upon us if we want to preserve the two-form, that is, $dz_1 \wedge dz_3 + dz_2 \wedge dz_4 = 2(d\xi_1 \wedge d\eta_1 + d\xi_1 \wedge d\eta_2)$.

Summarizing, we first transformed the original Hamiltonian function $\sqrt{2}H$ into the complex normal form

$$K = -iz_1 z_3 + iz_2 z_4 + z_1 z_2, \tag{39}$$

which we then transformed into the real normal form

$$\mathcal{K} = -\xi_1 \eta_2 + \xi_2 \eta_1 + \tfrac{1}{2}(\xi_1^2 + \xi_2^2). \tag{40}$$

The composite transformation from the original to the new real coordinates is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \frac{\sqrt{5\sqrt{2}}}{100} \begin{bmatrix} 4\sqrt{2} & 9\sqrt{23} & -10\sqrt{23} & -40\sqrt{2} \\ 0 & -5 & 50 & 0 \\ \sqrt{46}/2 & 49 & -10 & -5\sqrt{46} \\ -37\sqrt{2}/2 & 9\sqrt{23} & -10\sqrt{23} & -15\sqrt{2} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \eta_1 \\ \eta_2 \end{bmatrix}. \tag{41}$$

The transformations given above takes the matrix A in (34) to its normal form but $A = \sqrt{2}B$ where B is the true matrix at $\mathcal{L}_4$. Similarly the transformations takes $\sqrt{2}Q$ to normal form where Q is the true quadratic Hamiltonian at $\mathcal{L}_4$. The transformations take Q to $K = (\sqrt{2}/2)\{-iz_1 z_3 + iz_2 z_4 + z_1 z_2\}$ and $\mathcal{K} = (\sqrt{2}/2)\{-\xi_1 \eta_2 + \xi_2 \eta_1 + (\xi_1^2 + \xi_2^2)/2\}$. In order to get Q into its true normal form, one additional scaling, $\xi_i \to \sqrt[4]{2}\xi_i$, $\eta_i \to 1/\sqrt[4]{2}\eta_i$, is required. This scaling is symplectic, and the Q becomes $(\sqrt{2}/2)\{-\xi_1 \eta_2 + \xi_2 \eta_1\} + (\xi_1^2 + \xi_2^2)/2$.

H. Further Reading

Be careful in reading other texts because the terminology has only recently stabilized to what is found here. In some of the older texts "symplectic" becomes "canonical" and "Hamiltonian" becomes "infinitesimal canonical." Other variations are possible.

The most complete discussion of linear Hamiltonian systems is contained in Yakubovick and Stazhinskii (1975). This two-volume set gives a complete discussion of the theory of parametric stability and its applications. These books are well written, but a little wordy. A great deal of the basic information and special facts about symplectic matrices can be found with diligence in the almost unreadable text by Wintner (1944). Siegel and Moser (1971) contain a nice discussion of Hamiltonian and symplectic matrices with applications to the equilibrium points of the restricted problem. The material on linear systems is scattered in various chapters, but the treasure is worth the search.

The most complete treatment of the canonical forms for Hamiltonian matrices is found in Willianson (1936, 1938, 1939). These papers are a little hard to read, due to the old style of treating algebraic questions. The more geometric treatment given here is a part of the complete discussion given in Laub and Meyer (1974).

Appendix. Logarithm of a Symplectic Matrix

The simplest proof that a symplectic matrix has a Hamiltonian logarithm uses the theory of analytic functions of a matrix. Since this theory is not widely known and it would take us too far afield to develop it here, we will give the proof without the background development for those who know the theory. This proof and some of the background material is found in Yakubovich and Stazhinskii (1975). See Sibuya (1960) for a more algebraic, but not necessarily simpler, proof.

Lemma 1. *Let T be a real symplectic matrix which does not have -1 as an eigenvalue. Then there exists a real, Hamiltonian matrix B such that $T = \exp B$.*

Proof Outline

The logarithm of a nonsingular matrix, T, is given by the formula

$$\ln T = \frac{1}{2\pi i} \oint_{\Gamma} (\zeta I - T)^{-1} \ln \zeta \, d\zeta, \tag{1}$$

where Γ is a smooth, closed curve which contains in its interior all the eigenvalues of T but not the origin. There are many branches of the ln functions used in (1), and so there are many branches of the matrix logarithm also.

Let T be symplectic and have distinct eigenvalues $\lambda_1, \ldots, \lambda_{2k}$, with $\lambda_i \neq -1$ for all i. The set of eigenvalues of T are symmetric with respect to the real axis and the unit circle by Proposition C.1. Let $\Gamma_1, \ldots, \Gamma_{2k}$ be small, nonintersecting circles in the complex plane centered at $\lambda_1, \ldots, \lambda_{2k}$, respectively, which are symmetric with respect to the real axis and the unit circle. Thus, conjugation, $z \to \bar{z}$, and inversion, $z \to 1/z$, take the set of circles $\Gamma_1, \ldots, \Gamma_{2k}$ into itself (possibly permutating the order).

Let Ln be a branch of the logarithm function defined by slitting the complex plane alone the negative real axis and $-\pi < \arg(\text{Ln } z) < \pi$. Then a logarithm of T is given by

$$B = \ln T = \frac{1}{2\pi i} \sum_{j=1}^{2k} \oint_{\Gamma_j} (\zeta I - T)^{-1} \text{Ln } \zeta \, d\zeta. \tag{2}$$

Let conjugation take Γ_j to $-\Gamma_s = \bar{\Gamma}_j$ (the minus indicates that conjugation reverses orientation). Then

$$\begin{aligned} \overline{\frac{1}{2\pi i} \oint_{\Gamma_j} (\zeta I - T)^{-1} \text{Ln } \zeta \, d\zeta} &= -\frac{1}{2\pi i} \oint_{\bar{\Gamma}_j} (\bar{\zeta} I - T)^{-1} \text{Ln } \bar{\zeta} \, d\bar{\zeta} \\ &= -\frac{1}{2\pi i} \oint_{-\Gamma_s} (\zeta I - T)^{-1} \text{Ln } \zeta \, d\zeta \\ &= \frac{1}{2\pi i} \oint_{\Gamma_s} (\zeta I - T)^{-1} \text{Ln } \zeta \, d\zeta. \end{aligned} \tag{3}$$

So conjugation takes each term in (2) into another, which implies that B is real.

Now we claim that B is Hamiltonian. Let inversion take Γ_j into Γ_s (inversion is orientation preserving). Make the change of variables $\zeta = 1/\xi$ in the integrals in (2) and recall that $T^{-1} = -JT^TJ$. Then

$$\begin{aligned} (\zeta I - T)^{-1} \text{Ln } \zeta \, d\zeta &= [(1/\xi)I - T]^{-1}(-\text{Ln } \xi)(-d\xi/\xi^2) \\ &= (I - \xi T)^{-1} \xi^{-1} \text{Ln } \xi \, d\xi \\ &= \{T(I - \xi T)^{-1} + \xi^{-1} I\} \text{Ln } \xi \, d\xi \\ &= \{(T^{-1} - \xi I)^{-1} + \xi^{-1} I\} \text{Ln } \xi \, d\xi \\ &= \{(-JT^TJ - \xi I)^{-1} + \xi^{-1} I\} \text{Ln } \xi \, d\xi \\ &= -J(T^T - \xi I)^{-1} J \text{ Ln } \xi \, d\xi + \xi^{-1} \text{ Ln } \xi \, d\xi. \end{aligned} \tag{4}$$

Since the circle Γ_j does not enclose the origin, $\oint \xi^{-1} \text{ Ln } \xi \, d\xi = 0$ on Γ_j for all j. Making the substitution $\zeta = 1/\xi$ in (2) and using (4) shows that $B = JB^TJ$ or $JB^T + BJ = 0$. Thus, B is Hamiltonian. ∎

Theorem 2. *Let T be a real, symplectic matrix. T has a real, Hamiltonian logarithm if and only if it has a real, symplectic square root.*

PROOF. By the discussion in Section D, the symplectic space has a direct sum decomposition $\mathbb{V} = \mathbb{Y} \oplus \mathbb{L}$, where both $\mathbb{Y}$ and $\mathbb{L}$ are T-invariant, symplectic subspaces, and T restricted to $\mathbb{Y}$ has only the eigenvalue -1, whereas T restricted to $\mathbb{L}$ does not have the eigenvalue -1. By the lemma given above, T restricted to $\mathbb{L}$ has a real Hamiltonian logarithm; so, it must be shown that T restricted to $\mathbb{Y}$ has a real Hamiltonian logarithm.

Thus, we can assume that T has only the eigenvalue -1. If T has a real, symplectic square root, Q, so $T = Q^2$, then Q has eigenvalues $\pm i$. By Lemma 1, there is a real Hamiltonian matrix B' so that $Q = \exp B'$. But $\exp(2B') = \exp B' \exp B' = Q^2 = T$. Thus, $B = 2B'$ is the real, Hamiltonian logarithm of T. Conversely, if $T = \exp B$, where B is a real, Hamiltonian matrix, then $\exp(B/2)$ is a real, symplectic square root of T. ∎

Problems

1. Supply proofs to the Lemmas and Corollaries C.8 to C.13.
2. Prove that the two symplectic matrices in formula (12) in Lemma C.11 are not symplectically similar.
3. Consider the linear fractional (or Möbius transformation)

$$\Phi: z \to w = \frac{1+z}{1-z}, \qquad \Phi^{-1}: w \to z = \frac{w-1}{w+1}.$$

   a. Show that Φ maps the left half-plane into the interior of the unit circle. What are $\Phi(0)$, $\Phi(1)$, $\Phi(i)$, and $\Phi(\infty)$?
   b. Show that Φ maps the set of $m \times m$ matrices with no eigenvalue $+1$ bijectively onto the set of $m \times m$ matrices with no eigenvalue -1.
   c. Let $B = \Phi(A)$ where A and B are $2n \times 2n$. Show that B is symplectic if and only if A is Hamiltonian.
   d. Apply Φ to each of the canonical forms for Hamiltonian matrices to obtain canonical forms for symplectic matrices.
   e. Prove the analog of Theorem F.3 for symplectic matrices by using Theorem F.3 for Hamiltonian matrices and Φ.

4. Consider the system (*) $M\ddot{q} + Vq = 0$, where M and V are $n \times n$ symmetric matrices and M is positive definite. From matrix theory there is a nonsingular matrix P such that $P^T M P = I$ and an orthogonal matrix R such that $R^T(P^T V P)R = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Show that the above equation can be reduced to $\ddot{p} + \Lambda p = 0$. Discuss the stability and asymptotic behavior of these systems. Write (*) as a Hamiltonian system with Hamiltonian matrix H. Use the above results to obtain a symplectic matrix T such that

$$T^{-1}HT = \begin{pmatrix} 0 & I \\ -\Lambda & 0 \end{pmatrix}.$$

5. a. Let M and V be as in Problem 4. Show that if V has one negative eigenvalue, then some solutions of (*) in Problem 4 tend to infinity as $t \to \pm\infty$.

b. Consider the system (**) $M\ddot{q} + U(q) = 0$, where M is positive definite and $U\colon \mathbb{R}^n \to \mathbb{R}$ is smooth. Let q_0 be a critical point of U such that the Hessian of U at q_0 has one negative eigenvalue (so q_0 is not a local minimum of U). Show that q_0 is an unstable critical point for the system (**).

6. Let $H(t, z) = \frac{1}{2} z^T S(t) z$ and $\zeta(t)$ be a solution of the linear system with Hamiltonian H. Show that

$$\frac{d}{dt} H = \frac{\partial}{\partial t} H,$$

i.e.,

$$\frac{d}{dt} H(t, \zeta(t)) = \frac{\partial}{\partial t} H(t, \zeta(t)).$$

7. Let G be a set. A *product on* G is a function from $G \times G$ into G. A product is usually written using infix notation; so, if the product is denoted by $^\circ$, then one writes $a^\circ b$ instead of $^\circ(a, b)$. Addition and multiplication of real numbers define products on the reals, but the inner product of two vectors does not define a product since the inner product of two vectors is a scalar not a vector. A *group* is a set G with a product $^\circ$ on G which satisfies the following: (i) There is a unique element $e \in G$ such that $a^\circ e = e^\circ a = a$ for all $a \in G$; (ii) for every $a \in G$ there is a unique element $a^{-1} \in G$ such that $a^\circ a^{-1} = a^{-1\circ} a = e$; (iii) $(a^\circ b)^\circ c = a^\circ(b^\circ c)$ for all $a, b, c \in G$. e is called the identity, a^{-1} the inverse of a, and the last property is the associative law. Show that the following are groups.

a. $G = \mathbb{R}$, the reals, and $^\circ = +$, addition of real numbers (what is e? Answer: 0).

b. $G = \mathbb{C}$, the complex numbers, and $^\circ = +$, addition of complex numbers (what is 1^{-1}. Answer: -1).

c. $G = \mathbb{R}\backslash\{0\}$, the nonzero reals, and $^\circ = \cdot$, multiplication of reals.

d. $G = Gl(n, \mathbb{R})$, the set of all $n \times n$ real, nonsingular matrices, and $^\circ = \bullet$ matrix multiplication.

Show that the following are not groups.

e. $G = \mathbb{E}^3$, three dimensional geometric vectors, and $^\circ = \times$, the vector cross product.

f. $G = \mathbb{R}^+$, the positive reals, and $^\circ = +$, addition.

g. $G = \mathbb{R}$, and $^\circ = \bullet$, real multiplication.

8. A *subgroup* of a group G is a subset $H \subset G$, which is a group with the same product. A *matrix Lie group* is a closed subgroup of $Gl(m, \mathbb{F})$. Show that the following are matrix Lie groups.

a. $Gl(m, \mathbb{F})$ = *general linear group* = all $n \times n$ nonsingular matrices.

b. $Sl(m, \mathbb{F})$ = *special linear group* = set of all $A \in Gl(m, \mathbb{F})$ with $\det A = 1$.

c. $O(m, \mathbb{F})$ = *orthogonal group* = set of all $m \times m$ orthogonal matrices.

d. $So(m, \mathbb{F})$ = *special orthogonal group* = $O(m, \mathbb{F}) \cap Sl(m, \mathbb{F})$.

e. $Sp(n, \mathbb{F})$ = *symplectic group* = set of all $2n \times 2n$ symplectic matrices.

9. Show that the following are Lie subalgebras of $gl(m, \mathbb{F})$, see problem I.2.c.

a. $sl(m, \mathbb{F})$ = set of $m \times m$ matrices with trace = 0. (sl = special linear.)

b. $o(m, \mathbb{F})$ = set of $m \times m$ skew symmetric matrices. (o = orthogonal.)

c. $sp(n, \mathbb{F})$ = set of all $2n \times 2n$ Hamiltonian matrices.

10. Let $Q(n, \mathbb{F})$ be the set of all quadratic forms in $2n$ variables with coefficients in $\mathbb{F}$, so $q \in \mathcal{Q}(n, \mathbb{R})$, if $q(x) = \frac{1}{2}x^T S x$, where S is a $2n \times 2n$ symmetric matrix and $x \in \mathbb{F}^{2n}$.
a. Prove that $\mathcal{Q}(n, \mathbb{F})$ is a Lie algebra, where the product is the Poisson bracket.
b. Prove that $\Psi: \mathcal{Q}(n, \mathbb{F}) \to sp(n, \mathbb{F}) : q(x) = \frac{1}{2}x^T S x \to JS$ is a Lie algebra isomorphism.

11. Prove the theorem: $e^{At} \in \mathcal{G}$ for all t if and only if $A \in \mathcal{A}$ in the following cases:
a. When $\mathcal{G} = Gl(m, \mathbb{F})$ and $\mathcal{A} = gl(m, \mathbb{F})$.
b. When $\mathcal{G} = Sl(m, \mathbb{F})$ and $\mathcal{A} = sl(m, \mathbb{F})$.
c. When $\mathcal{G} = So(m, \mathbb{F})$ and $\mathcal{A} = so(m, \mathbb{F})$.
d. When $\mathcal{G} = Sp(n, \mathbb{F})$ and $\mathcal{A} = sp(n, \mathbb{F})$.

12. Show that the matrices

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -2 & 0 \\ 0 & -1/2 \end{pmatrix}$$

have no real logarithm.

13. Prove Lemma E.1 for the symplectic matrix T by using induction on the formula $\{\eta_k(\lambda), \eta_k(\mu)\} = 0$, where $\eta_k(\lambda) = \text{kernel}(T^k - \lambda I)$. (See Laub and Meyer, 1974.)

14. Write the fourth-order equation $x^{(4)} = 0$ as a Hamiltonian system. (Hint: see the canonical forms in D.3.)

15. Compute $\exp A$ for each canonical form given in D.3 and D.5.

16. Hill's lunar problem is defined by the Hamiltonian

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{1}{\|x\|} + (3x_1^2 - \|x\|^2),$$

where $x, y \in \mathbb{R}^2$. Show that it has two equilibrium points on the x_2 axis. Linearize the equations of motion about these equilibrium points and put the linearized equations in normal form by a real symplectic change of coordinates.

17. Consider a quadratic form $H = \frac{1}{2}x^T S x$, where $S = S^T$ is a real symmetric matrix. The *index* of the quadratic form H is the dimension of the largest linear space where H is negative. Show that the index of H is the same as the number of negative eigenvalues of S. Show that if S is nonsingular and H has odd index, then the linear Hamiltonian system $\dot{x} = JSx$ is unstable. (Hint: show that the determinant of JS is negative.)

CHAPTER III

Exterior Algebra and Differential Forms

Differential forms play an important part in the theory of Hamiltonian systems, but his theory is not universally known by scientists and mathematicians. It gives the natural higher-dimensional generalization of the results of classical vector calculus. We give a brief introduction with some, but not all, proofs and refer the reader to Flanders (1963) for another informal introduction but a more complete discussion with many applicatons, or to Spivak (1965) or Abraham and Marsden (1978) for more complete mathematical discussion. The reader conversant with the theory of differential forms can skip this chapter, and the reader not conversant with the theory should realize that what is presented here is not meant to be a complete development but simply an introduction to a few results that will be used sparingly later.

In this chapter we introduce and use the notation of classical differential geometry by using superscripts and subscripts to differentiate between a vector space and its dual. This convention helps sort out the multitude of different types of vectors encountered.

A. Exterior Algebra

Lev $\mathbb{V}$ be a vector space of dimension m over the real numbers $\mathbb{R}$. The best examples to keep in mind are the space of directed line segments in Euclidean 3-space, $\mathbb{E}^3$, or the space of all forces which can act at a point. Let $\mathbb{V}^k$ denote k copies of $\mathbb{V}$, i.e., $\mathbb{V}^k = \mathbb{V} \times \cdots \times \mathbb{V}$ (k times). A function $\phi\colon \mathbb{V}^k \to \mathbb{R}$ is called *k-multilinear* if it is linear in each argument; so,

$$\begin{aligned}\phi(a_1, \ldots, a_{r-1}, \alpha u + \beta v, a_{r+1}, \ldots, a_k) \\ = \alpha\phi(a_1, \ldots, a_{r-1}, u, a_{r+1}, \ldots, a_k) + \beta\phi(a_1, \ldots, a_{r-1}, v, a_{r+1}, \ldots, a_k)\end{aligned} \tag{1}$$

for all $a_1, \ldots, a_k$, $u, v \in \mathbb{V}$, all $\alpha, \beta \in \mathbb{R}$, and all arguments, $r = 1, \ldots, k$. A 1-multilinear map is a linear functional which we sometimes call a *covector* or 1-*form*. In $\mathbb{R}^m$ the scalar product $(a, b) = a^T b$ is 2-multilinear, in $\mathbb{R}^{2n}$ the symplectic product $\{a, b\} = a^T J b$ is 2-multilinear, and in $\mathbb{R}^m$ the determinant is m-multilinear in its m rows (or columns). A k-multilinear function ϕ is *skew symmetric* or *alternating* if interchanging any two arguments changes its sign; so,

$$\phi(a_1, \ldots, a_r, \ldots, a_s, \ldots, a_k) = -\phi(a_1, \ldots, a_s, \ldots, a_r, \ldots, a_k) \tag{2}$$

for all $a_1, \ldots, a_k \in \mathbb{V}$ and all $r, s = 1, \ldots, k, r \neq s$. Clearly, ϕ is zero if two of its arguments are the same. We shall call an alternating k-multilinear function a *k-linear form* or *k-form* for short. In $\mathbb{R}^{2n}$ the symplectic product, $\{a, b\} = a^T J b$, and in $\mathbb{R}^m$ the determinant are alternating. Let $\mathbb{A}^0 = \mathbb{R}$ and $\mathbb{A}^k = \mathbb{A}^k(\mathbb{V})$ be the space of all k-forms for $k \geq 1$. It is easy to verify that $\mathbb{A}^k$ is a vector space when using the usual definition of addition of functions and multiplication of functions by a scalar.

In $\mathbb{E}^3$, as we have seen, a linear functional (a 1-form or an alternating 1-multilinear function) acting on a vector v can be thought of as the scalar project of v in a particular direction. A physical example is work. The work done by a uniform force is a linear functional on the displacement vector of a particle. (See Figure A.1).

Physical examples which are 2-forms are torque, angular momentum and magnetic field. Given two vectors in $\mathbb{E}^3$, they determine a plane through the origin and a parallelogram in that plane. The oriented area of this parallelogram is a 2-form. Two vectors in $\mathbb{E}^3$ determine (i) a plane, (ii) an orientation in the plane, and (iii) a magnitude—the area of the parallelogram. Physical quantities that also determine a plane, an orientation, and a magnitude are torque, angular momentum, and magnetic field.

Three vectors in $\mathbb{E}^3$ determine a parallelepiped, and its oriented volume is a 3-form. The flux of a uniform vector field, v, crossing a parallelogram determined by two vectors a and b is a 3-form.

If ψ is a 2-multilinear function, then ϕ defined by $\phi(a, b) = \{\psi(a, b) - \psi(b, a)\}/2$ is alternating and is sometimes called the alternating part of ψ. If ψ is already alternating, then $\phi = \psi$. If α and β are 1-forms, then $\phi(a, b) = \alpha(a)\beta(b) - \alpha(b)\beta(a)$ is a 2-form. This construction can be generalized. Let P_k be the set of all permutations of the k numbers $1, 2, \ldots, k$ and sign: $P_k \to \{+1, -1\}$ the function that assigns $+1$ to an even permutation

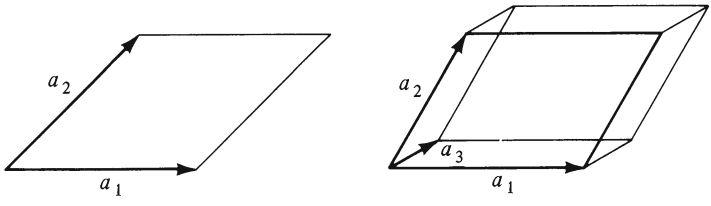

Figure A.1. Multilinear functions.

and -1 to an odd permutation. So if ϕ is alternating, $\phi(a_{\sigma(1)}, \ldots, a_{\sigma(k)}) = \operatorname{sign}(\sigma)\phi(a_1, \ldots, a_k)$. If ψ is a k-mutilinear function, then ϕ defined by

$$\phi(a_1, \ldots, a_k) = \frac{1}{k!} \sum_{\sigma \in P_k} \operatorname{sign}(\sigma)\psi(a_{\sigma(1)}, \ldots, a_{\sigma(k)}) \tag{3}$$

is alternating. We write $\phi = \operatorname{Alt}(\psi)$. If ψ is already alternating, then $\psi = \operatorname{alt}(\psi)$. If $\alpha \in \mathbb{V}^k$ and $\beta \in \mathbb{V}^r$, then define $\alpha \wedge \beta \in \mathbb{V}^{k+r}$ by

$$\alpha \wedge \beta = \frac{(k+r)!}{k!r!} \operatorname{alt}(\alpha\beta)$$

or

$$\alpha \wedge \beta(a_1, \ldots, a_{k+r}) = \sum_{\sigma \in P} \operatorname{sign}(\sigma)\alpha(a_{\sigma(1)}, \ldots, a_{\sigma(k)})\beta(a_{\sigma(k+1)}, \ldots, a_{\sigma(k+r)}). \tag{4}$$

The operator $\wedge : \mathbb{A}^k \times \mathbb{A}^r \to \mathbb{A}^{k+r}$ is called the *exterior product* or *wedge product.*

Lemma 1. *For all k-forms α, r-forms β, and s-forms γ:*

(i) $\alpha \wedge (\beta + \gamma) = \alpha \wedge \beta + \alpha \wedge \gamma$,
(ii) $\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma$,
(iii) $\alpha \wedge \beta = (-1)^{kr}\beta \wedge \alpha$.

PROOF. Parts (i) and (ii) are fairly easy and are left as exercises. Let τ be the permutation $\tau: (1, \ldots, k, k+1, \ldots, k+r) \to (k+1, \ldots, k+r, 1, \ldots, k)$, i.e., τ interchanges the first k entries and the last r entries. By thinking of τ as being the sequence $(1, \ldots, k, k+1, \ldots, k+r) \to (k+1, 1, \ldots, k, k+2, \ldots, k+r) \to (k+1, k+2, 1, \ldots, k+3, \ldots, k+r) \to \cdots \to (k+1, \ldots, k+r, 1, \ldots, k)$, it is easy to see that $\operatorname{sign}(\tau) = (-1)^{rk}$. Now

$$\begin{aligned}
\alpha \wedge \beta(a_1, \ldots, a_{k+r}) &= \sum_{\sigma \in P} \operatorname{sign}(\sigma)\alpha(a_{\sigma(1)}, \ldots, a_{\sigma(k)})\beta(a_{\sigma(k+1)}, \ldots, a_{\sigma(k+r)}) \\
&= \sum_{\sigma \in P} \operatorname{sign}(\sigma \circ \tau)\alpha(a_{\sigma\circ\tau(1)}, \ldots, a_{\sigma\circ\tau(k)})\beta(a_{\sigma\circ\tau(k+1)}, \ldots, a_{\sigma\circ\tau(k+r)}) \\
&= \sum_{\sigma \in P} \operatorname{sign}(\sigma)\operatorname{sign}(\tau)\beta(a_{\sigma(1)}, \ldots, a_{\sigma(kr)})\alpha(a_{\sigma(k+1)}, \ldots, a_{\sigma(k+r)}) \\
&= (-1)^{rk}\beta \wedge \alpha. \qquad \blacksquare
\end{aligned}$$

Let $e_1, \ldots, e_m$ be a basis for $\mathbb{V}$ and $f^1, \ldots, f^m$ be the dual basis for the dual space $\mathbb{V}^*$; so, $f^i(e_j) = \delta^i_j = 0$ if $i \neq j$, and $f^i(e_j) = \delta^i_j = 1$ if $i = j$. This is our first introduction to the subscript-superscript convention of differential geometry and classical tensor analysis.

Lemma 2. $\dim \mathbb{A}^k = \binom{m}{k}$. *In particular a basis for $\mathbb{A}^k$ is*

$$\{f^{i_1} \wedge f^{i_2} \wedge \cdots \wedge f^{i_k} : 1 \leq i_1 < i_2 < \cdots < i_k \leq m\}.$$

PROOF. Let I denote the set $\{(i_1, \ldots, i_k)$: i_j's are positive integers with $1 \leq i_1 < \cdots < i_k \leq m\}$ and $f^i = f^{i_1} \wedge \cdots \wedge f^{i_k}$ when $i \in I$. From the definition, $f^{i_2} \wedge f^{i_2} \wedge \cdots \wedge f^{i_k}(e_{j_1}, \ldots, e_{j_k})$ equals 1 if $i, j \in I$ and $i = j$ and equals 0 otherwise—in short, $f^i(e_j) = \delta^i_j$.

Let ϕ be a k-form and define

$$\psi = \sum_{i \in I} \phi(e_{i_1}, \ldots, e_{i_k}) f^{i_1} \wedge f^{i_2} \wedge \cdots \wedge f^{i_k} = \sum_{i \in I} \phi(e_i) f^i. \tag{5}$$

Let $v_i = \sum a^j_i e_j$, $i = 1, \ldots, k$, be k arbitrary vectors. By the multilinearity of ϕ and ψ, one sees that $\phi(v_1, \ldots, v_k) = \psi(v_1, \ldots, v_k)$; so, they agree on all vectors and, therefore, are equal. Thus, the set $\{f^i: i \in I\}$ spans.

Assume that

$$\sum_{i \in I} a_{i_1 \cdots i_k} f^{i_1} \wedge f^{i_2} \wedge \cdots \wedge f^{i_k} = 0. \tag{6}$$

For a fixed set of indices $s_1, \ldots, s_k$, let $r_{k+1}, \ldots, r_m$ be a complementary set, i.e., $s_1, \ldots, s_k, r_{k+1}, \ldots, r_m$ is just a permutation of the integers $1, \ldots, m$. Take the wedge product of (6) with $f^{r_{k+1}} \wedge \cdots \wedge f^{r_m}$ to get

$$\sum_{i \in I} a_{i_1 \cdots i_k} f^{i_1} \wedge f^{i_2} \wedge \cdots \wedge f^{i_k} \wedge f^{r_{k+1}} \wedge \cdots \wedge f^{r_m} = 0. \tag{7}$$

The only term in the above sum without a repeated f in the wedge is the one with $i_1 = s_1, \ldots, i_k = s_k$, and so it is the only nonzero term. Since $s_1, \ldots, s_k$, $r_{k+1}, \ldots, r_m$ is just a permutation of the integers $1, \ldots, m$, $f^{s_1} \wedge f^{s_2} \wedge \cdots \wedge f^{s_k} \wedge f^{r_{k+1}} \wedge \cdots \wedge f^{r_m} = \pm f^1 \wedge \cdots \wedge f^m$. Thus, applying the sum in (6) to $e_1, \ldots, e_m$ gives $\pm a_{s_1 \ldots s_k} = 0$. Thus, the f^i, $i \in I$, are independent. ∎

In particular, the dimension of $\mathbb{V}^m$ is 1, and the space has as a basis the single element $f^1 \wedge \cdots \wedge f^m$.

Lemma 3. *Let $g^1, \ldots, g^r \in \mathbb{V}^*$. Then $g^1, \ldots, g^r$ are linearly independent if and only if $g^1 \wedge \cdots \wedge g^r \neq 0$.*

PROOF. If the g's are dependent, then one of them is a linear combination of the others say, $g^r = \sum_{s=1}^{r-1} \alpha_s g^s$. Then $g^1 \wedge \cdots \wedge g^r = \sum_{s=1}^{r-1} \alpha_s g^1 \wedge \cdots \wedge g^{r-1} \wedge g^s$. Each term in this last sum is a wedge product with a repeated entry, and so by the alternating property, each term is zero. Therefore, $g^1 \wedge \cdots \wedge g^r = 0$. Conversely, if $g^1, \ldots, g^r$ are linearly independent, then extend them to a basis $g^1, \ldots, g^r, \ldots, g^m$. By Lemma 2, $g^1 \wedge \cdots \wedge g^r \wedge \cdots \wedge g^m \neq 0$, so $g^1 \wedge \cdots \wedge g^r \neq 0$. ∎

A linear map $L: \mathbb{V} \to \mathbb{V}$ induces a linear map $L_k: \mathbb{A}^k \to \mathbb{A}^k$ by the formula $L_k \phi(a_1, \ldots, a_k) = \phi(La_1, \ldots, La_k)$. If M is another linear map of $\mathbb{V}$ onto itself, then $(LM)_k = M_k L_k$ because $(LM)_k \phi(a_1, \ldots, a_k) = \phi(LMa_1, \ldots, LMa_k) = L_k \phi(Ma_1, \ldots, Ma_k) = M_k L_k \phi(a_1, \ldots, a_k)$. Recall that $\mathbb{A}^1 = \mathbb{V}^*$ is the dual space, and $L_1 = L^*$ is called the dual map.

If $\mathbb{V} = \mathbb{R}^m$ (column vectors), then we can identify the dual space $\mathbb{V}^* = \mathbb{A}^1$ with $\mathbb{R}^m$ by the convention $f \Leftrightarrow \not{f}$, where $f \in \mathbb{V}^*, \not{f} \in \mathbb{R}^m$, and $f(x) = \not{f}^T x$. In this

case, L is an $m \times m$ matrix, and Lx is the usual matrix product. $L_1 f$ is defined by $L_1 f(x) = f(Lx) = f^T Lx = (L^T f)^T x$; so, the matrix representation of L_1 is the transpose of L, i.e., $L_1(f) = L^T f$. The matrix representation of L_k is discussed in Flanders (1963).

By Lemma 1, dim $\mathbb{A}^m = 1$, and so every element in $\mathbb{A}^m$ is a scalar multiple of a single element. L_m is a linear map; so, there is a constant ℓ such that $L_m f = \ell f$ for all $f \in \mathbb{A}^m$. Define the determinant of L to be this constant ℓ, and denote it by $\det(L)$; so, $L_m f = (\det(L))f$.

Lemma 4. *Let L and M: $\mathbb{V} \to \mathbb{V}$ be linear. Then* (i) $\det(LM) = \det(L)\det(M)$, (ii) $\det(I) = 1$, *where I: $\mathbb{V} \to \mathbb{V}$ is the identity map,* (iii) *L is invertible if and only if* $\det(L) \neq 0$, *and, in this case,* $\det(L^{-1}) = \det(L)^{-1}$.

PROOF. Part (i) follows from $(LM)_m = M_m L_m$ which was established above. (ii) follows from the definition. Let L be invertible; so, $LL^{-1} = I$, and by (i) and (ii), $\det(L)\det(L^{-1}) = 1$; so, $\det(L) \neq 0$ and $\det(L^{-1}) = 1/\det(L)$. Conversely assume L is not invertible so there is an $e \in \mathbb{V}$ with $e \neq 0$ and $Le = 0$. Extend e to a basis, $e_1 = e, e_2, \ldots, e_m$. Then for any m-form ϕ, $L_m\phi(e_1, \ldots, e_m) = \phi(Le_1, \ldots, Le_m) = \phi(0, \ldots, Le_m) = 0$. So $\det(L) = 0$. ∎

Let $\mathbb{V} = \mathbb{R}^m$, $e_1, e_2, \ldots, e_m$ be the standard basis of $\mathbb{R}^m$, and let L be the matrix $L = (L_i^j)$; so, $Le_i = \sum_j L_i^j e_j$. Let ϕ be a nonzero element of $\mathbb{A}^m$.

$$\begin{aligned}
[\det(L)]\phi(e_1, \ldots, e_m) &= L_m\phi(e_1, \ldots, e_m) \\
&= \phi(Le_1, \ldots, Le_m) \\
&= \sum_{j_1} \cdots \sum_{j_m} \phi(L_1^{j_1} e_{j_1}, \ldots, L_m^{j_m} e_{j_m}) \\
&= \sum_{j_1} \cdots \sum_{j_m} L_1^{j_1} \cdots L_m^{j_m} \phi(e_{j_1}, \ldots, e_{j_m}) \\
&= \sum_{\sigma \in P} L_1^{\sigma(1)} \cdots L_m^{\sigma(m)} \operatorname{sign}(\sigma)\phi(e_1, \ldots, e_m).
\end{aligned}$$

In the second to last sum above the only nonzero terms are the ones with distinct e's. Thus, the sum over the nonzero terms is the sum over all permutations of the e's. From the above,

$$\det(L) = \sum_{\sigma \in P} L_1^{\sigma(1)} \cdots L_m^{\sigma(m)} \operatorname{sign}(\sigma), \tag{8}$$

which is the classical formula for the determinant of a matrix.

B. The Symplectic Form

In this section, let $(\mathbb{V}, \omega)$ be a symplectic space of dimension $2n$. Recall that in Chapter 2 a symplectic form, ω (on a vector space $\mathbb{V}$), was defined to be a nondegenerate, alternating bilinear form on $\mathbb{V}$, and the pair $(\mathbb{V}, \omega)$ was called a symplectic space.

Theorem 1. *There exists a basis $f^1, \ldots, f^{2n}$ for $\mathbb{V}^*$ such that*

$$\omega = \sum_{i=1}^{n} f^i \wedge f^{n+i}. \tag{1}$$

PROOF. By Theorem II.B.1 and its corollary, there is a symplectic basis $e_1, \ldots, e_{2n}$ so that the matrix of the form ω is the standard $J = (J_{ij})$ or $J_{ij} = \omega(e_i, e_j)$. Let $f^1, \ldots, f^{2n} \in \mathbb{V}^*$ be the basis dual to the symplectic basis $e_1, \ldots, e_{2n}$. The 2-form given on the right in (1) above agrees with ω on the basis $e_1, \ldots, e_{2n}$. ∎

The basis $f^1, \ldots, f^{2n}$ is a symplectic basis for the dual space $\mathbb{V}^*$. By the above, $\omega^n = \omega \wedge \omega \wedge \cdots \wedge \omega$ (n times) $= \pm n! f^1 \wedge f^2 \wedge \cdots \wedge f^{2n}$, where the sign is plus if n is even and minus if n is odd. Thus, ω^n is a nonzero element of $\mathbb{A}^{2n}$. Since a symplectic linear transformation preserves ω, it preserves ω^n, and, therefore, its determinant is $+1$.

Corollary 2. *The determinant of a symplectic linear transformation (or matrix) is $+1$.*

Actually, using the above arguments and the full statement of Theorem II.B.1, we can prove that a 2-form v on a linear space of dimension $2n$ is nondegenerate if and only if v^n is nonzero.

C. Tangent Vectors and Cotangent Vectors

Let $\mathbb{O}$ be an open set in an m-dimensional vector space $\mathbb{V}$ over $\mathbb{R}$, $e_1, \ldots, e_m$ a basis for $\mathbb{V}$, and $f^1, \ldots, f^m$ the dual basis. Let $x = (x^1, \ldots, x^m)$ be coordinates in $\mathbb{V}$ relative to $e_1, \ldots, e_m$ and also coordinates in V^* relative to the dual basis. Let $I = (-1, 1) \subset \mathbb{R}^1$, and let t be a coordinate in $\mathbb{R}^1$. Think of $\mathbb{V}$ as $\mathbb{R}^m$—we use the more general notation because it is helpful to keep a space and its dual distinct. $\mathbb{R}^m$ and its dual are often identified with each other, which can lead to confusion.

Much of analysis reduces to studying maps from an interval in $\mathbb{R}^1$ into $\mathbb{O}$ (curves, solutions of differential equations, etc.) and the study of maps from $\mathbb{O}$ into $\mathbb{R}^1$ (differentials of functions, potentials, etc.). The linear analysis of these two types of maps is, therefore, fundamental. The linearization of a curve at a point gives rise to a tangent vector, and the linearization of a function at a point gives rise to a cotangent vector. These are the concepts of this section.

A tangent vector to $\mathbb{O}$ at p is to be thought of as the tangent vector to a curve through p. Let $g, g' \colon I \to \mathbb{O} \subset \mathbb{V}$ be smooth curves with $g(0) = g'(0) = p$. We say g and g' are *equivalent at p* if $Dg(0) = Dg'(0)$. Since $Dg(0) \in \mathscr{L}(\mathbb{R}, \mathbb{V})$, we can identify $\mathscr{L}(\mathbb{R}, \mathbb{V})$ with $\mathbb{V}$ by letting $Dg(0)(1) = dg(0)/dt \in \mathbb{V}$. Being equivalent at p is an equivalence relation on curves, and an equivalence class (a maximal set of curves equivalent to each other) is defined to be a *tangent*

vector or a *vector to* $\mathbb{O}$ *at* p. That is, a tangent vector, $\{g\}$, is the set of all curves equivalent to g at p, i.e., $\{g\} = \{g' : I \to \mathbb{O} : g'(0) = p \text{ and } dg(0)/dt = dg'(0)/dt\}$. In the x coordinates, the derivative is $dg(0)/dt = (dg^1(0)/dt, \ldots, dg^m(0)/dt) = (\gamma^1, \ldots, \gamma^m)$; so, $(\gamma^1, \ldots, \gamma^m)$ are coordinates for the tangent vector $\{g\}$ relative to the x coordinates. The set of all tangent vectors to $\mathbb{O}$ at p is called the *tangent space to* $\mathbb{O}$ *at* p and is denoted by $T_p\mathbb{O}$. This space can be made into a vector space by using the coordinate representation given above. The curve $\xi_i : t \to p + te_i$ has $d\xi_i(0)/dt = e_i$ which is $(0, \ldots, 0, 1, 0, \ldots, 0)$—1 in the ith position—in the x coordinates. The tangent vector consisting of all curves equivalent to ξ_i at p is denoted by $\partial/\partial x^i = \partial/\partial x^i$. $\partial/\partial x^1, \ldots, \partial/\partial x^m$ form a basis for $T_p\mathbb{O}$. A typical vector $v_p \in T_p\mathbb{O}$ can be written $v_p = \gamma^1(\partial/\partial x^1) + \cdots + \gamma^m(\partial/\partial x^m)$. In classical tensor notation, one writes $v_p = \gamma^i(\partial/\partial x^i)$; it was understood that a repeated index, one as a superscript and one as a subscript, was to be summed over from 1 to m. This was called the *Einstein* or *summation convention.*

A cotangent vector (or covector for short) to $\mathbb{O}$ at p is to be thought of as the differential of a function at p. Let $h, h' : \mathbb{O} \to \mathbb{R}^1$ be two smooth functions. We say h and h' are *equivalent at* p if $Dh(p) = Dh'(p)$. [$Dh(p)$ is the same as the differential $dh(p)$.] This is an equivalence relation. A *cotangent vector* or a *covector to* $\mathbb{O}$ *at* p is by definition an equivalence class of functions. That is, a covector $\{h\}$ is the set of functions equivalent to h at p, i.e., $\{h\} = \{h' : \mathbb{O} \to \mathbb{R}^1 : Dh'(p) = Dh(p)\}$. In the x coordinate, $Dh(p) = (\partial h(p)/\partial x^1, \ldots, \partial h(p)/\partial x^m) = (\eta_1, \ldots, \eta_m)$; so, $(\eta_1, \ldots, \eta_m)$ are coordinates for the covector $\{h\}$. The set of all covectors at p is called the *cotangent space to* $\mathbb{O}$ *at* p and is denoted by $T_p^*\mathbb{O}$. This space can be made into a vector space by using the coordinate representation given above. The function $x^i : \mathbb{O} \to \mathbb{R}^1$ has a cotangent vector at p, which is $(0, \ldots, 1, \ldots, 0)$—1 in the ith position—in the x coordinates. The covector consisting of all functions equivalent to x^i at p is denoted by dx^i. $dx^1, \ldots, dx^m$ forms a basis for $T_p^*\mathbb{O}$. A typical covector $v^p \in T_p^*\mathbb{O}$ can be written $\eta_1\, dx^1 + \cdots + \eta_m\, dx^m$ or $\eta_i\, dx^i$ using the Einstein convection.

In the above two paragraphs there is clearly a parallel construction being carried out. In fact, they are dual constructions. Let g and h be as above; so, $h \circ g : I \subset \mathbb{R}^1 \to \mathbb{R}^1$. By the chain rule, $D(h \circ g)(0)(1) = Dh(p) \circ Dg(0)(1)$ which is a real number; so $Dh(p)$ is a linear functional on tangents to curves. In coordinates, if

$$\{g\} = v_p = \frac{dg^1}{dt}(0)\frac{\partial}{\partial x^1} + \cdots + \frac{dg^m}{dt}(0)\frac{\partial}{\partial x^m} = \gamma^1\frac{\partial}{\partial x^1} + \cdots + \gamma^m\frac{\partial}{\partial x^m}$$

and

$$\{h\} = v^p = \frac{\partial h}{\partial x^1}(p)\, dx^1 + \cdots + \frac{\partial h}{\partial x^m}(p)\, dx^m = \eta_1\, dx^1 + \cdots + \eta_m\, dx^m,$$

then

$$v^p(v_p) = D(h \circ g)(0) = \frac{dg^1}{dt}(0)\frac{\partial h}{\partial x^1}(p) + \cdots + \frac{dg^m}{dt}(0)\frac{\partial h}{\partial x^m}(p)$$

$$= \gamma^1\eta_1 + \cdots + \gamma^m\eta_m \quad (= \gamma^i\eta_i \text{ in Einstein notation}).$$

Thus, $T_p\mathbb{O}$ and $T_p^*\mathbb{O}$ are dual spaces.

At several points in the above discussion the corrdinates $x^1, \ldots, x^m$ were used. The natural question to ask is to what extent do these definitions depend of the choice of coordinates. Let $y^1, \ldots, y^m$ be another coordinate system which may not be linearly related to the x's. Assume that we can change coordinates by $y = \phi(x)$ and back by $x = \psi(y)$, where ϕ and ψ are smooth functions with nonvanishing Jacobians, $D\phi$ and $D\psi$. In classical notation, one writes $x^i = x^i(y)$, $y^j = y^j(x)$, and $D\phi = \{\partial y^j/\partial x^i\}$, $D\psi = \{\partial x^i/\partial y^j\}$.

Let $g\colon \mathbb{I} \to \mathbb{O}$ be a curve. In x coordinates let $g(t) = (a^1(t), \ldots, a^m(t))$ and in y coordinates let $g(t) = (b^1(t), \ldots, b^m(t))$. The x coordinate for the tangent vector $v_p = \{g\}$ is $\mathbf{a} = (da^1(0)/dt, \ldots, da^m(0)/dt) = (\alpha^1, \ldots, \alpha^m)$, and the y coordinate for $v_p = \{g\}$ is $\mathbf{b} = (db^1(0)/dt, \ldots, db^m(0)/dt) = (\beta^1, \ldots, \beta^m)$. Recall that we write vectors in the text as row vectors, but they are to be considered as column vectors. Thus $\mathbf{a}$ and $\mathbf{b}$ are column vectors. By the change of variables, $a(t) = \psi(b(t))$; so, differentiating gives $\mathbf{a} = D\psi(p)\mathbf{b}$. In classical notation, $a^i(t) = x^i(b(t))$; so, $da^i/dt = \sum_i (\partial x^i/\partial y^j)(db^j/dt)$ or

$$\alpha^i = \sum_{j=1}^{m} \frac{\partial x^i}{\partial y^j}\beta^j \quad \left(= \frac{\partial x^i}{\partial y^j}\beta^j \text{ in Einstein notation}\right). \tag{1}$$

This formula tells how the coordinates of a tangent vector are transformed. In classical tensor jargon, this is the transformation rule for a *contravariant vector*.

Let $h\colon \mathbb{O} \to \mathbb{R}^1$ be a smooth function. Let h be $a(x)$ in x coordinates and $b(y)$ in y coordinates. The cotangent vector $v^p = \{h\}$ in x coordinates is $\mathbf{a} = (\partial a(p)/\partial x^1, \ldots, \partial a(p)/\partial x^m) = (\alpha_1, \ldots, \alpha_m)$ and in y coordinates it is $\mathbf{b} = (\partial b(p)/\partial x^1, \ldots, \partial b(p)/\partial x^m) = (\beta_1, \ldots, \beta_m)$. By the change of variables $a(x) = b(\phi(x))$; so, differentiating gives $\mathbf{a} = D\phi(p)^T\mathbf{b}$. In classical notation $a(x) = b(y(x))$; so, $\alpha_i = \partial a/\partial x^i = \sum_j (\partial b/\partial y^j)(\partial y^j/\partial x^i) = \sum_j \beta_j(\partial y^j/\partial x^i)$ or

$$\alpha_i = \sum_{j=1}^{m} \frac{\partial y^j}{\partial x^i}\beta_j \quad \left(= \frac{\partial y^j}{\partial x^i}\beta_j \text{ in Einstein notation}\right). \tag{2}$$

This formula shows how the coordinates of a cotangent vector are transformed. In classical tensor jargon this is the transformation rule for a *covariant vector*.

D. Vector Fields and Differential Forms

Continue the notation of the Section C. A *tangent* (*cotangent*) *vector field on* $\mathbb{O}$ is a smooth choice of a tangent (cotangent) vector at each point of $\mathbb{O}$. That is, in coordinates, a tangent vector field, V, can be written in the form

$$V = V(x) = \sum_{i=1}^{m} v^i(x) \frac{\partial}{\partial x^i} \quad \left(= v^i(x) \frac{\partial}{\partial x^i} \right), \tag{1}$$

where the v^i: $\mathbb{O} \to \mathbb{R}^i$, $i = 1, \ldots, m$, are smooth functions, and a cotangent vector field, U, can be written in the form

$$U = U(x) = \sum_{i=1}^{m} u_i(x)\, dx^i \quad (= u_i(x)\, dx^i), \tag{2}$$

where u_i: $\mathbb{O} \to \mathbb{R}^1$, $i = 1, \ldots, m$, are smooth functions.

A tangent vector field V gives a tangent vector $V(p) \in T_p\mathbb{O}$ which was defined as the tangent vector of some curve. A different curve might be used for each point of $\mathbb{O}$; so, a natural question to ask is, Does there exist a curve g: $\mathbb{I} \subset \mathbb{R} \to \mathbb{O}$ such that $dg(t)/dt = V(g(t))$? In coordinates this is $dg^i(t) = v^i(g(t))$. This is the same as asking for a solution of the differential equation $\dot{x} = V(x)$. Thus, a tangent vector field is an ordinary differential equation. In classical tensor jargon it is also called a *contravariant vector field.*

A cotangent vector field U gives a cotangent vector $U(p) \in T_p^*\mathbb{O}$ which was defined as the differential of a function at p. A different function might be used for each point of $\mathbb{O}$; so, a natural question to ask is, Does there exist a function h: $\mathbb{O} \to \mathbb{R}^1$ such that $dh(x) = U(x)$? The answer to this question is no, in general. Certain integrability conditions discussed below must be satisfied before a cotangent vector field is a differential of a function. If this cotangent vector field is a field of work elements, i.e., a field of forces, then if $dh = -U$, the U would be a potential and the field would be conservative. But, as we shall see, not all forces are conservative.

Let $p \in \mathbb{O}$, and denote by $\mathbb{A}_p^k\mathbb{O}$ the space of k-forms on the tangent space $T_p\mathbb{O}$. A *k-differential form* or *k-form on* $\mathbb{O}$ is a smooth choice of a k-linear form in $\mathbb{A}_p^k\mathbb{O}$ for all $p \in \mathbb{O}$. That is, a k-form, F, can be written

$$F = \sum_{1 \le i_1 < \cdots < i_k \le m} f_{i_1 \ldots i_k}(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \sum_{i \in I} f_i(x)\, dx^i, \tag{3}$$

where the functions $f_{i_1 \ldots i_k}$: $\mathbb{O} \to \mathbb{R}$ are smooth. In the last expression in (3), I denotes the set $\{(i_1, \ldots, i_k)$: i_j's are positive integers with $1 \le i_1 < \cdots < i_k \le m\}$, and $dx^i = dx^{i_1} \wedge \cdots \wedge dx^{i_k}$. Since $\mathbb{A}_p^0\mathbb{O} = \mathbb{R}$, 0-forms are simply smooth functions, and since $\mathbb{A}_p^1\mathbb{O} = T_p^*\mathbb{O}$, 1-forms are covector fields.

In classical analysis, everything was a vector. In $\mathbb{R}^3$, 1-forms are often identified with (or confused with) vector fields. For example, the differential of a function, $df = f_x\, dx + f_y\, dy + f_z\, dz$, is treated as a vector field by writing $\nabla f = \text{grad } f = f_x i + f_y j + f_z k$. That is why one calls a force a vector and not a covector even when it is the gradient of a potential function.

Also, since the dimension of the space of 2-linear forms in a 3-dimensional space is $\binom{3}{2} = 3$ classically 2-forms in $\mathbb{R}^3$ were identified with (or confused with) vector fields. Usually one identifies $aj \wedge k + bk \wedge i + ci \wedge j$ with $ai + bj + ck$. Think about the cross product of vectors. This is why angular momentum and magnetic fields are sometimes misrepresented as vectors.

Given a 0-form F, i.e., a function, dF is a 1-form. The natural generalization is the *exterior derivative operator* d which converts a k-form F as given in (3) into a $(k+1)$-form dF by the formula

$$dF = \sum_{j=1}^{m} \sum_{1 \le i_1 \dots < i_k \le m} \frac{\partial f_{i_1 \dots i_k}(x)}{\partial x_j} dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \sum_{j=1}^{m} \sum_{I} df_i(x) \wedge dx^i. \tag{4}$$

Lemma 1. *Let F and G be smooth forms defined on an open set $\mathbb{O}$. Then*

(i) $d(F+G) = dF + dG$,
(ii) $d(F \wedge G) = dF \wedge G + (-1)^{\deg(F)} F \wedge dG$,
(iii) $d(dF) = 0$ *for all* F,
(iv) *if F is a function, then dF agrees with the standard definition of the differential of F,*
(v) *the operator d is uniquely defined by the properties given above.*

PROOF. Part (iv) is obvious, and parts (i), (ii), and (v) are left as exercises. Part (iii) will be proved here. Let i be a multiple index, and so the summations on i range over I. Let $F = \sum_i f_i \, dx^i$. Then

$$dF = \sum_i \sum_{j=1}^{n} (\partial f_i / \partial x_j) \, dx^j \wedge dx^i,$$

$$d(dF) = \sum_i \sum_{j=1}^{n} \sum_{k=1}^{n} \left(\frac{\partial^2 f_i}{\partial x_j \partial x_k} \right) dx^k \wedge dx^j \wedge dx^i$$

$$= \sum_i \sum_{j<k} \left(\frac{\partial^2 f_i}{\partial x_j \partial x_k} - \frac{\partial^2 f_i}{\partial x_k \partial x_j} \right) dx^k \wedge dx^j \wedge dx^i = 0.$$

The last sum is zero by the equality of mixed partial derivatives. ∎

Remark. The first four can be used as a coordinate free definition of the operator d. Formula (4) shows its existence, and part (v) shows its uniqueness.

Let (x, y, z) be the standard coordinates in $\mathbb{R}^3$ and i, j, k the usual unit vectors. If $F(x, y, z)$ is a function, then

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial z} dz$$

is the usual differential. The classical approach is to make the differential a vector field by defining

$$\nabla F = \operatorname{grad} F = \frac{\partial F}{\partial x} i + \frac{\partial F}{\partial y} j + \frac{\partial F}{\partial z} k.$$

Next consider a 1-form $F = a(x, y, z) \, dx + b(x, y, z) \, dy + c(x, y, z) \, dz$, then

$$dF = \left(\frac{\partial c}{\partial y} - \frac{\partial b}{\partial z}\right) dy \wedge dz + \left(\frac{\partial a}{\partial z} - \frac{\partial c}{\partial x}\right) dz \wedge dx + \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right) dx \wedge dy.$$

The classical approach is to make this F a vector field $F = ai + bj + ck$ and to define a new vector field by

$$\nabla \times F = \operatorname{curl} F = \left(\frac{\partial c}{\partial y} - \frac{\partial b}{\partial z}\right) i + \left(\frac{\partial a}{\partial z} - \frac{\partial c}{\partial x}\right) j + \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right) k.$$

Now let F be a 2-form so $F = a\, dy \wedge dz + b\, dz \wedge dx + c\, dx \wedge dy$ and

$$dF = \left(\frac{\partial a}{\partial x} + \frac{\partial b}{\partial y} + \frac{\partial c}{\partial x}\right) dx \wedge dy \wedge dz.$$

The classical approach would have considered F as a vector field $F = ai + bj + ck$ and defined a scalar function

$$\nabla \cdot F = \operatorname{div} F = \left(\frac{\partial a}{\partial x} + \frac{\partial b}{\partial y} + \frac{\partial c}{\partial x}\right).$$

The statement that $d(dF) = 0$, or $d^2 = 0$, contains the two classical statement $\operatorname{curl}(\operatorname{grad} F) = 0$ and $\operatorname{div}(\operatorname{curl} F) = 0$.

A k-form, F, *is closed* if $dF = 0$. A k-form, F, *is exact* if there is a $(k-1)$-form G such that $F = dG$. Part (iii) of Lemma 1 says that an exact form is closed. A partial converse is also true as we shall see.

Theorem 2 (Poincare's Lemma). *Let $\mathbb{O}$ be a ball in $\mathbb{R}^m$ and F a k-form such that $dF = 0$. Then there is a $(k-1)$-form f on $\mathbb{O}$ such that $F = df$.*

Remark. This is a partial converse to $d(df) = 0$. Note that the domain of definition, $\mathbb{O}$, of the form F is required to be a ball. The theorem says that in a ball, a closed form is exact. The 1-form, $F = (x\, dy - y\, dx)/(x^2 + y^2)$, satisfies $dF = 0$, but there does not exist a function, f, defined on all of $\mathbb{R}^2 \backslash (0, 0)$ such that $df = F$. The form F is the differential of the polar angle $\theta = \arctan(y/x)$ which is not a single-valued function defined on all of $\mathbb{R}^2 \backslash (0, 0)$. However, it can be made single valued in a neighborhood of any point in $\mathbb{R}^2 \backslash (0, 0)$, e.g., for any point not on the negative x axis, one can take $-\pi < \theta < \pi$, and for points on the negative x axis, one can take $0 < \theta < 2\pi$. Since F locally defines a function, we have $dF = 0$.

Poincarè's lemma contains classical theorems: (i) if F is a vector field defined on a ball in $\mathbb{R}^3$ with $\operatorname{curl}(F) = 0$, then there is a smooth function f such that $F = \operatorname{grad}(f)$, and (ii) if F is a smooth vector field such that $\operatorname{div}(F) = 0$, then there is a smooth vector field f such that $F = \operatorname{curl}(f)$.

Proof. The full statement of the Poincarè Lemma will not be needed here—only the case when $k = 1$ will be used in subsequent chapters. Therefore, only

that case will be proved here. The proof of the full theorem can be found in Flanders (1963), or Spivak (1965), or Abraham and Marsden (1978).

Let $F = \sum_i f_i(x)\, dx^i$ be a given 1-form.

$$dF = \sum_i \sum_j \left(\frac{\partial f_i}{\partial x^j}\right) dx^j \wedge dx^i = \sum_{i<j} \left(\frac{\partial f_i}{\partial x^j} - \frac{\partial f_j}{\partial x^i}\right) dx^j \wedge dx^i.$$

So $dF = 0$ if and only if $\partial f_i/\partial x^j = \partial f_j/\partial x^i$. Define $g(x) = \int_0^1 \sum_i f_i(tx) x^i\, dt$. So

$$\begin{aligned}
\frac{\partial g(x)}{\partial x^j} &= \int_0^1 \left\{\sum_i \left(\frac{\partial f_i(tx)}{\partial x^j}\right) tx^i + f_j(tx)\right\} dt \\
&= \int_0^1 \left\{\sum_i \left(\frac{\partial f_j(tx)}{\partial x^i}\right) tx^i + f_j(tx)\right\} dt \\
&= \int_0^1 \left\{\frac{t\, df_j(tx)}{dt} + f_j(tx)\right\} dt \\
&= tf_j(tx)\big|_0^1 = f_j(x).
\end{aligned}$$

Thus, $dg = F$. ∎

Note that the function g defined in the proof given above is a line integral and the condition $dF = 0$ is the condition that a line integral be independent of path.

Corollary 3. *Let $F = (F^1, \ldots, F^m)$ be a vector field defined in a ball $\mathbb{O}$ in $\mathbb{R}^m$. Then a necessary and sufficient condition for F to be the gradient of a function $f: \mathbb{O} \to \mathbb{R}$ is that the Jacobian matrix $(\partial F^i/\partial x^j)$ be symmetric.*

Proof. First, to see that it is a corollary, consider F as the differential form $F = F^1\, dx^1 + \cdots + F^m\, dx^m$. Then by the above,

$$dF = \sum_{i<j} \left(\frac{\partial F^i}{\partial x^j} - \frac{\partial F^j}{\partial x^i}\right) dx^i \wedge dx^j.$$

So $dF = 0$ if and only if the Jacobian $(\partial F^i/\partial x^j)$ is symmetric. Corollary 3 follows from Lemma 1 (iii) and Theorem 2. ∎

E. Changing Coordinates and Darboux's Theorem

To change coordinates for vector fields or differential forms, simply transform the coordinates as was done in Section C using the Jacobian of the transformation. In particular, let x and y be coordinates on $\mathbb{O}$, and assume that the change of coordinates is given by $x = \phi(y)$ and the change back by $y = \psi(x)$, or in classical notation $x = x(y)$ and $y = y(x)$. Assume the Jacobians, $D\phi = \{\partial y^j/\partial x^i\}$ and $D\psi = \{\partial x^i/\partial y^j\}$, are nonsingular.

If a vector field V is given by

$$V = \sum_{i=1}^{m} \alpha^i(x) \frac{\partial}{\partial x^i} = \sum_{i=1}^{m} \beta^i(x) \frac{\partial}{\partial y^i}, \tag{1}$$

and we set $\mathbf{a}(x) = (\alpha^1(x), \ldots, \alpha^m(x))$, $\mathbf{b}(y) = (\beta^1(y), \ldots, \beta^m(y))$, then

$$\mathbf{a} = D\psi(\mathbf{b}) \quad \text{or} \quad \alpha^i = \sum_{j=1}^{m} \frac{\partial x^i}{\partial y^j} \beta^j. \tag{2}$$

If a differential 1-form is given by

$$F = \sum_{i=1}^{m} \alpha_i(x)\, dx^i = \sum_{i=1}^{m} \beta_i(y)\, dy^i, \tag{3}$$

and we set $\mathbf{a}(x) = (\alpha_1(x), \ldots, \alpha_m(x))$ and $\mathbf{b}(y) = (\beta_1(y), \ldots, \beta_m(y))$, then

$$\mathbf{a} = bD\phi \quad \text{or} \quad \alpha_i = \sum_{j=1}^{m} \frac{\partial y^j}{\partial x^i} \beta_j. \tag{4}$$

If a differentiable 2-form F is given by

$$F = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij}(x)\, dx^i \wedge dx^j = \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij}(y)\, dy^i \wedge dy^j, \tag{5}$$

and we set $\mathbf{A} = \{\alpha_{ij}\}$, $\mathbf{B} = \{\beta_{ij}\}$ ($\mathbf{A}$ and $\mathbf{B}$ are skew-symmetric matrices), then

$$\mathbf{A} = D\phi^T \mathbf{B} D\phi \quad \text{or} \quad \alpha_{ij} = \sum_{s=1}^{m} \sum_{r=1}^{m} \frac{\partial y^s}{\partial x^i} \frac{\partial y^r}{\partial x^i} \beta_{sr}. \tag{6}$$

Let $\mathbb{O}$ be an open set in $\mathbb{R}^{2n}$. A 2-form F on $\mathbb{O}$ is *nondegenerate* if $F^n = F \wedge F \wedge \cdots \wedge F$ (n times) is nonzero. As we saw above, the coefficients in a coordinate system of a 2-form can be represented as a skew-symmetric matrix. As we saw in Section B, a linear 2-form is nondegenerate if and only if the coefficient matrix is nonsingular. Thus, the 2-form F in (5) is nondegenerate if and only if $\mathbf{A}$ (or $\mathbf{B}$) is nonsingular on all of $\mathbb{O}$. A *symplectic structure* or *symplectic form on* $\mathbb{O}$ is a closed nondegenerate 2-form. The *standard symplectic structure* in $\mathbb{R}^{2n}$ is

$$\Omega = \sum_{i=1}^{n} dz^i \wedge dz^j = \sum_{i=1}^{n} dq^i \wedge dp^i, \tag{7}$$

where $z = (z^1, \ldots, z^{2n}) = (q^1, \ldots, q^n, p^1, \ldots, p^n)$ are coordinates in $\mathbb{R}^{2n}$. The coefficient matrix of Ω is just J. By Corollary II.B.2, there is a linear change of coordinates so that the coefficient matrix of a nondegenerate 2-form is J at one point. A much more powerful result which will not be needed in the subsequent chapters is the following.

Theorem 1 (Darboux's Theorem). *If F is a symplectic structure on an open ball in $\mathbb{R}^{2n}$, then there exists a coordinate system z such that F in this coordinate system is the standard symplectic structure Ω.*

PROOF. See Abraham and Marsden (1978). ∎

A coordinate system for which a symplectic structure is Ω is called *symplectic coordinate (for this form)*. A *symplectic transformation*, ϕ, is one which preserves the form Ω or preserves the coefficient matrix J, i.e., $D\phi^T J D\phi = J$.

F. Integration and Stokes' Theorem

We shall not need any result from integration theory on manifolds, and so we will not develop the theory here. To tease the reader into learning more about this subject, consider a weak form of the general Stokes' theorem. It will illustrate the power and beauty of differential forms. Let M be an n-dimensional oriented manifold with an $(n-1)$-dimensional boundary ∂M. Let the boundary ∂M be oriented consistently with M. Let ω be an $(n-1)$-form on M; so, $d\omega$ is an n-form on M. One can define the integral of an n-form on an n-manifold in a logical way, and then one has:

$$\int_{\partial M} \omega = \int_M d\omega \quad \text{(Stokes' Theorem)}.$$

This one general theorem contains Green's theorem, the divergence theorem, and the classical Stokes' theorem of classical vector calculus. See Spivak (1965) for a complete discussion of the general Stokes' theorem and all its ramifications.

Problems

1. Show that if $f^1, \ldots, f^k$ are 1-forms, then

$$f^1 \wedge \cdots \wedge f^k(a_1, \ldots, a_k) = \det \begin{bmatrix} f^1(a_1) & \cdots & f^k(a_1) \\ \vdots & & \vdots \\ f^1(a_k) & \cdots & f^k(a_k) \end{bmatrix}.$$

2. Show that the mapping $(f^1, f^2) \to f^1 \wedge f^2$ is a skew-symmetric bilinear map from $\mathbb{V}^* \times \mathbb{V}^* \to \mathbb{A}^2$.

3. Let F and G be 0-, 1-, or 2-forms in $\mathbb{R}^3$. Verify Lemma D.1 in this case.

4. a. Let F be a 1-form in $\mathbb{R}^3$ such that $dF = 0$. Verify that if $F = a\,dx + b\,dy + c\,dz$, then $\partial a/\partial y = \partial a/\partial z$, $\partial b/\partial y = \partial b/\partial x$, $\partial c/\partial x = \partial c/\partial y$. Also verify that if

$$f(x, y, z) = \int_0^1 (a(tx, ty, tz)x + b(tx, ty, tz)y + c(tx, ty, tz)z)\,dt,$$

then $F = df$.

b. Let F be a 2-form in $\mathbb{R}^3$ such that $dF = 0$. Verify that if $F = a\,dy \wedge dz + b\,dz \wedge dx + c\,dx \wedge dy$, then $\partial a/\partial x + \partial b/\partial y + \partial c/\partial z = 0$. Also verify that $F = df$ where

$$f = \left(\int_0^1 (a(tx, ty, tz)t\,dt\right)(y\,dz - z\,dy) + \left(\int_0^1 (b(tx, ty, tz)t\,dt\right)(z\,dx - x\,dz) + \left(\int_0^1 (c(tx, ty, tz)t\,dt\right)(x\,dy - y\,dx).$$

5. Prove that the $\wedge$ operator is bilinear and associative. (See Lemma A.1.)

6. **a.** Show that the operator d which operates on smooth forms is linear, i.e., $d(F + G) = dF + dG$.

b. Show that d satisfies a product rule, i.e., $d(F \wedge G) = dF \wedge G + (-1)^{\deg(F)}F \wedge dG$.

c. Show that if δ is a mapping which takes smooth k-forms to $(k + 1)$-forms and satisfies (a) $\delta(F + G) = \delta F + \delta G$, (b) $\delta(F \wedge G) = \delta F \wedge G + (-1)^{\deg(F)}F \wedge \delta G$, (c) $\delta(\delta F) = 0$ for all F, and (d) if F is a function, then δF agrees the standard definition of the differential of F, then δ is the same as the operator d given by the Formula (D.4).

7. Let $Q(q, p)$ and $P(q, p)$ be smooth functions defined on an open set in $\mathbb{R}^2$. Consider the four differential forms $\Omega_1 = P\,dQ - p\,dq$, $\Omega_2 = P\,dQ + q\,dp$, $\Omega_3 = Q\,dP + p\,dq$, $\Omega_4 = Q\,dP - q\,dp$.

a. Show that Ω_i is exact if and only if Ω_j is exact for $j, j = 1, 2, 3, 4$.

b. Show that Ω_i is closed if and only if Ω_j is closed for $i, j = 1, 2, 3, 4$.

c. Show that if Ω_i is exact (or closed), then so is $\Theta = (Q - q)d(P + p) - (P - p)d(Q + q)$. [Hint: $d(qp) = q\,dp + p\,dq$ is exact.]

CHAPTER IV

Symplectic Transformations and Coordinates

A. Symplectic Transformations

The form of Hamilton's equations is very special, and the special form is not preserved by an arbitrary change of variables; so, the change of variables that preserve that special form are very important in the theory. The classical subject of celestial mechanics is replete with special coordinate systems which bear the names of some of the greatest mathematicians. We shall consider some of them in this chapter.

1. General Definition

Let $\Xi: O \to \mathbb{R}^{2n}: (t, z) \to \zeta = \Xi(t, z)$ be a smooth function where O is some open set in $\mathbb{R}^{2n+1}$; Ξ is called *a symplectic function* (*or transformation or map*, etc.) if the Jacobian of Ξ with respect to z, $D_2\Xi(t, z) = \partial\Xi/\partial z$, is a symplectic matrix at every point of $(t, z) \in O$. Sometimes we will use the notation $D_2\Xi$ for the Jacobian of Ξ, and sometimes the notation $\partial\Xi/\partial z$ will be used. In the first case we think of the Jacobian $D_2\Xi$ as a map from O into the space $\mathscr{L}(\mathbb{R}^{2n}, \mathbb{R}^{2n})$ of linear operators from $\mathbb{R}^{2n}$ to $\mathbb{R}^{2n}$, and in the second case, we think of $\partial\Xi/\partial z$ as the matrix

$$\frac{\partial\Xi}{\partial z} = \begin{bmatrix} \dfrac{\partial\Xi_1}{\partial z_1} & \cdots & \dfrac{\partial\Xi_1}{\partial z_{2n}} \\ \dfrac{\partial\Xi_{2n}}{\partial z_1} & \cdots & \dfrac{\partial\Xi_{2n}}{\partial z_{2n}} \end{bmatrix}.$$

Thus, Ξ is symplectic if and only if

$$\frac{\partial \Xi^T}{\partial z} J \frac{\partial \Xi}{\partial z} = J. \tag{2}$$

Since the product of two symplectic matrices is symplectic, the composition of two symplectic maps is symplectic by the chain rule of differentiation. Since a symplectic matrix is invertible, and its inverse is symplectic, the inverse function theorem implies that a symplectic map is locally invertible and its inverse, $Z(t, \zeta)$, is symplectic where defined. Since the determinant of a symplectic matrix is $+1$, the transformation is orientation and volume preserving.

If the transformation $z \to \zeta = \Xi(t, z)$ is considered a change of variables, then one calls ζ symplectic or canonical coordinates. Consider a nonlinear Hamiltonian system

$$\dot{z} = J\nabla_z H(t, z) \tag{3}$$

where H is defined and smooth in some open set $O \subset \mathbb{R}^{2n+1}$. Make a symplectic change of variables from z to ζ by

$$\zeta = \Xi(t, z) \quad \text{with inverse } z = Z(t, \zeta) \tag{4}$$

[so $\zeta \equiv \Xi(t, Z(t, \zeta)), z \equiv Z(t, \Xi(t, z))$]. Let $\mathbb{O} \in \mathbb{R}^{2n+1}$ be the image of O under this transformation. Then the Hamiltonian $H(t, z)$ transforms to the function $\hat{H}(t, \zeta) = H(t, Z(t, \zeta))$. Later we will abuse notation and write $H(t, \zeta)$ instead of introducing a new symbol, but now we will be careful to distinguish H and $\hat{H}$. Equation (3) transforms to

$$\begin{aligned}
\dot{\zeta} &= \frac{\partial \Xi}{\partial t}(t, z) + \frac{\partial \Xi}{\partial z}(t, z)\dot{z} = \frac{\partial \Xi}{\partial t}(t, z) + \frac{\partial \Xi}{\partial z}(t, z) J \left(\frac{\partial H}{\partial z}(t, z)\right)^T \\
&= \frac{\partial \Xi}{\partial t}(t, z) + \frac{\partial \Xi}{\partial z}(t, z) J \left(\frac{\partial \hat{H}}{\partial \zeta}(t, \zeta) \frac{\partial \Xi}{\partial z}(t, z)\right)^T \\
&= \frac{\partial \Xi}{\partial t}(t, z) + J\left(\frac{\partial \hat{H}}{\partial \zeta}\right)^T = \frac{\partial \Xi}{\partial t}(t, z)\bigg|_{z=Z(t,\zeta)} + J\nabla_\zeta \hat{H}(t, \zeta).
\end{aligned} \tag{5}$$

The notation in the second to last term in (5) means that you are to take the partial derivative with respect to t first and then substitute in $z = Z(t, \zeta)$. If the change of coordinates, Ξ, is independent of t, then the term $\partial \Xi / \partial t$ is missing in (5); so, the equation in the new coordinates is simply $\dot{\zeta} = J\nabla_\zeta \hat{H}$, a Hamiltonian system with Hamiltonian $\hat{H}$. In this case one simply substitutes the change of variables into the Hamiltonian H to get the new Hamiltonian $\hat{H}$. In this case the Hamiltonian character of the equations is preserved. Actually the system (5) is still Hamiltonian even if Ξ depends on t, provided $\mathbb{O}$ is a nice set as we shall see in the next paragraph.

For each fixed t, let the set $\mathbb{O}_t = \{\zeta : (t, \zeta) \in \mathbb{O}\}$ be a ball in $\mathbb{R}^{2n}$. We will show that there is a smooth function $R : \mathbb{O} \to \mathbb{R}^1$ such that

$$\frac{\partial \Xi}{\partial t}(t, z)\bigg|_{z=Z(t,\zeta)} = J\nabla_\zeta R(t, \zeta). \tag{6}$$

R is called the *remainder function.* Therefore, in the new coordinates, Equation (5) is Hamiltonian with Hamiltonian $R(t, \zeta) + \hat{H}(t, \zeta)$. [In case $\mathbb{O}_t$ is not a ball, the above holds locally, i.e., at each point of $p \in \mathbb{O}$ there is a function R defined in a neighborhood of p such that (6) holds in the neighborhood, but R may not be globally defined as a single-valued function on all of $\mathbb{O}$.] By Corollary III.D.3, we must show that J times the Jacobian of the left-hand side of (6) is symmetric. That is, we must show

$$\Gamma = \Gamma^T, \tag{7}$$

where

$$\Gamma(t, \zeta) = J \frac{\partial^2 \Xi}{\partial t \partial z}(t, z)\bigg|_{z=Z(t,\zeta)} \frac{\partial Z}{\partial \zeta}(t, \zeta).$$

Differentiating (2) with respect to t gives

$$\begin{aligned} &\frac{\partial^2 \Xi^T}{\partial t \partial z}(t, z) J \frac{\partial \Xi}{\partial z}(t, z) + \frac{\partial \Xi}{\partial z}(t, z) J \frac{\partial^2 \Xi}{\partial t \partial z}(t, z) = 0, \\ &\frac{\partial \Xi^{-T}}{\partial z}(t, z) \frac{\partial^2 \Xi^T}{\partial t \partial z}(t, z) J + J \frac{\partial^2 \Xi}{\partial t \partial z}(t, z) \frac{\partial \Xi^{-1}}{\partial z}(t, z) = 0. \end{aligned} \tag{8}$$

Substituting $z = Z(t, \zeta)$ into (8) and noting that $(\partial \Xi^{-1}/\partial z)(t, Z(t, \zeta)) = \partial Z(t, \zeta)/\partial \zeta$ yields $-\Gamma^T + \Gamma = 0$. Thus we have shown:

Theorem 1. *A symplectic change of variables takes a Hamiltonian system of equations into a Hamiltonian system.*

A partial converse is also true. If a change of variables preserves the Hamiltonian form of all Hamiltonian equations, then it is symplectic. We will not need this result and leave it as an exercise.

2. The Variational Equations

Let $\phi(t, \tau, \zeta)$ be the general solution of (3); so, $\phi(\tau, \tau, \zeta) = \zeta$, and let $X(t, \tau, \zeta)$ be the Jacobian of ϕ with respect to ζ, i.e., $X(t, \tau, \zeta) = \partial \phi(t, \tau, \zeta)/\partial \zeta$. $X(t, \tau, \zeta)$ is called *the monodromy matrix.* Substituting ϕ into (3) and differentiating with respect to ζ gives

$$\dot{X} = JS(t, \tau, \zeta)X, \qquad S(t, \tau, \zeta) = \frac{\partial^2 H}{\partial x^2}(t, \phi(t, \tau, \zeta)). \tag{9}$$

Equation (9) is called the variational equation and is a linear Hamiltonian system. Differentiating the identity $\phi(\tau, \tau, \zeta) = \zeta$ with respect to ζ gives $X(\tau, \tau, \zeta) = I$, the $2n \times 2n$ identity matrix; so, X is a fundamental matrix solution of the variational equation. By Theorem II.A.3, X is symplectic.

Theorem 2. *Let $\phi(t, \tau, \zeta)$ be the general solution of the Hamiltonian system* (3). *Then for fixed t and τ, the map $\zeta \to \phi(t, \tau, \zeta)$ is symplectic. Conversely, if*

$\phi(t, \tau, \zeta)$ is the general solution of a differential equation $\dot{z} = f(t, z)$, where f is defined and smooth on $I \times O$, I an interval in $\mathbb{R}$ and O a ball in $\mathbb{R}^{2n}$, and the map $\zeta \to \phi(t, \tau, \zeta)$ is always symplectic, then the differential equations $\dot{z} = f(t, z)$ is Hamiltonian.

PROOF. The direct statement was proved above; now consider the converse. Let $\phi(t, \tau, \zeta)$ be the general solution of $\dot{z} = f(t, z)$, and let $X(t, \tau, \zeta)$ be the Jacobian of ϕ. Since X is symplectic, $\dot{X}X^{-1}$ is Hamiltonian, and $-J\dot{X}X^{-1}$ is symmetric. But $\dot{X}(t, \tau, \zeta) = (\partial f/\partial z)(t, \phi(t, \tau, \zeta))X$; so, $-J\partial f/\partial z$ is symmetric. Since O is a ball, $-Jf$ is a gradient of a function H by Corollary III.D.3. Thus, $f(t, z) = J\nabla H(t, z)$. ∎

This theorem says that the flow defined by an autonomous Hamiltonian system is volume preserving. So, in particular, there cannot be an asymptotically stable equilibrium point, periodic solution, etc. This makes the stability theory of Hamiltonian systems difficult and interesting. In general, it is difficult to construct a symplectic transformation with nice properties using definition (2). The theorem above gives one method of assuring that a transformation is symplectic, and this is the basis of the method of Lie transforms explored in Chapter 6.

3. Poisson Brackets

Let $F(t, z)$ and $G(t, z)$ be smooth, and recall the definition of the Poisson bracket $\{F, G\}_z(t, z) = \nabla_z F(t, z)^T J \nabla_z G(t, z)$. Here we subscript the bracket to remind us it is a coordinate-dependent definition. Let $\hat{F}(t, \zeta) = F(t, Z(t, \zeta))$ and $\hat{G}(t, \zeta) = G(t, Z(t, \zeta))$ where Z is symplectic; so,

$$\begin{aligned}
\{\hat{F}, \hat{G}\}_\zeta(t, \zeta) &= \nabla_\zeta \hat{F}(t, \zeta)^T J \nabla_\zeta \hat{G}(t, \zeta) \\
&= \left(\frac{\partial Z^T}{\partial \zeta}(t, \zeta)\nabla_z F(t, Z(t, \zeta))\right)^T J \frac{\partial Z^T}{\partial \zeta}(t, \zeta)\nabla_z F(t, Z(t, \zeta)) \\
&= \nabla_z F(t, Z(t, \zeta))^T \frac{\partial Z}{\partial \zeta}(t, \zeta) J \frac{\partial Z^T}{\partial \zeta}(t, \zeta)\nabla_z F(t, Z(t, \zeta)) \qquad (10) \\
&= \nabla_z F(t, Z(t, \zeta))^T J \nabla_z F(t, Z(t, \zeta)) \\
&= \{F, G\}_z(t, Z(t, \zeta)).
\end{aligned}$$

Equation (10) shows that the Poisson bracket operation is invariant under symplectic changes of variables. That is, you can commute the operations of computing Poisson brackets and making a symplectic change of variables.

Theorem 3. *A symplectic change of coordinates preserves Poisson brackets.*

Ξ is symplectic if its Jacobian is a symplectic matrix or it satisfies (2). But by Theorem II.A.2, Ξ is sympectic if and only if the transpose of the Jacobian of Ξ is symplectic, i.e.,

$$\frac{\partial \Xi}{\partial z} J \frac{\partial \Xi^T}{\partial z} = J. \tag{11}$$

Let $\zeta_i = \Xi_i(t, z)$ be the ith component of the transformation. In components, (11) says

$$\{\Xi_i, \Xi_j\} = J_{ij}, \tag{12}$$

where $J = (J_{ij})$.

If the transformation (4) is given in the classical notation

$$Q_i = Q_i(q, p), \qquad P_i = P_i(q, p), \tag{13}$$

then (12) becomes

$$\{Q_i, Q_j\} = 0, \qquad \{P_i, P_j\} = 0, \qquad \{Q_i, P_j\} = \delta_{ij}, \tag{14}$$

where δ_{ij} is the Kronecker delta.

Theorem 4. *The transformation* (4) *is symplectic if and only if* (12) *holds, or the transformation* (13) *is symplectic if and only if* (14) *holds.*

B. Applications

1. The N-body Problem in Rotating Coordinates

Let

$$K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \exp(\omega K t) = \begin{pmatrix} \cos \omega t & \sin \omega t \\ -\sin \omega t & \cos \omega t \end{pmatrix} \tag{1}$$

be 2×2 matrices, and consider the planar N-body problem; so, the vectors q_i, p_i in Section I.C are 2-vectors. Introduce a set of coordinates which uniformly rotate with frequency ω by

$$u_i = \exp(\omega K t) q_i, \qquad v_i = \exp(\omega K t) p_i. \tag{2}$$

Since K is skew symmetric, $\exp(\omega K t)$ is orthogonal for all t; so, the change of variables is symplectic. The remainder function is $-\sum \omega u_i^T K v_i$, and so the Hamiltonian of the N-body problem in rotating coordinates is

$$H = \sum_{i=1}^{N} \frac{\|v\|^2}{2m_i} - \sum_{i=1}^{N} \omega u_i^T K v_i - \sum_{1 \le i, j \le N} \frac{m_i m_j}{\|u_i - u_j\|}. \tag{3}$$

The remainder term give rise to extra terms in the equations of motion which are sometimes called Coriolis forces.

2. Jacobi Coordinates

Jacobi coordinates are ideal coordinates for many investigations in the N-body problem. First, one coordinate locates the center of mass of the system, and so it can be set to zero and ignored in subsequent considerations. One of the other coordinates is the vector from one particle to another, and this is useful when studying the case when two particles are close together. Let q_i, $p_i \in \mathbb{R}^3$ for $i = 1, \ldots, N$ be the coordinates of the N-body problem as discussed in Section I.C. Define a sequence of transformations starting with $g_1 = q_1$ and $\mu_1 = m_1$ and proceed inductively by

$$T_k: \begin{cases} u_k = q_k - g_{k-1}, \\ g_k = (1/\mu_k)(m_k q_k + \mu_{k-1} g_{k-1}), \\ \mu_k = \mu_{k-1} + m_k \end{cases} \tag{4}$$

for $k = 2, \ldots, N$. μ_k is the total mass, and g_k is the position vector of the center of mass of the system of particles with indices $1, 2, \ldots, k$. The vector u_k is the position of the kth particle relative to the center of mass of the previous $k-1$ particles. See Figure B.1. Consider T_k as a change of coordinates from g_{k-1}, $u_2, \ldots, u_{k-1}, q_k, \ldots, q_N$ to $g_k, u_2, \ldots, u_k, q_{k+1}, \ldots, q_N$ or simply from g_{k-1}, q_k to g_k, q_{k+1}. The inverse of T_k is

$$T_k^{-1}: \begin{cases} q_k = (\mu_{k-1}/\mu_k)u_k + g_k, \\ g_{k-1} = (-m_k/\mu_k)u_k + g_k. \end{cases} \tag{5}$$

This is a linear transformation on the q variables only, i.e., on a Lagrangian subspace; so, Lemma II.B.14 forces the transformation on the p variables in order to have a symplectic transformation. To make the symplectic comple-

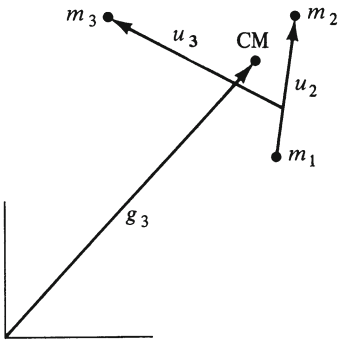

Figure B.1. Jacobi coordinate for the 3-body problem.

tion of T^{-1}, define $G_1 = p_1$ and

$$Q_k: \begin{cases} v_k = (\mu_{k-1}/\mu_k)p_k - (m_k/\mu_k)G_{k-1}, \\ G_k = p_k + G_{k-1}, \end{cases} \tag{6}$$

$$Q_k^{-1}: \begin{cases} p_k = v_k + (m_k/\mu_k)G_k, \\ G_{k-1} = -v_k + (\mu_{k-1}/\mu_k)G_k. \end{cases} \tag{7}$$

Here G_k is the total linear momentum of the system of particles with indices $1, 2, \ldots, k$.

If we denote the coefficient matrix in (4) by A, then the coefficient matrices in (5), (6), and (7) are A^{-1}, A^{-T}, and A^T, respectively; so, the pair T_k, Q_k is a symplectic change of variables. Thus, the composition of all these changes is symplectic, and the total set $g_N, u_2, \ldots, u_N, G_N, v_2, \ldots, v_N$ forms a symplectic coordinate system known as the Jacobi coordinates.

These variables satisfy the identities

$$g_{k-1} \times G_{k-1} + q_k \times p_k = g_k \times G_k + u_k \times v_k \tag{8}$$

and

$$\frac{\|G_{k-1}\|^2}{2\mu_{k-1}} + \frac{\|p_k\|^2}{2m_k} = \frac{\|G_k\|^2}{2\mu_k} + \frac{\|v_k\|^2}{2M_k}, \tag{9}$$

where $M_k = m_k\mu_{k-1}/\mu_k$. Thus, kinetic energy is

$$\mathrm{KE} = \sum_{k=1}^{N} \frac{\|p_k\|^2}{2m_k} = \frac{\|G_N\|^2}{2\mu_N} + \sum_{k=2}^{N} \frac{\|v_k\|^2}{2M_k}, \tag{10}$$

and total angular momentum is

$$A = \sum_{1}^{N} q_k \times p_k = g_N \times G_N + \sum_{2}^{N} u_k \times v_k. \tag{11}$$

Also, g_N is the center of mass of the system, and G_N is the total linear momentum.

Unfortunately, the formulas for the variables u_k and v_k are not simply expressed in terms of the variables q_k and p_k. Note that

$$u_2 = q_2 - q_1. \tag{12}$$

Let $d_{ij} = q_i - q_j$ and so the Hamiltonian of the N-body problem in Jacobi coordinates is

$$H = \frac{\|G_N\|^2}{2\mu_N} + \sum_{k=2}^{N} \frac{\|v_k\|^2}{2M_k} - \sum_{1 \le i < j \le N} \frac{m_i m_j}{\|d_{ij}\|}. \tag{13}$$

Note that the Hamiltonian is independent of g_N, and so, $\dot{G}_N = 0$ or G_N is an integral of the system. When a variable does not appear in the Hamiltonian, it is called *ignorable*, and its conjugate variable is an integral. Since $\dot{g}_N = G_N/\mu_N$, the center of gravity moves with uniform rectilinear motion. In general, one may assume that the center of mass if fixed at the origin of the system and so

sets $g_N = G_N = 0$, which reduces the problem by three degrees of freedom in the spatial problem.

In the planar problem, one verifies

$$\sum_2^N q_k^T K p_k = g_N^T K G_N + \sum_2^N u_k^T K v_k, \tag{14}$$

which is the same as Formula (8). So the Hamiltonian of the planar N-body problem in rotating coordinates with the center of mass fixed at the origin is

$$H = \sum_{k=2}^N \frac{\|v_k\|^2}{2M_k} - \sum_2^N u_k^T K v_k - \sum_{1 \le i < j \le N} \frac{m_i m_j}{\|d_{ij}\|}. \tag{15}$$

3. The 2-Body Problem in Jacobi Coordinates

When $N = 2$, then (13) with $g_2 = G_2 = 0$ takes the simple form

$$H = \frac{\|v\|^2}{2M} - \frac{m_1 m_2}{\|u\|}, \tag{16}$$

where $v = v_2$, $u = u_2 = q_2 - q_1$, $M = m_1 m_2/(m_1 + m_2)$. This is just the Kepler problem, and so in Jacobi coordinates the 2-body problem is just the Kepler problem. This says that the motion of one body, say the moon, when viewed from another, say the earth, is as if the earth were a fixed body and the moon was attracted to the earth as a central force.

4. The 3-Body Problem in Jacobi Coordinates

In the 3-body problems the distances between the bodies, and hence the potential, are not too complicated in Jacobi coordinates. Moreover, the Hamiltonian of the 3-body problem in Jacobi coordinates will be transformed to polar coordinates in Section IV.C.8, which will be used in Chapter V to understand reduction of the 3-body problem and to establish the existence of periodic solutions for two small masses (Poincaré's periodic solutions of the first kind).

Let

$$\begin{aligned} M_2 &= \frac{m_1 m_2}{m_1 + m_2}, & M_3 &= \frac{m_3(m_1 + m_2)}{m_1 + m_2 + m_3}, \\ \alpha_0 &= \frac{m_2}{m_1 + m_2}, & \alpha_1 &= \frac{m_1}{m_1 + m_2}; \end{aligned} \tag{17}$$

then the Hamiltonian of the 3-body problem with center of mass fixed at the origin and zero linear momentum in Jacobi coordinates is

$$H = \frac{\|v_2\|^2}{2M_2} + \frac{\|v_3\|^2}{2M_3} - \frac{m_1 m_2}{\|u_2\|} - \frac{m_1 m_3}{\|u_3 + \alpha_0 u_2\|} - \frac{m_2 m_3}{\|u_3 - \alpha_1 u_2\|}. \tag{18}$$

See Figure B.1. Sometimes one numbers the N-bodies from 0 to $N-1$. In this case all the subscripts in (18) except the subscripts of the α's are reduced by 1, which looks nicer to some people.

C. Differential Forms and Generating Functions

Definition (A.2) is easy enough to check *a posteriori*, but it is difficult to use this definition to generate a symplectic transformation with desired properties. This section contains only a local analysis; so, we shall assume that everything is defined in some ball about the origin in $\mathbb{R}^{2n}$.

1. The Symplectic Form

Recall that in Chapter III we defined the (standard) symplectic form to be

$$\Omega = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} J_{ij}\, dz^i \wedge dz^j = \sum_{i=1}^{n} dz^i \wedge dz^{i+n} = \sum_{i=1}^{n} dq^i \wedge dp^i = dq \wedge dp. \tag{1}$$

Here we have used the noation of differential geometry and Chapter III by using superscripts for components instead of subscripts. Also we have $z = (z^1, \ldots, z^{2n}) = (q^1, \ldots, q^n, p^1, \ldots, p^n)$ as usual. Ω is closed, $d\Omega = 0$, but, in fact, it is exact because

$$\Omega = d\alpha, \qquad \alpha = \sum_{i=1}^{n} q^i\, dp^i = q\, dp. \tag{2}$$

In short, Ω is a closed, nondegenerate (the coefficient matrix is nonsingular) 2-form. By Darboux's theorem discussed in Chapter III, for any closed, nondegenerate 2-form, there are local coordinates such that in these coordinates the 2-form is given by (1). This says that $\mathbf{J}$ is simply the coefficient matrix of a closed, nondegenerate 2-form in Darboux coordinates. The left-hand side of (A.2) is just the transformation law for a 2-form with coefficient matrix J; so, a symplectic transformation is a transformation which preserves the special form of the differential form Ω. In two dimensions, $\Omega = dq \wedge dp$, the area form in $\mathbb{R}^2$, and so we see again that a two-dimensional symplectic transformation is area preserving. In higher dimensions, being symplectic is far more restrictive than being volume preserving.

2. Action-Angle Variables

The change from rectangular coordinates q, p to polar coordinates r, ϕ is not symplectic, but the following is symplectic

$$dq \wedge dp = r\,dr \wedge d\phi = d(r^2/2) \wedge d\phi = \delta I \wedge d\phi,$$

$$I = \tfrac{1}{2}(q^2 + p^2), \qquad \phi = \arctan\left(\frac{p}{q}\right), \tag{3}$$

$$q = \sqrt{2I}\cos\phi, \qquad p = \sqrt{2I}\sin\phi.$$

Therefore, I, ϕ are symplectic (or canonical) coordinates called action-angle coordinates. In Chapter II, we saw that the harmonic oscillator could be written as a Hamiltonian system with Hamiltonian

$$H = \tfrac{1}{2}\omega(q^2 + p^2) = \omega I, \tag{4}$$

and in action-angle coordinates, the equations of motion are

$$\dot{I} = \frac{\partial H}{\partial \phi} = 0, \qquad \dot{\phi} = -\frac{\partial H}{\partial I} = -\omega. \tag{5}$$

So the solutions move on the circles I = constant with uniform angular frequency ω in a counterclockwise direction.

Action-angle variables are used quite often in perturbation theory, for example, Duffing's equation has a Hamiltonian

$$H = \tfrac{1}{2}(q^2 + p^2) + \tfrac{1}{4}\gamma q^4, \tag{6}$$

where γ is a constant. Writing Duffing's Hamiltonian in action-angle variables gives

$$H = I + \gamma I^2 \cos^4\phi = I + \tfrac{1}{8}\gamma I^2\{3 + 4\cos 2\phi + \cos 4\phi\}. \tag{7}$$

The last form in (7) was named a Poisson series by Deprit for linguistic reasons. He claims that "Poisson" was a French name that is almost impossible for his English speaking friends to pronounce correctly. A *Poisson series in* $r = \sqrt{2I}$ *and* ϕ is a Fourier series in ϕ with coefficients which are polynomials in r. Such series arise from substituting action-angle variables into a power series expansion in q and p, but not all Poisson series come about in this manner. Action-angle variables are used and misused so often in celestial mechanics that we shall investigate this point in a little detail in the next subsection.

3. d'Alembert Character

Consider a Poisson series $g(r, \phi) = \sum_i a_{i0} r^i + \sum\sum (a_{ij} r^i \cos j\phi + b_{ij} r^i \sin j\phi)$. The Poisson series $g(r, \phi)$ comes from a power series $f(q, p) = \sum f_{ij} q^i p^j$ if $g(r, \theta) = f(r\cos\phi, r\sin\phi)$. The Poisson series g has the d'Alembert character if $a_{ij} = 0, b_{ij} = 0$ unless $i \geq j$ and $i \equiv j \bmod 2$ (i.e., i and j have the same parity).

Theorem 1. *The Poisson series g comes from a power series if and only if it has the d'Alembert character.*

PROOF. $x^i y^i = r^{i+j} \cos^i \phi \sin^j \phi$. *Claim*: $\cos^j \phi \sin^j \phi$ has a Fourier polynomial of the form $a_0 + \sum \{a_k \cos k\phi + b_k \sin k\phi\}$, where $a_k = b_k = 0$ unless $k \leq i + j$ and $k \equiv i + j \bmod 2$. The claim is clearly true for $i + j = 1$; so, assume it is true for $i + j < N$, and let $i + j = N$. Let $i \neq 0$; then

$$\begin{aligned}\cos^i \phi \sin^j \phi &= \cos \phi[\cos^{i-1} \phi \sin^j \phi] \\ &= \cos \phi[\alpha_0 + \sum \{\alpha_k \cos k\phi + \beta_k \sin k\phi\}] \\ &= \alpha_0 \cos \phi + \sum \frac{\alpha_k}{2}(\cos(k+1)\phi + \cos(k-1)\phi) \\ &\quad + \sum \frac{\beta_k}{2}(\sin(k+1)\phi + \sin(k-1)\phi).\end{aligned}$$

The induction hypothesis gives $\alpha_k = 0$, $\beta_k = 0$ unless $k \leq i + j - 1$, and $k \equiv i + j - 1 \bmod 2$. The last polynomial above shows that the induction hypothesis is true for $i + j = N$. Similar formulas hold when $j \neq 0$. So a power series gives rise to a Poisson series with the d'Alembert character.

Conversely, if $r^a \cos b\phi$ satisfies $a \geq b$ and $a \equiv b \bmod 2$, then $\cos b\phi = \mathscr{R}(\exp i\phi)^b =$ a sum of terms like $\cos(b - 2s)\phi \sin 2s\phi$. The d'Alembert character gives $a = b + 2p$, so $r^a \cos b\phi =$ a sum of terms like $r^{2p}\{r^{b-2p} \cos(b - 2p\Phi)\}\{r^{2p} \sin(2p\phi)\} = (x^2 + y^2)^p x^{b-2p} y^{2p}$. ∎

A perturbation analysis is often done in action-angle variables—see later chapters—keeping track of the d'Alembert character of the change of variables in action-angle variables is important in order to keep track of the analyticity of the change of variables in rectangular variables.

Say you want an analytic Hamiltonian with a fivefold symmetry. In polar coordinates the functions r^k, $\cos 5\theta$, and $\sin 5\theta$ are all invariant under the rotation $\theta \to \theta + 2\pi/5$ but are not analytic in rectangular coordinates. The functions r^{2k}, $r^{2k+5} \cos 5\theta$, and $r^{2k+5} \sin 5\theta$ are all invariant under the rotation $\theta \to \theta + 2\pi/5$ and have the d'Alembert character; therefore, any linear combination (finite or uniformly convergent for $r < \rho$, $\rho > 0$) gives an analytic function in rectangular coordinates with a fivefold symmetry.

4. Generating Functions

Use classical notation $z = (q, p)$ so that the standard symplectic form is

$$\Omega = \sum_{i=1}^{n} dq^i \wedge dp^i = dq \wedge dp. \tag{8}$$

Let $Q = Q(q, p)$, $P = P(q, p)$ be a change of variables, and assume the functions Q and P are defined in a ball in $\mathbb{R}^{2n}$. This change of variables is symplectic if and only if

$$dq \wedge dp = dQ \wedge dP. \tag{9}$$

This is equivalent to $d(q\,dp - Q\,dP) = 0$ or that $\sigma_1 = q\,dp - Q\,dP$ is exact. σ_1 is exact if and only if $\sigma_2 = \sigma_1 + d(QP) = q\,dp + P\,dQ$ is exact. In a similar manner the change of variables $Q = Q(q, p)$, $P = P(q, p)$ is symplectic if and only if any one of the following forms is exact:

$$\begin{aligned} \sigma_1 &= q\,dp - Q\,dP, & \sigma_2 &= q\,dp + P\,dQ, \\ \sigma_3 &= p\,dq - P\,dQ, & \sigma_4 &= p\,dq + Q\,dP. \end{aligned} \tag{10}$$

Since the functions Q and P are defined in a ball, exact forms are closed by the Poincaré lemma; so, the change of variables is symplectic if and only if one of the functions S_1, S_2, S_3, S_4 exists and satisfies

$$\begin{aligned} dS_1(p, P) &= \sigma_1, & dS_2(p, Q) &= \sigma_2, \\ dS_3(q, Q) &= \sigma_3, & dS_4(q, P) &= \sigma_4. \end{aligned} \tag{11}$$

In the above formulas, there is an implied summation over the components.

These statements give an easy way to construct a symplectic change of variables. Assume that there exists a function $S_1(p, P)$ such that $dS_1 = \sigma_1$; so,

$$dS_1 = \frac{\partial S_1}{\partial p}\,dp + \frac{\partial S_1}{\partial P}\,dP = q\,dp - Q\,dp.$$

So if

$$q = \frac{\partial S_1}{\partial p}(p, P), \qquad Q = -\frac{\partial S_1}{\partial P}(p, P) \tag{12}$$

defines a changes of variables from (q, p) to (Q, P), then it is symplectic. By the implicit function theorem, the equations in (12) are solvable for P as a function of q and p and for p as a function of Q and P when the Hessian of S_1 in nonsingular. Thus, in a similar manner we have

Theorem 2. *The following define a local symplectic change of variables*:

$$q = \frac{\partial S_1}{\partial p}(p, P), \qquad Q = -\frac{\partial S_1}{\partial P}(p, P) \quad \textit{when } \frac{\partial^2 S_1}{\partial p\,\partial P} \textit{ is nonsingular}; \tag{13a}$$

$$q = \frac{\partial S_2}{\partial p}(p, Q), \qquad P = \frac{\partial S_2}{\partial Q}(p, Q) \quad \textit{when } \frac{\partial^2 S_2}{\partial p\,\partial Q} \textit{ is nonsingular}; \tag{13b}$$

$$p = \frac{\partial S_3}{\partial q}(q, Q), \qquad P = -\frac{\partial S_3}{\partial Q}(q, Q) \quad \textit{when } \frac{\partial^2 S_3}{\partial q\,\partial Q} \textit{ is nonsingular}; \tag{13c}$$

$$p = \frac{\partial S_4}{\partial q}(q, P), \qquad Q = \frac{\partial S_4}{\partial P}(q, P) \quad \textit{when } \frac{\partial^2 S_4}{\partial q\,\partial P} \textit{ is nonsingular}. \tag{13d}$$

The functions S_i are called generating functions. For example, if $S_2(p, Q) = pQ$, then (13b) says that the identity transformation $Q = q$, $P = p$ is symplectic, or if $S_1(p, P) = pP$, then the switching of variables $Q = -p$, $P = q$ is symplectic.

5. Mathieu Transformations

If you are given a point transformation $Q = f(q)$, with $\partial f/\partial q$ invertible, then the transformation can be extended to a symplectic transformation by defining $S_4(q, P) = f(q)^T P$ and

$$p = \frac{\partial f}{\partial q}(q)^T P, \qquad Q = f(q). \tag{14}$$

These transformations were studied by Mathieu (1874).

6. Polar Coordinates

Let x, y be the usual coordinates in the plane and X, Y their conjugate momentum. Suppose we wish to change to polar coordinates, r, θ in the x, y plane and to extend this point transformation to a symplectic change of variables. Let R, Θ be conjugate to r, θ. By the above, we take $S = S_4 = Xr\cos\theta + Yr\sin\theta$, and so

$$\begin{aligned} x &= \frac{\partial S}{\partial X} = r\cos\theta, \qquad y = \frac{\partial S}{\partial Y} = r\sin\theta, \\ R &= \frac{\partial S}{\partial r} = X\cos\theta + Y\sin\theta = \frac{xX + yY}{r}, \\ \Theta &= \frac{\partial S}{\partial \theta} = -Xr\sin\theta + Yr\cos\theta = xY - yX. \end{aligned} \tag{15}$$

If we think of a particle of mass m moving in the plane, then $X = m\dot{x}$ and $Y = m\dot{y}$ are linear momenta in the x and y directions; so, $R = m\dot{r}$ is the linear momentum in the r direction, and $\Theta = mx\dot{y} - my\dot{x} = mr^2\dot{\theta}$ is angular momentum. The inverse transformation is

$$\begin{aligned} X &= R\cos\theta - \left(\frac{\Theta}{r}\right)\sin\theta, \\ Y &= R\sin\theta + \left(\frac{\Theta}{r}\right)\cos\theta. \end{aligned} \tag{16}$$

7. Kepler's Problem in Polar Coordinates

The Hamiltonian of Kepler's problem (I.C.7) in polar coordinates is

$$H = \frac{1}{2}(X^2 + Y^2) - \frac{\mu}{(x^2 + y^2)} = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \frac{\mu}{r}. \tag{17}$$

Since H is independent of θ, it is an ignorable coordinate, and Θ is an integral.

The equations of motion are

$$\begin{aligned} \dot{r} &= R, \qquad \dot{\theta} = \frac{\Theta}{r^2}, \\ \dot{R} &= -\frac{\Theta^2}{r^3} + \frac{\mu}{r^2}, \qquad \dot{\Theta} = 0. \end{aligned} \tag{18}$$

These equations imply that Θ, angular momentum, is constant, say c; so,

$$\ddot{r} = \dot{R} = -\frac{c^2}{r^3} + \frac{\mu}{r^2}. \tag{19}$$

This is a one-degree-of-freedom equation for r; so, it is solvable by the method discussed in Section I.B.2. Actually, this equation for r can be solved explicitly.

Assume $c \neq 0$; so, the notion is not collinear. In (19) make the changes of variables $u = 1/r$ and $dt = (r^2/c)\, d\theta$ so

$$\begin{aligned} \ddot{r} &= \frac{c}{r^2}\frac{d}{d\theta}\left\{\frac{c}{r^2}\frac{dr}{d\theta}\right\} = c^2u^2\frac{d}{d\theta}\left\{u^2\frac{du^{-1}}{d\theta}\right\} = -c^2u^2u'' \\ &= -\frac{c^2}{r^3} + \frac{\mu}{r^2} = -c^2u^2 + \mu u^2, \end{aligned} \tag{20}$$

or

$$u'' + u = \frac{\mu}{c^2}, \tag{21}$$

where $' = d/d\theta$. Equation (21) is just the nonhomogeneous harmonic oscillator which has the general solution $u = \mu/c^2(1 + \varepsilon\cos(\theta - g))$, where ε and g are integration constants. Let $f = \theta - g$; so,

$$r = \frac{c^2/\mu}{1 + \varepsilon\cos f}. \tag{22}$$

Equation (22) is the equation of a conic section in polar coordinates. Consider a line ℓ in Figure C.1 which is perpendicular to the ray at angle g through the origin and at a distance c^2/μ. Rewrite (22) as $r = \varepsilon(c^2/\mu\varepsilon - r\cos f)$, which says that the distance of the particle to the origin, r, is equal to ε times the distance of the particle to the line ℓ, $c^2/\mu\varepsilon$-rcosf. This is one of the many definitions of a conic section. One focus is at the origin. ε is the eccentricity, and the locus is circle if $\varepsilon = 0$, an ellipse if $0 < \varepsilon < 1$, a parabola if $\varepsilon = 1$, and a hyperbola if $\varepsilon > 1$.

The point of closest approach, p in Figure C.1, is called the *perihelion* if the sun is the attractor at the origin or the *perigee* if the earth is. The angle f is called the *true anomaly* and g the *argument of the perihelion* (*perigee*).

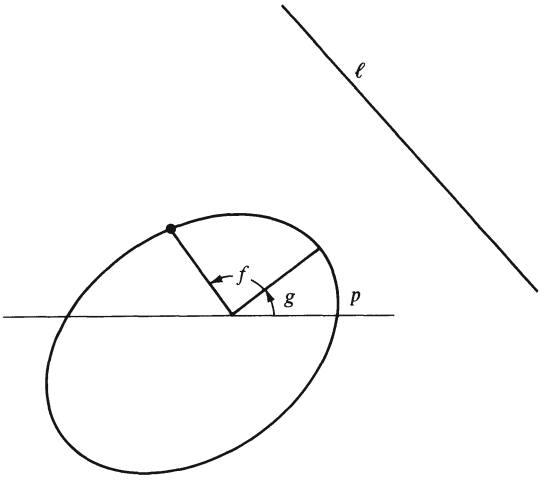

Figure C.1. An elliptic orbit.

8. The 3-Body Problem in Jacobi-Polar Coordinates

Consider the 3-body problem in Jacobi coordinates with center of mass at the origin and linear momentum zero, i.e, the Hamiltonian (B.18). Introduce polar coordinates for u_2 and u_3 as in Subsection 6). That is, let

$$\begin{aligned} u_2 &= (r_1 \cos\theta_1, r_1 \cos\theta_1), \qquad u_3 = (r_2 \cos\theta_2, r_2 \cos\theta_2), \\ v_2 &= \left(R_1 \cos\theta_1 - \left(\frac{\Theta_1}{r_1}\right)\sin\theta_1, R_1 \sin\theta_1 + \left(\frac{\Theta_1}{r_1}\right)\cos\theta_1\right), \\ v_3 &= \left(R_2 \cos\theta_2 - \left(\frac{\Theta_2}{r_2}\right)\sin\theta_2, R_2 \sin\theta_2 + \left(\frac{\Theta_2}{r_2}\right)\cos\theta_2\right); \end{aligned} \tag{23}$$

so, the Hamiltonian (B.18) becomes

$$\begin{aligned} H = \frac{1}{2M}\left\{R_1^2 + \left(\frac{\Theta_1^2}{r_1^2}\right)\right\} + \frac{1}{2M_2}\left\{R_2^2 + \left(\frac{\Theta_2^2}{r_2^2}\right)\right\} - \frac{m_0 m_1}{r_1} \\ - \frac{m_0 m_2}{\sqrt{r_2^2 + \alpha_0^2 r_1^2 - 2\alpha_0 r_1 r_2 \cos(\theta_2 - \theta_1)}} \\ - \frac{m_1 m_2}{\sqrt{r_2^2 + \alpha_1^2 r_1^2 - 2\alpha_1 r_1 r_2 \cos(\theta_2 - \theta_1)}}. \end{aligned} \tag{24}$$

The constants are the same as in (B.17). Note that the Hamiltonian only depends on the difference of the polar angles, $\theta_2 - \theta_1$.

D. Symplectic Transformations with Multipliers and Scaling

If instead of satisfying (A.2) a transformation $\zeta = \Xi(t, z)$ satisfies

$$J = \mu \frac{\partial \Xi^T}{\partial z} J \frac{\partial \Xi}{\partial z}, \tag{1}$$

where μ is some nonzero constant, then $\zeta = \Xi(t, z)$ is called a *symplectic transformation (map, change of variables, etc.) with multiplier* μ. Equations (A.3) become

$$\dot{\zeta} = \mu J \nabla_\zeta H(t, \zeta) + J \nabla_\zeta R(t, \zeta), \tag{2}$$

where all the symbols have the same meaning as in Section IV.A. In the time-independent case, you simply multiply the Hamiltonian by μ. Let us look at some examples and applications.

1. Universal Gravitational Constant

When the N-body problem was introduced in Section I.C, the equations contained the universal gravitational constant G. Later we set $G = 1$. This can be accomplished by a symplectic change of variables with multiplier. The change of distance only. A better way to make the universal gravitational constant unity is to change the unit of mass. This scaling is simply given as an example.

$$H = \sum_{i=1}^{N} \frac{\|p_i'\|^2}{2m_i} - \sum_{1 \le i < j \le N} \frac{G}{\alpha^3} \frac{m_i m_j}{\|q_i' - q_j'\|}. \tag{3}$$

If we take $\alpha^3 = G$, then in the prime coordinates the gravitational constant will be 1. q has the dimensions of distance, p has the dimensions of distance-mass/time; and so the change of variables can be done by changing the units of distance only. A better way to make the universal gravitational constant unity is to change the unit of mass. This scaling is simply given as an example.

2. Equations Near an Equilibrium Point

Consider a Hamiltonian which has a critical point at the origin; so,

$$H(z) = \tfrac{1}{2} z^T S z + K(z), \tag{4}$$

where S is the Hessian of H at $z = 0$, and K vanishes along with its first and second partial derivatives at the origin. The change of variables $z = \varepsilon w$ is a symplectic change of variables with multiplier ε^{-2}; so, the Hamiltonian becomes

$$H(w) = \tfrac{1}{2} w^T S w + \varepsilon^{-2} K(\varepsilon w) = \tfrac{1}{2} w^T S w + O(\varepsilon). \tag{5}$$

In the above, the classical notation, $O(\varepsilon)$, of perturbation theory is used. Since K is at least third order at the origin, there is a constant C such that $|\varepsilon^{-2}K(\varepsilon w)| \leq C\varepsilon$ for w in a neighborhood of the origin and ε small, which is written $\varepsilon^{-2}K(\varepsilon w) = O(\varepsilon)$. The equations of motion become

$$\dot{w} = Aw + O(\varepsilon), \qquad A = JS. \tag{6}$$

If $\|w\|$ is about 1 and ε is small, then z is small. Thus, the above transformation is useful in studying the equations near the critical point. To the lowest order in ε the equations are linear; so, close to the critical point the linear terms are the most important terms. This is an example of what is called scaling variables, and ε is called the scale parameter. To avoid the growth of symbols, one often says: scale by $z \to \varepsilon z$ which means replace z by εz everywhere. This would have the effect of changing w back to z in (6). It must be remembered that scaling is really changing variables.

3. The Restricted 3-Body Problem

In the traditional derivation of the restricted 3-body problem, one is asked to consider the motion of a particle of infinitesimal mass moving in the plane under the influence of the gravitational attraction of two finite particles which move around each other on a circular orbit of the Kepler problem. Although this description is picturesque, it hardly clarifies the relationship between the restricted 3-body problem and the full problem. Consider the 3-body problem in rotating coordinates (B.3) with $N = 3$ and $\omega = 1$. Let the third mass be small by setting $m_3 = \varepsilon^2$ and considering ε as a small positive parameter. Making this substitution into (B.3) and rearranging terms gives

$$H_3 = \frac{\|v_3\|^2}{2\varepsilon^2} - u_3^T K v_3 - \sum_{i=1}^{2} \frac{\varepsilon^2 m_i}{\|u_i - u_3\|} + H_2. \tag{7}$$

Here H_2 is the Hamiltonian of the 2-body problem in rotating coordinates, i.e., (B.3) with $N = 2$. ε is a small parameter which measures the smallness of one mass. A small mass should make a small perturbation on the other particles; thus, we should attempt to make ε measure the deviation of the motion of the two finite particles from a circular orbit. That is, ε should measure the smallness of the mass and how close the two finite particles' orbits are to circular. To accomplish this we must prepare that Hamiltonian so that one variable represents the deviation from a circular orbit.

Let $Z = (u_1, u_2, v_1, v_2)$; so H_2 is a function of the 8-vector Z. A circular solution of the 2-body problem is a critical point of the Hamiltonian of the 2-body problem in rotating coordinates, i.e., H_2. Let $Z^* = (a_1, a_2, b_1, b_2)$ be such a critical point (later we will specify Z^*). By Taylor's theorem

$$H_2(Z) = H_2(Z^*) + \tfrac{1}{2}(Z - Z^*)^T S(Z - Z^*) + O(\|Z - Z^*\|^3), \tag{8}$$

where S is the Hessian of H_2 at Z^*. Since the equations of motion do not

depend on constants, drop the constant term in (8). If the motions of the two finite particles were nearly circular, the $Z - Z^*$ would be small; this suggests that one should change variables by $Z - Z^* = \varepsilon U$, but to make the change of variables symplectic, you must also change coordinates by $u_3 = \xi$, $v_3 = \varepsilon^2 \eta$, which gives a symplectic change of variables with multiplier ε^{-2}. The Hamiltonian becomes

$$H_3 = \left\{ \frac{\|\eta\|^2}{2} - \xi^T K \eta - \sum_{i=1}^{2} \frac{m_i}{\|\xi - a_i\|} \right\} + \frac{1}{2} U^T S U + O(\varepsilon). \tag{9}$$

The quantity in the braces in (9) is the Hamiltonian of the restricted 3-body problem if we take $m_1 = \mu$, $m_2 = 1 - \mu$, $a_1 = (1 - \mu, 0)$, and $a_2 = (\mu, 0)$. The quadratic term above in (9) is simply the linearized equations about the circular solutions of the 2-body problem in rotating coordinates. Thus, to first order in ε, the Hamiltonian of the full 3-body problem is the sum of the Hamiltonian for the restricted problem and the Hamiltonian of the linearized equations about the circular solution. So, to first order, the equations of the full 3-body problem decouples into the equations for the restricted problem and the linearized equations about the circular solution.

In Chapter VI, this scaled version of the restricted problem will be used to prove that nondegenerate periodic solutions of the restricted problem can be continued into the full 3-body problem for small mass.

E. Delaunay and Poincaré Elements

There is an old saying in celestial mechanics that "no set of coordinates is good enough." Indeed, classical and modern literature are replete with endless coordinate changes. There are two set of coordinates which make the 2-body problem particularly simple and, thus, simplify perturbation arguments. These coordinates will only be used in examples, so the complete justification of their validity will not be given, and the reader is referred to the specialized literature for complete details. The first set of variables, the Delaunay elements, are valid for the elliptic orbits, and the second set, the Poincaré elements, are valid near the circular orbits of the 2-body problem.

1. The Delaunay Elements

Let $D \subset \mathbb{R}^4$ be the set of initial conditions (q_0, p_0) for the 2-body problem, (I.C.7) with $\mu = 1$, which give rise to elliptic solutions (circular and collision orbits excluded). This domain is known as the *elliptic domain.* It can be coordinatized by two angles and two rectilinear coordinates. For example, the semi-major axis, a, the eccentricity, ε, and the argument of the perihelion, g,

specify the ellipse, and the true anomaly, f, specifies where the particle is on the ellipse. These are not symplectic coordinates, whereas the Delaunay elements to be introduced are.

Tradition holds that the Delaunay elements are denoted ℓ, g, L, and G. The symplectic change of coordinates from the polar coordinates (r, θ, R, Θ) to (ℓ, g, L, G) is obtained from the generating function

$$W(r, \theta, L, G) = \theta G + \int_z^r \left\{ -\frac{G^2}{x^2} + \frac{2}{x} - \frac{1}{L^2} \right\}^{1/2} dx, \tag{1}$$

where $z = L\{L - (L^2 - G^2)^{1/2}\}$. The transformation is given implicitly by the formulas

$$\begin{aligned} \ell &= \frac{\partial W}{\partial L}, \qquad R = \frac{\partial W}{\partial r} = \left\{ -\frac{G^2}{r^2} + \frac{2}{r} - \frac{1}{L^2} \right\}^{1/2}, \\ g &= \frac{\partial W}{\partial G}, \qquad \Theta = \frac{\partial W}{\partial \theta} = G. \end{aligned} \tag{2}$$

Since $\Theta = G$, G is angular momentum. From the formula for R it follows that the Hamiltonian of Kepler's problem (C.17) becomes

$$H = \frac{1}{2}(X^2 + Y^2) - \frac{1}{(x^2 + y^2)} = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2} \right) - \frac{1}{r} = -\frac{1}{2L^2}. \tag{3}$$

So the Hamiltonian of Kepler's problem depends on only one variable, the unnamed variable L. The equations of motion become

$$\begin{aligned} \dot{\ell} &= \frac{1}{L^3}, \qquad \dot{L} = 0, \\ \dot{g} &= 0, \qquad \dot{G} = 0; \end{aligned} \tag{4}$$

so, L, g, and G are integrals of the motion, and ℓ is swept out uniformly.

The Hamiltonian of Kepler's problem in rotating coordinates becomes

$$\begin{aligned} H &= \frac{1}{2}(X^2 + Y^2) - (xY - yX) - \frac{1}{(x^2 + y^2)} = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2} \right) - \Theta - \frac{1}{r} \\ &= -\frac{1}{2L^2} - G, \end{aligned} \tag{5}$$

and the equations of motion become

$$\begin{aligned} \dot{\ell} &= \frac{1}{L^3}, \qquad \dot{L} = 0, \\ \dot{g} &= -1, \qquad \dot{G} = 0. \end{aligned} \tag{6}$$

After considerable effort, it can be shown that (ℓ, g, L, G) define a valid set of coordinates on D and that both ℓ and g are angular variables defined

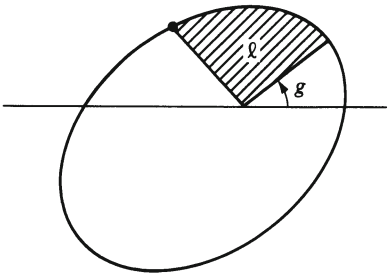

Figure D.1. Delaunay angles.

modulo 2π. In fact, it can be shown that z is the perihelion on the elliptic orbit and that g is the *argument of the perihelion.* The angle ℓ is new and is known as the *mean anomaly.* It is measured from the perihelion. It is the ratio of the area swept out by the ray from the origin to the particle from its passage by the perihelion to the total area (see Figure D.1). The variable L is unnamed in the classical literature and is related to the semi-major axis, a, by $L = a^{1/2}$.

2. Poincaré Elements

The argument of the perihelion is clearly undefined for circular orbits; so, Delaunay elements are not valid coordinates in a neighborhood of the circular orbits. To overcome this problem Poincarè introduced what he called Kepler variables but which have become known as *Poincaré elements.* Make the symplectic change of variables from the Delaunay variables (ℓ, g, L, G) to the Poincaré variables (Q_1, Q_2, P_1, P_2) by

$$\begin{aligned} Q_1 &= \ell + g, & Q_2 &= [2(L - G)]^{1/2} \cos \ell, \\ P_1 &= L, & P_2 &= [2(L - G)]^{1/2} \sin \ell. \end{aligned} \tag{7}$$

The Hamiltonian of the Kepler problem (I.C.7) becomes

$$H = \frac{1}{2}(X^2 + Y^2) - \frac{1}{(x^2 + y^2)} = -\frac{1}{2P_1^2}, \tag{8}$$

and the Hamiltonian of the Kepler problem in rotating coordinates becomes

$$H = \frac{1}{2}(X^2 + Y^2) - (xY - yX) - \frac{1}{(x^2 + y^2)} = -\frac{1}{2P_1^2} - P_1 + \frac{1}{2}(Q_2^2 + P_2^2). \tag{9}$$

Q_1 is an angular coordinate defined modulo 2π, and the remaining coordinates Q_2, P_1, and P_2 are rectangular variables. $Q_2 = P_2 = 0$ correspond to the circular orbits of the 2-body problem. Even though these new coordinates are defined from the Delaunay elements, which are not defined on the circular

orbits, it can be shown that these are valid coordinates in a neighborhood of the direct circular orbits. There is a similar set for the retrograde orbits.

F. Further Reading

Almost any advanced book on mechanics has some material on symplectic (or canonical) transformations, but their presentation may be a little slipshod for a mathematician. Pollard (1966) has a very short, clean presentation that is well worth reading. Arnold (1978) and Siegel and Moser (1971) have most of the general material presented here in one form or another. The most detailed presentation is found in Wintner (1944). As we have said before, Wintner is not easy to read.

Szebehely (1967) has a nice presentation of Delaunay and Poincaré elements. The examples from Section D were taken from the survey article by Meyer (1984a). More examples and applications are given in the paper.

Problems

1. Show that if you scale time by $t \to \mu t$, then you should scale the Hamiltonian by $H \to \mu^{-1} H$.

2. Scale the Hamiltonian on the N-body prolem in rotating coordinates (III.A.11) so that ω is 1.

3. Consider the restricted 3-body problem (I.C.9). To investigate solutions near ∞, scale by $x \to \varepsilon^{-2} x$, $y \to \varepsilon y$. Show that the Hamiltonian becomes $H = -x^T K y + \varepsilon^3 \{ \|y\|^2/2 - 1/\|x\| \} + O(\varepsilon^2)$. Justify this result on physical grounds.

4. Consider the restricted 3-body problem (I.C.9). To investigate solutions near one of the primaries first shift the origin to one primary by $x \to x - (1 - \mu, 0)$. Then scale by $x \to \varepsilon^2 x$, $y \to \varepsilon^{-1} y$, $t \to \varepsilon^3 t$.

5. In Section C, the equations for the Kepler problem were written in polar coordinates (see C.18). Since angular momentum, Θ, is a constant, set $\Theta = c$ and investigate the equation for r, $\ddot{r} = \dot{R} = -c^2/r^3 + \mu/r^2$, using the geometric methods discussed in Section I.B.2.

6. Write the functions r^{2k}, $r^{2k+5} \cos 5\theta$, and $r^{2k+5} \sin 5\theta$ in rectangular coordinates. Sketch the level curves of $r^2 + r^5 \cos 5\theta$.

7. Use the notation used to discuss the crystal model in Section E. Let $T(x, y) = (X(x, y), Y(x, y))$. Show that $\partial X(x, y)/\partial y > 0$. This implies that T is a monotone twist map as defined and discussed in Chapter VII.

8. Give an example of a linear symplectic transformation which is not given by a generating function as given in Theorem C.2.

9. Let T: $(q, p) \to (Q, P) = (Q(q, p), P(q, p))$ be a symplectic transformations defined on an open set in $\mathbb{R}^2$.

a. Show that $\omega = (Q - q)d(P + p) - (P - p)d(Q + q)$ is closed.

b. Assume that domain of definition of Q and P is such that Poincaré's lemma applies, so $\omega = dS$. Assume also that $\xi = P + p$, $\eta = Q + q$ is a valid change of coordinates (not necessary symplectic). Show that the critical points of S are fixed points of transformation T: $(q, p) \to (Q, P)$.

c. Let $S = q^2/2 + \mu p + p^3/3$ where μ is a parameter. Find the critical points of S as μ varies. Compute the map T corresponding to this S. What can you say about T's fixed points as μ varies.

CHAPTER V

Introduction to the Geometric Theory of Hamiltonian Dynamical Systems

This chapter gives an introduction to the geometric theory of autonomous Hamiltonian systems by studying some local questions about the nature of the solutions in a neighborhood of a point or a periodic solution. The dependences of periodic solutions on parameters is also presented in the case when no drastic changes occur, i.e., when there are no bifurcations. Bifurcations are addressed in Chapter VIII. Several applications to the 3-body problem are given. The chapter ends with a brief introduction to hyperbolic objects and homoclinic phenomena.

The geometric theory of Hamiltonian systems is vast and far from complete. Some of the basic definitions and results from the theory of dynamical systems are given to put the topic in context. In most cases, the background theory for ordinary (non-Hamiltonian) equations is given first. Since the non-Hamiltonian theory is fairly well documented in the literature, the more lengthy proofs will be given by referral.

A. Introduction to Dynamical Systems

Consider an autonomous system of ordinary differential equations of the form

$$\dot{x} = f(x), \tag{1}$$

where $f: \mathbb{O} \to \mathbb{R}^m$ is smooth and $\mathbb{O}$ is an open set in $\mathbb{R}^m$. Let $\psi(t)$ be a solution of (1) defined for $t \in (\alpha, \omega)$. A geometric representation of a solution (for a nonautonomous as well as an autonomous system) is the graph of ψ, $\{(t, \psi(t)): t \in (\alpha, \omega)\}$, in $\mathbb{O} \times (\alpha, \omega) \subset \mathbb{R}^{m+1}$, position-time space. See Figure A.1. The fundamental existence and uniqueness theorem for differential equa-

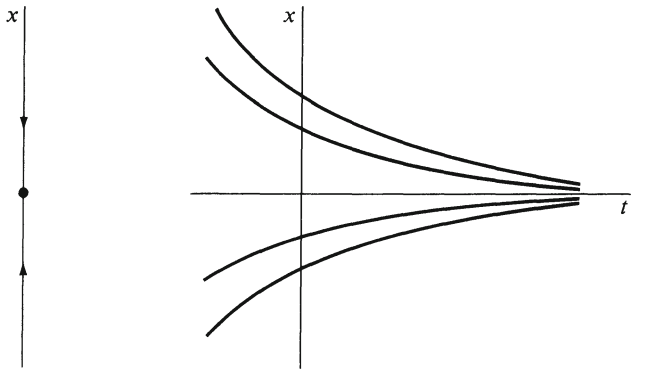

Figure A.1. Solutions and orbits of $\dot{x} = -x$.

tions asserts that there is one and only one solution through a point $\xi \in \mathbb{O}$ when $t = t_0$; so, there is one and only one graph of a solution through a point $(\xi, t_0) \in \mathbb{O} \times \mathbb{R}$.

Since (1) is independent of t, any translate of a solution, $\psi(t - \tau)$, is a solution also. (There is no clock for an autonomous equation and so no initial epoch.) If one thinks of ψ as a curve in $\mathbb{O} \subset \mathbb{R}^m$, then all translates of the solution ψ give the same curve in $\mathbb{R}^m$. The parameterized curve $\psi(t)$ in $\mathbb{R}^n$ is called a *trajectory*, and the oriented but unparameterized curve $\psi(t)$ is called an *orbit*. An orbit is the set $\{\psi(t)\colon t \in (\alpha, \omega)\}$ with the orientation coming from the orientation of (α, ω) in $\mathbb{R}$, and a trajectory is the map $\psi\colon (\alpha, \omega) \to \mathbb{R}$. In dynamical systems, the geometry of the trajectories (orbits) in $\mathbb{R}^m$ is the object of study.

If $\psi_i(t)$, $i = 1, 2$, are two solutions with $\psi_1(t_1) = \psi_2(t_2)$ then $\chi(t) = \psi_2(t - t_1 + t_2)$ is also a solution of (1) with $\chi(t_1) = \psi_2(t_2) = \psi_1(t_1)$; so, by the uniqueness theorem for differential equations, $\chi(t) = \psi_2(t - t_1 + t_2) \equiv \psi_1(t)$. So if two solutions meet, they are simply time translates of each other and are the same orbit in $\mathbb{O}$. Thus, orbits never cross in $\mathbb{O}$.

Let $\phi(t, \xi)$ denote the general solution of (1), that is, the maximal solution of (1) which satisfies $\phi(0, \xi) = \xi$ for $\xi \in \mathbb{O}$.

Lemma 1. *If t and τ are such that $\phi(\tau, \xi)$ and $\phi(t + \tau, \xi)$ are defined, then*

$$\phi(t, \phi(\tau, \xi)) = \phi(t + \tau, \xi). \tag{2}$$

PROOF. Both sides of (2) are solutions of (1) and are equal to $\phi(\tau, \xi)$ when $t = 0$. Thus, by the uniqueness theorem for differential equations, they are equal where defined. ∎

Lemma 2. *Let $\xi_0 \in \mathbb{O}$ be a equilibrium point; so, $f(\xi_0) = 0$. If $\phi(t, \xi') \to \xi_0$ as $t \to t' +$ (respectively $t \to t'-$) and $\xi' \neq \xi_0$, then $t' = +\infty$ (respectively $t' = -\infty$.) It takes on infinite amount of time to come to rest!*

PROOF. Assume not; so t' is finite. Then $\eta(t) \equiv \xi_0$ is one solution through ξ_0 at time t' and so is $\psi(t)$, where $\psi(t) = \phi(t, \xi')$ for $t < t'$ and $\psi(t) = \xi_0$ for $t \geq t'$. But this contradicts the uniqueness theorem for differential equations since $\xi' \neq \xi_0$. ∎

Let $g: \mathbb{O} \to \mathbb{R}$ be smooth and positive; then a reparameterization of the solutions of (1) is defined by

$$dt = g(x)\, d\tau, \tag{3}$$

and if a prime denotes $d/d\tau$, then Equations (1) become

$$x' = f(x)g(x). \tag{4}$$

The solution curves of (1) and (3) are the same, only their parameterizations are different.

Lemma 3. *There exists a reparameterization of* (1) *such that all solutions are defined for all* t.

PROOF. We shall only prove this theorem when $\mathbb{O} = \mathbb{R}^m$. Let $g(x) = 1/(1 + \|f(x)\|)$. The equation $x' = f(x)g(x) = h(x)$ satisfies $\|h(x)\| \leq 1$ for all x. By the discussion in I.A.2, a solution $\psi(t)$ is either defined for all time or tends to ∞ in finite time. But $\|\dot{\psi}(t)\| \leq 1$ implies $\|\psi(t)\| \leq \|\psi(0)\| + t$; so, ψ must be defined for all t. ∎

In the general case when $\mathbb{O} \neq \mathbb{R}^m$, one can construct a smooth function $g: \mathbb{O} \to \mathbb{R}$ such that $g(x) \to 0$ and $f(x)g(x) \to \mathbb{O}$ as $x \to \partial\mathbb{O}$ where $\partial\mathbb{O}$ is the boundary of $\mathbb{O}$. By the above argument, the solutions of $x' = h(x) = f(x)g(x)$ will be defined for all t. By defining $h(x) = 0$ for $x \notin \mathbb{O}$, the equations and solutions would be defined for all $x \in \mathbb{R}^m$ also.

Assume that the function $f(x)$ in (1) is defined and smooth on all of $\mathbb{O}$ and that all the solutions of (1) are defined for all $t \in \mathbb{R}$. By the discussion given above, these assumptions are always valid after a reparameterization, so these assumptions do not limit the discussion of the geometry of the orbits. Let $\phi(t, \xi)$ be the solution of (1) which satisfies $\phi(0, \xi) = \xi$. By the definition and Lemma 1, the family $\{\phi_t\}$ satisfies

$$\begin{aligned} \phi_0 &= \text{id} = \text{the identity map on } O, \\ \phi_t \circ \phi_\tau &= \phi_{t+\tau}. \end{aligned} \tag{5}$$

This implies ϕ_t has an inverse ϕ_{-t} and so ϕ_t is a homeomorphism for all t. Any family of smooth mappings satisfying (5) defines a *dynamical system* or a *flow* on $\mathbb{O}$. If (1) is a Hamiltonian system of equations, then ϕ_t is symplectic for all t by Theorem IV.A.2. In this case, the family of smooth maps ϕ_t defines a *Hamiltonian dynamical system* or a *Hamiltonian flow*. Sometimes the name dynamical system is used even if the solutions are not defined for all t.

A trajectory $\phi_t(\xi^0) = \phi(t, \xi_0)$ is periodic if there is a $T \neq 0$ such that $\phi(t + T, \xi_0) = \phi(t, \xi_0)$ for all $t \in \mathbb{R}$. The number T is called *a period*, and the least positive T is called *the period.*

Two dynamical systems ϕ_t: $\mathcal{O} \to \mathcal{O}$ and ψ_t: $\mathcal{Q} \to \mathcal{Q}$ are (*topologically*) *equivalent* if there is a homeomorphism h: $\mathcal{O} \to \mathcal{Q}$ which carries orbits of ϕ_t onto orbits of ψ_t and vice versa. Usually it is required to preserve the sense or orientation of the orbits also. Thus, two dynamical systems are equivalent if the geometry of their orbits is the same, but the timing may not be the same. The homeomorphism will take equilibrium points to equilibrium points and periodic orbits to periodic orbits. The dynamical systems defined by the two harmonic oscillators $\dot{x} = \omega_i y$, $\dot{y} = -\omega_i x$, $i = 1, 2$, are equivalent, since the identity map takes orbits to orbits, but since their periods are, in general, unequal, they would not be equivalent if the parameterization were required to be preserved.

Lemma 4. *$\phi_t(\xi_0)$ is periodic with period T if and only if $\phi_T(\xi_0) = \xi_0$.*

PROOF. If $\phi_t(\xi_0)$ is periodic, then set $t = 0$ in $\phi(t + T, \xi_0) = \phi(t, \xi_0)$ to get $\phi(T, \xi_0) = \phi(0, \xi_0) = \xi_0$. If $\phi_T(\xi_0) = \xi_0$, then apply ϕ_t to both sides and apply (5) to get $\phi(t + T, \xi_0) = \phi_t \circ \phi_T(\xi_0) = \phi_t(\xi_0) = \phi(t, \xi_0)$. ∎

An *invariant set* is a subset $Q \subset \mathcal{O}$ such that if $\xi \in Q$, then $\phi(t, \xi) \in Q$ for all t. That is an invariant set is a union of orbits.

A linear equation $\dot{x} = Ax$, A a constant $m \times m$ matrix, defines a linear dynamical system $\psi_t(x) = e^{At}x$ on $\mathbb{R}^m$. If A is a Hamiltonian matrix, then the map is a symplectomorphism. The origin is an equilibrium point. If $x_0 = u_0 + iv_0$ is an eigenvector corresponding to a pure imaginary eigenvalue $\lambda = i\omega$, $\omega \neq 0$, then $e^{At}u$ is a $2\pi/\omega$ periodic solution. In fact, the two-dimensional real linear space span$\{u, v\}$ is filled with $2\pi/\omega$ periodic solutions.

If none of the eigenvalues of the matrix A are pure imaginary, then the matrix A is called *hyperbolic*, and the equilibrium point at the origin is called *hyperbolic* also. If all the eigenvalues of A have real parts less (greater) than zero, then $e^{At}x \to 0$ as $t \to +\infty$ (respectively as $t \to -\infty$) for all x. Neither of these cases happens for a Hamiltonian matrix A because the eigenvalues of a Hamiltonian matrix are symmetric with respect to the imaginary axis. If A has k eigenvalues with negative real parts and $m - k$ eigenvalues with positive real parts, then by the Jordan canonical form theorem there is a nonsingular $m \times m$ matrix P such that $P^{-1}AP = A' = \text{diag}(B, C)$, where B is a $k \times k$ matrix with eigenvalues with negative real parts and C is an $(m - k) \times (m - k)$ matrix with eigenvalues with positive real parts. The matrix P can be thought of as the matrix of a change of variables, so that in the new variables, A has the form $A' = \text{diag}(B, C)$. Thus, in this case A preserves the splitting $\mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^{m-k}$, i.e., the coordinate planes $\mathbb{R}^k \times \{0\}$ and $\{0\} \times \mathbb{R}^{m-k}$ are invariant sets. If $x \in \mathbb{R}^k \times \{0\}$, then $e^{A't} \to 0$ as $t \to +\infty$, and if $x \in \{0\} \times \mathbb{R}^{m-k}$, then

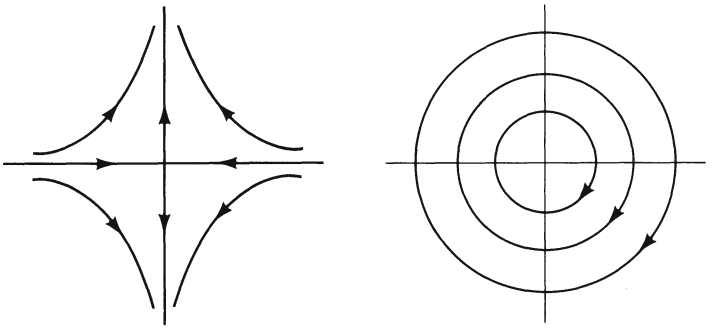

Figure A.2. Linear dynamical systems. (a) Hyperbolic; (b) elliptic.

$e^{A't}x \to 0$ as $t \to -\infty$. Figure A.2 (a) indicates the orbit structure for the hyperbolic, Hamiltonian matrix $A' = \text{diag}(-1, +1)$.

If all the eigenvalues of the matrix A are pure imaginary and A is simple (diagonalizable), then A is called *elliptic*, and the equilibrium point at the origin is called *elliptic*. By the Jordan canonical form theorem, there is a nonsingular, $m \times m$ matrix P such that $P^{-1}AP = A' = \text{diag}(\omega_1 J, \ldots, \omega_k J, 0, \ldots, 0)$ where J is the 2×2 matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Then $e^{A't} = \text{diag}(R(\omega_1 t), \ldots, R(\omega_k t), 1, \ldots, 1)$, where $R(\theta)$ is the rotation matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Figure A.2(b) indicates the orbit structure for the elliptic, symplectic matrix $R(\theta)$.

If p is an equilibrium point for the nonlinear equation (1), then the equation $\dot{x} = Ax$, where $A = Df(p) = \partial f(p)/\partial x$, is called the *linearization of* (1) at p. The equilibrium point p is called *hyperbolic* or *elliptic* as matrix A is called.

A famous theorem of Hartman (1964) says that the flow near a hyperbolic equilibrium point is equivalent to a linear flow. That is, if p is a hyperbolic equilibrium for (1), then there are neighborhoods $\mathbb{O}$ of p and $\mathbb{Q}$ of the $0 \in \mathbb{R}^m$ and a homeomorphism h: $\mathbb{O} \to \mathbb{Q}$ such that h maps orbits of (1) onto orbits of $\dot{x} = Ax$. No such theorem is true for elliptic equilibrium points.

B. Discrete Dynamical Systems

Closely related to differential equations are diffeomorphisms which define discrete dynamical systems. Since discrete dynamical systems are first introduced in this section, several examples will be given.

1. Diffeomorphisms and Symplectomorphisms

A map $\psi\colon \mathcal{O} \to \mathbb{R}^m$, $\mathcal{O}$ open in $\mathbb{R}^m$, is a *diffeomorphism* if ψ is differentiable and has a differentiable inverse. In particular, a diffeomorphism is a homeomorphism of $\mathcal{O}$ onto $\psi(\mathcal{O})$. In many cases it is required that ψ take $\mathcal{O}$ into $\mathcal{O}$, $\psi(\mathcal{O}) = \mathcal{O}$, in which case ψ is said to be a *diffeomorphism* of $\mathcal{O}$. Let k be a positive integer, and let $\psi^k = \psi \circ \psi \circ \cdots \circ \psi$, k times, be the kth composition of ψ with itself. So $\psi^1 = \psi$. Define $\psi^0 = \mathrm{id}$, the identity map $[\mathrm{id}(x) = x]$ and $\psi^{-k} = \psi^{-1} \circ \psi^{-1} \circ \cdots \circ \psi^{-1}$, k times, be the kth composition of ψ^{-1}, the inverse of ψ. If ψ is a diffeomorphism of $\mathcal{O}$, then ψ^k is defined for all k and is a diffeomorphism of $\mathcal{O}$ for all k. In general, ψ^k may be defined for some k and on only a part of $\mathcal{O}$. In either case, it is easy to verify that $\psi^{k+s} = \psi^k \circ \psi^s$ whenever the two sides are defined. If ψ is a symplectic diffeomorphism, then ψ is called a *symplectomorphism.*

A *discrete dynamical system* is simply a diffeomorphism ψ of a set $\mathcal{O}$. A *discrete Hamiltonian dynamical system* is simply a symplectomorphism of a set $\mathcal{O}$. If we let $\mathbb{Z}$ be the integers and $\Psi(k, \xi) = \psi^k(\xi)$, then $\Psi\colon \mathbb{Z} \times \mathbb{R}^m \to \mathbb{R}^m$ is analogous to the general solution of a differential equation. In fact, $\Psi(k, \xi)$ is the general solution of the difference equation $x(k+1) = \psi(x(k))$, $x(0) = \xi$.

The set $\{\psi^n(p)\colon -\infty < n < +\infty\}$ is called the *orbit of the point p.* A point $p \in \mathcal{O}$ such that $\psi(p) = p$ is called *a fixed point (of ψ)*, and a point $p \in \mathcal{O}$ such that $\psi^k(p) = p$, for some positive integer k, is called *a periodic point (of ψ)*, and k is call *a period.* The least positive integer k such that $\psi^k(p) = p$ is called *the period.*

Two discrete dynamical systems $\phi\colon \mathcal{O} \to \mathcal{O}$ and $\psi\colon \mathcal{Q} \to \mathcal{Q}$ are *(topologically) equivalent* if there is a homeomorphism $h\colon \mathcal{O} \to \mathcal{Q}$ which carries orbits of ϕ onto orbits of ψ and vice versa. This is the same as $h \circ \phi = \psi \circ h$.

A nonsingular linear map $x \to Ax$, A a constant $m \times m$ matrix, defines a discrete dynamical system $\psi(x) = Ax$ on $\mathbb{R}^m$. If A is a symplectic matrix, then the map is a symplectomorphism. The origin is a fixed point. If x_0 is an eigenvector corresponding to an eigenvalue λ which is a kth root of unity, $\lambda^k = 1$, then x_0 is a periodic point of period k because $A^k x_0 = \lambda^k x_0 = x_0$.

If none of the eigenvalues of the matrix A have modulus 1, then the matrix A is called *hyperbolic*, and the fixed point at the origin is called *hyperbolic* also. If all the eigenvalues of A have modulus less (respectively greater) than 1, then $A^n x \to 0$ as $n \to +\infty$ (respectively as $n \to -\infty$) for all x. Neither of these cases happens for a symplectic matrix A. If A has k eigenvalues with modulus less than 1 and $m - k$ eigenvalues with modulus greater than 1, then by the Jordan canonical form theorem, there is a nonsingular, $m \times m$ matrix P such that $P^{-1}AP = A = \mathrm{diag}(B, C)$, where B is a $k \times k$ matrix with eigenvalues of modules less than 1, and C is an $(m-k) \times (m-k)$ matrix with eigenvalues of modules greater than 1. The matrix P can be thought of as the matrix of a change of variables, so that in the new variables, A has the form $A = \mathrm{diag}(B, C)$. Thus, in this case, A preserves the splitting $\mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^{m-k}$, i.e., A

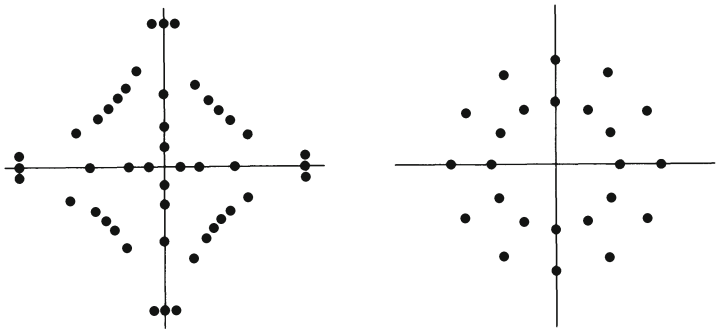

Figure B.1. Linear dynamical systems. (a) Hyperbolic; (b) elliptic.

carries the coordinate plane $\mathbb{R}^k \times \{0\}$ into itself and the coordinate plane $\{0\} \times \mathbb{R}^{m-k}$ into itself. If $x \in \mathbb{R}^k \times \{0\}$, then $A^n x \to 0$ as $n \to +\infty$ and if $x \in (0) \times \mathbb{R}^{m-k}$, then $A^n x \to 0$ as $n \to -\infty$. Figure B.1(a) indicates the orbit structure for the hyperbolic, symplectic matrix $A = \text{diag}(\frac{1}{2}, 2)$.

If all the eigenvalues of the matrix A have modulus 1 and A is diagonalizable, then A is called *elliptic* and the fixed point at the origin is called *elliptic*. By the Jordan canonical form theorem, there is a nonsingular, $m \times m$ matrix P such that $P^{-1}AP = A = \text{diag}(R(\theta_1), \ldots, R(\theta_k), \pm 1, \ldots, \pm 1)$ where $R(\theta)$ is the rotation matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

A is the direct sum of rotations in coordinate planes and reflections in coordinate axes. Figure B.1(b) indicates the orbit structure for the elliptic, symplectic matrix $R(\theta)$.

If ψ is a general nonlinear diffeomorphism with fixed point p (respectively periodic point p of period k), then p is call *hyperbolic* or *elliptic* as matrix $D\psi(p)$ [respectively $D\psi^k(p)$] is called.

A famous theorem of Hartman (1964) says that near a hyperbolic fixed point, a diffeomorphism is equivalent to its linear part. That is, if p is a hyperbolic point for ψ, then there are neighborhoods $\mathbb{O}$ of p and $\mathbb{Q}$ of the $0 \in \mathbb{R}^m$ and a homeomorphism h: $\mathbb{O} \to \mathbb{Q}$ such that h maps orbits of ψ onto orbits of $x \to Ax$. No such theorem is true for elliptic fixed points.

2. The Henon Map

The Henon map is the quadratic map of $\mathbb{R}^2$ into itself defined by H: $(x, y) \to (x', y')$, where

$$\begin{aligned} x' &= \alpha - y - x^2, & x &= y', \\ y' &= x, & y &= \alpha - x' - y'^2, \end{aligned} \tag{1}$$

and α is simply a parameter. The map is one-to-one and onto since its inverse is a quadratic map also. The Jacobian of this map is clearly $+1$, so the map is area preserving, and (1) defines a discrete Hamiltonian dynamical system.

The Henon map has fixed points at $(-1 \pm \sqrt{1+\alpha}, -1 \pm \sqrt{1+\alpha})$. The one at $(-1+\sqrt{1+\alpha}, -1+\sqrt{1-\alpha})$ is hyperbolic for all α, and the one at $(-1-\sqrt{1+\alpha}, -1-\sqrt{1+\alpha})$ is elliptic for $-1 < \alpha \leq 3$ and hyperbolic otherwise. The Henon map has been extensively studied by computer simulation.

3. The Time τ Map

If $\phi(t, \xi)$ is the general solution of the autonomous differential equation $\dot{x} = f(x)$, then for a fixed τ the map $\psi: \xi \to \phi(\tau, \xi)$ is a diffeomorphism since its inverse is $\psi^{-1}: \xi \to \phi(-\tau, \xi)$. It is called the *time τ map*. If the differential equation is Hamiltonian, then ψ is a symplectomorphism.

Let *p be a fixed point of ψ*, $\psi(p) = \phi(\tau, p) = p$. Then by Lemma A.4, p is an initial condition for a periodic solution of period τ. In a like manner a periodic point of period k is an initial condition for a periodic solution of period $k\tau$. This example is somewhat artificial because the choice of τ was arbitrary. There is no clock in an autonomous system.

The harmonic oscillator $\dot{q} = \partial H/\partial p = \omega p$, $\dot{p} = -\partial H/\partial q = -\omega q$, $H = (\omega/2)(q^2 + p^2)$ defines the discrete Hamiltonian system

$$\begin{pmatrix} q \\ p \end{pmatrix} \to \begin{pmatrix} \cos \omega\tau & \sin \omega\tau \\ -\sin \omega\tau & \cos \omega\tau \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}, \tag{2}$$

a rotation of the plane by an angle $\omega\tau$. The origin is an elliptic fixed point for this system.

4. The Period Map

Consider a periodic differential equation

$$\dot{x} = f(t, x), \qquad f(t + \tau, x) \equiv f(t, x), \quad \tau > 0. \tag{3}$$

Let $\phi(t, \xi)$ be the general solution; so, $\phi(0, \xi) = \xi$. The mapping $\psi: \xi \to \phi(\tau, \xi)$ is called the *period map* (*sometimes the Poincaré map*). If Equation (1) is defined for all $x \in \mathbb{R}^m$ and the solutions for all t, $0 \leq t \leq \tau$, then ψ defines a discrete dynamical system; and if the equation is Hamiltonian, then ψ defines a discrete Hamiltonian system. By the same argument as above, a fixed point of ψ is the initial condition of a periodic solution of period τ, and a periodic point of period k is the initial condition of a periodic solution of period $k\tau$.

A natural qestion to ask is whether all diffeomorphisms and symplectomorphisms are time τ-maps of autonomous equations or period maps of periodic systems. Later we shall show that time τ-maps are much simpler than general diffeomorphisms but that period maps are essentially the same as diffeomorphisms.

A diffeomorphism $\psi\colon \mathbb{R}^m \to \mathbb{R}^m$ is *isotopic to the identity through diffeomorphisms* if there exists a smooth function $\Phi\colon [0, \tau] \times \mathbb{R}^m \to \mathbb{R}^m$ such that for each $t \in [0, \tau]$ the map $\Phi(t, \cdot)\colon \mathbb{R}^m \to \mathbb{R}^m$ is a diffeomorphism and $\Phi(0, \xi) \equiv \xi$ and $\Phi(\tau, \xi) \equiv \psi(\xi)$ for all $\xi \in \mathbb{R}^m$. There is a similar definition where "diffeomorphism" and "m" are replaced by "symplectomorphism" and "$2n$" throughout. The period map $\psi(\xi) = \phi(\tau, \xi)$ of a periodic system is clearly isotopic to the identity through diffeomorphisms. In fact:

Theorem 1. *A necessary and sufficient condition for a diffeomorphism $\psi\colon \mathbb{R}^m \to \mathbb{R}^m$ to be isotopic to the identity through diffeomorphisms is that ψ is the period map of a periodic system of the form* (3). *Also if ψ is isotopic to the identity through symplectomorphisms, then Equation* (3) *is Hamiltonian.*

PROOF. First a little trickery with smooth functions. The function α defined by $\alpha(t) = 0$ for $t \leq 0$ and $\alpha(t) = \exp(-1/t)$ for $t > 0$ is a smooth function. (It is an easy argument to show that the right and left derivatives of α are zero at $t = 0$ by l'Hôpital's rule.) The function $\beta(t) = \alpha(t - \tau/3)/(\alpha(t - \tau/3) + \alpha(2\tau/3 - t))$ is smooth, and $\beta(t) \equiv 0$ for $t \leq \tau/3$, and $\beta(t) \equiv 1$ for $t \geq 2\tau/3$.

Let $\Phi\colon [0, \tau] \times \mathbb{R}^m \to \mathbb{R}^m$ be the isotopy. Define $\Xi(t, \xi) = \Phi(\beta(t), \xi)$; so, $\Xi(t, \cdot)$ is the identity map for $0 \leq t \leq \tau/3$ and is the diffeomorphism ψ for $2\tau/3 \leq t \leq \tau$. Let $X(t, \eta)$ be the inverse of $\Xi(t, \xi)$; so, $X(t, \Xi(t, \xi)) \equiv \xi$. Now

$$\frac{\partial \Xi}{\partial t}(t, \xi) = \frac{\partial \Xi}{\partial t}(t, X(t, \Xi(t, \xi))) = F(t, \Xi(t, \xi)),$$

where

$$F(t, x) = \frac{\partial \Xi}{\partial t}(t, X(t, x)).$$

So $\Xi(t, \xi)$ is the general solution of $\dot{x} = F(t, x)$. Since Ξ is constant in t for $0 \leq t \leq \tau/3$ and $2\tau/3 \leq t \leq \tau$, F is identically zero for $0 \leq t \leq \tau/3$ and $2\tau/3 \leq t \leq \tau$. Therefore, the τ-periodic extension of F is smooth. Thus, Ξ is the general solution of a τ-periodic system, and ψ is a period map since $\psi(\xi) = \Xi(\tau, \xi)$.

If Φ is symplectic, then F is Hamiltonian by Theorem IV.A.2. ∎

For example, let $\psi(x) = Ax + g(x)$, where $g(0) = \partial g(0)/\partial x = 0$; so the origin is a fixed point. If A has a logarithm, so $A = \exp B$, then $\Phi(t, x) = \exp(Bt) + tg(x)$ is an isotopy through diffeomorphisms near the origin.

For a symplectic map you must be a little more careful. First, if $\psi'(x) = x + g'(x)$, where $g'(0) = \partial g'(0)/\partial x = 0$, then by Theorem IV.C.2, $\psi\colon (q, p) \to (Q, P)$, where $q = \partial S(p, Q)/\partial p$, $P = \partial S(p, Q)/\partial Q$, $S(p, Q) = p^T Q + s(p, Q)$,

where s is second order at the origin. Then $S(t, p, Q) = p^T Q + ts(p, Q)$ generates an isotopy to the identity through symplectomorphisms for ψ, $\Phi'(t, x)$ [i.e., $\Phi(1, \cdot) = \psi$ and $\Phi(0, \cdot) =$ identity]. Now if $\psi(x) = Ax + g(x)$, write $\psi(x) = A(x + g'(x))$, and an isotopy for ψ is $\Phi(t, x) = \exp(Bt)\Phi'(t, x)$, where B is a Hamiltonian logarithm of A.

5. The Convex Billiards Table

Let Γ be a smooth, closed, convex curve in the plane. Imagine a point moving in the interior of Γ like a billiard ball on a table with boundary Γ. In the interior, the point moves in a straight line and is reflected off the boundary by Snell's law of reflection—the angle of incidence is equal to the angle of reflection. Let the curve Γ be parameterized by arc length, s, measured from a fixed point on the curve in the counterclockwise direction. Since the curve is closed, the parameter s can be considered as an angle. A contact of the moving point (the billiard ball) with the boundary curve Γ can be coordinatized by s, the point of contact, and α, the angle of incidence. The angle of incidence α is measured by the sign convention show in Figure B.2; so, $0 < \alpha < \pi$. Thus, the contacts are parameterized by the points (s, α) in the annulus $A = \Gamma \times (0, \pi)$. Define a map B: $A \to A$ which takes a contact (s, α) to the next contact (s', α') as shown in Figure B.3. Let ℓ be the length of the path of the moving point between successive contacts (s, α) and (s', α'). Then $d\ell = \cos \alpha \, ds - \cos \alpha' \, ds'$. Since $d^2\ell = 0$, $\sin \alpha \, d\alpha \wedge ds = \sin \alpha' \, d\alpha' \wedge ds'$ or $dc \wedge ds = dc' \wedge ds'$, where $c = \cos \alpha$ and $c' = \cos \alpha'$. Thus, if we use the arc length s and the cosine of the angle of incidence, $c = \cos \alpha$, as coordinates, the billiards map B is area preserving and defines a discrete Hamiltonian system on the annulus $\Gamma \times (-1, 1)$. A periodic point of this map corresponds to a closed path of the billiard ball.

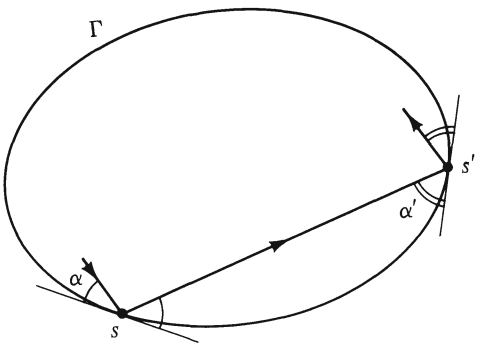

Figure B.2. The billiard table.

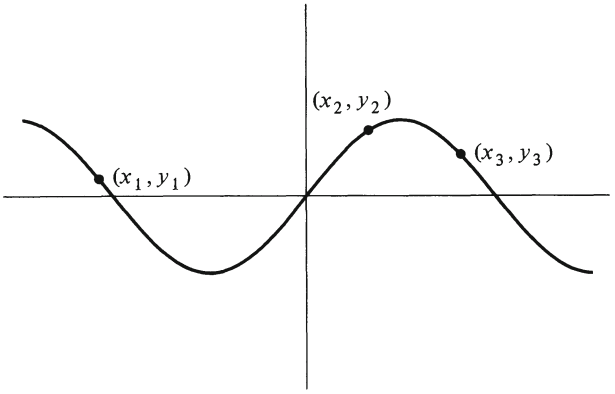

Figure B.3. Crystal model.

6. A Linear Crystal Model

The following mechanical system was suggested as a model for a one-dimensional crystal; see Nabaro (1967) for a discussion of the underlying physics. Consider an infinite wire bent into the shape of the sine curve $\{(x, y) \in \mathbb{R}^2 : y = (k/2\pi) \sin(2\pi x)\}$, where $k > 0$ is simply a parameter. The wire is placed parallel to the ground, the x axis, so that the force of gravity acts in the negative y direction as shown in Figure B.3. On this wire there are a countable number of beads (atoms) which can slide freely without friction, but each is subjected to the force of gravity and a linear attractive force to its nearest neighbors. The attractive force is not proportional to the distance between the beads, but to the projection of the distance on the x axis. The problem is to find the equilibrium states of the system.

Let (x_i, y_i), $y_i = (k/2\pi) \sin(2\pi x_i)$ be the position of the ith bead, and so the sequence $\{x_i\}_{-\infty}^{\infty}$ represents the state of the system. The physical assumptions imply that the total force on the ith bead, f_i, is

$$f_i = (x_{i-1} - x_i) + (x_{i+1} - x_i) + k \cos(2\pi x_i). \tag{4}$$

The three terms in (4) are the forces on the ith bead due to the bead on the left, the bead on the right, and the force due to gravity, in that order. At an equilibrium state $f_{i+1} = 0$ or

$$0 = (x_{i-1} - x_i) + (x_{i+1} - x_i) + k \cos(2\pi x_i). \tag{5}$$

Define the local energy or generating function by

$$h(x_i, x_{i+1}) = \frac{1}{2}(x_{i+1} - x_i)^2 + \frac{k}{2\pi} \sin(2\pi x_i) \tag{6}$$

and the total energy by the formal sum

$$H = \sum_{-\infty}^{\infty} h(x_i, x_{i+1}). \tag{7}$$

The formal condition for a critical point of H, $\nabla H = 0$, gives

$$\begin{aligned} -D_1 h(x_i, x_{i+1}) + D_2 h(x_{i-1}, x_i) &= 0, \\ -(x_{i+1} - x_i) - k\cos(2\pi x_i) + (x_i - x_{i-1}) &= 0, \end{aligned} \tag{8}$$

for all i. Here, and below, D_i, $i = 1, 2$, denotes the partial derivative with respect to the ith argument. So a formal solution of $\nabla H = 0$ is an equilibrium state.

Consider h as a generating function defining an area-preserving mapping (see Theorem IV.C.2) T: $\mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$: $(x, y) \to (x', y')$, where

$$\begin{aligned} y &= -D_1 h(x, x') \\ y' &= D_2 h(x, x'). \end{aligned} \tag{9}$$

From the form of h, it follows that if $T(x, y) = (x', y')$, then $T(x + 1, y) = (x' + 1, y')$ and vice versa. So T is well defined when the first argument is defined modulo 1, or we can consider T as a map of $\mathbb{S}^1 \times \mathbb{R}$, where $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$. It can be shown that T is one-to-one and onto.

The interesting fact about T is that a T-orbit defines an equilibrium state for the crystal model. In fact we have:

Theorem 2. *$\{x_i\}_{-\infty}^{\infty}$ is an equilibrium state, i.e., satisfies (4), if and only if $\{(x_i, y_i)\}_{-\infty}^{\infty}$ is a T-orbit.*

PROOF. $T(x_i, y_i) = (x_{i+1}, y_{i+1})$ for all i if and only if $y_i = -D_1 h(x_i, x_{i+1})$ and $y_{i+1} = D_2 h(x_i, x_{i+1})$ for all i if and only if $-D_1 h(x_i, x_{i+1}) = y_i = D_2 h(x_{i-1}, x_i)$ for all i if and only if $\{x_i\}_{-\infty}^{\infty}$ satisfies (8) or $\{x_i\}_{-\infty}^{\infty}$ is an equilibrium state. ∎

The last two examples are area-preserving mappings of the annulus $\mathbb{S}^1 \times \mathbb{R}$, where $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$. The rich theory of these maps is the topic of Chapter X.

C. The Flow Box Theorem and Local Integrals

This section investigates the local flow and local integrals near a nonequilibrium point. Consider first an ordinary differential equation

$$\dot{x} = f(x), \tag{1}$$

where f is smooth on $\mathbb{O}$, an open set in $\mathbb{R}^m$, and let $\phi(t, \xi)$ be the solution of (1) such that $\phi(0, \xi) = \xi$. The analogous results for diffeomorphism will be devel-

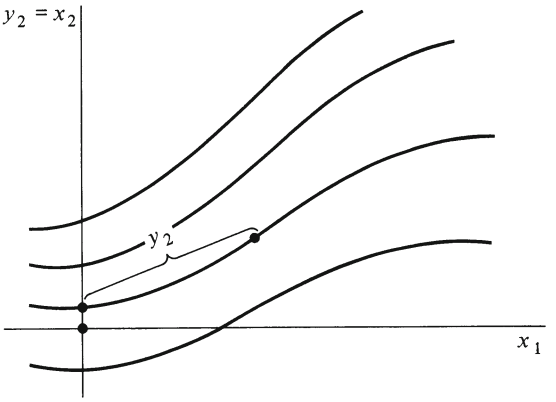

Figure C.1. The flow box.

oped in the problems. A point $r \in \mathbb{O}$ is an *ordinary point* for (1) if $f(r) \neq 0$, otherwise r is a *critical point, an equilibrium point,* or *a rest point.*

Theorem 1 (The Flow Box Theorem). *Let $r \in \mathbb{O}$ be an ordinary point for* (1), *then there exists a change of coordinates $y = \psi(x)$ defined near r such that in the new coordinates, Equations* (1) *define a parallel flow; in particular, the equations become*

$$\dot{y}_1 = 1, \qquad \dot{y}_i = 0, \quad i = 2, \ldots, m. \tag{2}$$

PROOF. Let r be the origin in $\mathbb{R}^m$. Since $f(0) = a \neq 0$, one component of a is nonzero, say $a_1 \neq 0$. The solutions cross the hyperplane $x_1 = 0$ transversely, and so the new coordinates will be the time from the crossing of this hyperplane and the $n - 1$ coordinates where the solution hits the hyperplane; see Figure C.1. That is, define the change of variables by

$$x = \phi(y_1, 0, y_2, y_3, \ldots, y_n). \tag{3}$$

Since $\phi(0, \xi) = \xi$, $\partial\phi(0, 0)/\partial x = I$, and so

$$\frac{\partial x}{\partial y}(0) = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ a_n & 0 & \cdots & 1 \end{bmatrix}, \tag{4}$$

which is nonsingular since $a_1 \neq 0$. Thus, (3) defines a valid change of coordinates near 0. The first variable, y_1, is time; so, $y_1 = 1$ and the variables $y_2, \ldots, y_n$ are initial conditions; so, $\dot{y}_2 = \dot{y}_3 = \cdots = \dot{y}_n = 0$. ∎

A set of smooth functions $F_1, \ldots, F_k$ defined on $\mathbb{O} \subset \mathbb{R}^m$ are independent at $r \in \mathbb{O}$ if the vectors $\nabla F_1(r), \ldots, \nabla F_k(r)$ are linearly independent.

Corollary 2. *Near an ordinary point, the system* (1) *has* $m-1$ *independent integrals.*

PROOF. In the y coordinates, $y_2, \ldots, y_m$ are constants of the motion and clearly independent. ∎

Assume that (1) admits an integral $F(x)$, where $F: \mathbb{O} \to \mathbb{R}$ is smooth and I is nondegenerate at $r \in \mathbb{O}$, i.e., $\nabla F(r) \neq 0$.

Theorem 3. *If* (1) *admits a nondegenerate integral* F *at* $r \in \mathbb{O}$, *where* r *is an ordinary point for* (1), *then the flow box coordinates given in Theorem* 1 *can be chosen so that* $F(y) = y_2$.

PROOF. Let y' be the coordinate system given by Theorem 1. Since F is an integral, F is independent of y_1', and since F is nondegenerate $\partial F(0)/\partial y_i \neq 0$, for some $i = 2, \ldots, m$, say for $i = 2$. Change coordinates by $y_i = y_i'$ for $i = 1, 3, 4, \ldots, m$ and $y_2 = F(y_2', \ldots, y_m')$. ∎

Consider a Hamiltonian system

$$\dot{z} = J\nabla H(z) \tag{5}$$

or

$$\dot{q} = H_p, \qquad \dot{p} = -H_q,$$

where $z = (q, p)$ and H is a smooth function defined on the open set $\mathbb{O} \subset \mathbb{R}^{2n}$.

Theorem 4 (The Hamiltonian Flow Box Theorem). *Let* $r \in \mathbb{O} \subset \mathbb{R}^{2n}$ *be an ordinary point for* (5); *then there exist symplectic coordinates* $\{y\}$ *defined near* r *such that the Hamiltonian becomes* $H(y) = y_{n+1}$, *and the equations of motion become*

$$\dot{y}_1 = \frac{\partial H}{\partial y_{n+1}} = 1, \qquad \dot{y}_i = 0 \quad \text{for } i = 2, \ldots, 2n. \tag{6}$$

PROOF. Again let r be the origin in $\mathbb{R}^{2n}$. Make a linear symplectic coordinates change so that $J\nabla H(0) = (1, 0, \ldots, 0)$—see the Problem Section. Let $q = \bar{q}(t, \xi, \eta)$, $p = \bar{p}(t, \xi, \eta)$ be the general solution of (5) with $\bar{q}(0, \xi, \eta) = \xi$ and $\bar{q}(0, \xi, \eta) = \eta$. For small values of t, these functions can be inverted to give $\xi = \bar{\xi}(t, q, p), \eta = \bar{\eta}(t, q, p)$. Since $J\nabla H(0) = (1, 0, \ldots, 0)$, we can solve the equation $\bar{\xi}_1(t, q, p) = 0$ for t to give $t = \bar{t}(q, p)$.

Define the new coordinates by

$$\begin{aligned} y_1 &= \bar{t}(q, p), & y_{n+1} &= H(q, p), \\ y_i &= \bar{\xi}(\bar{t}(q, p), q, p), & y_{n+i} &= \bar{\eta}_i(\bar{t}(q, p), q, p), \quad i = 2, \ldots, n. \end{aligned} \tag{7}$$

By Theorems IV.A.2 and IV.A.3, for fixed t, $\{\bar{\xi}_i, \bar{\xi}_j\} = \{\bar{\eta}_i, \bar{\eta}_j\} = 0$ and $\{\bar{\xi}_i, \bar{\eta}_j\} = \delta_{ij}$. Now check that (7) is symplectic. Let $2 \leq i, j \leq n$; then

$$
\begin{aligned}
\{y_i, y_j\} &= \sum_{k=1}^{n} \left(\frac{\partial y_i}{\partial q_k} \frac{\partial y_j}{\partial p_k} - \frac{\partial y_i}{\partial p_k} \frac{\partial y_j}{\partial q_k} \right) \\
&= \sum_{k=1}^{n} \left[\left(\frac{\partial \bar{\xi}_i}{\partial t} \frac{\partial \bar{t}}{\partial q_k} + \frac{\partial \bar{\xi}_i}{\partial q_k} \right) \left(\frac{\partial \bar{\xi}_j}{\partial t} \frac{\partial \bar{t}}{\partial p_k} + \frac{\partial \bar{\xi}_j}{\partial p_k} \right) \right. \\
&\quad \left. - \left(\frac{\partial \bar{\xi}_i}{\partial t} \frac{\partial \bar{t}}{\partial p_k} + \frac{\partial \bar{\xi}_i}{\partial p_k} \right) \left(\frac{\partial \bar{\xi}_j}{\partial t} \frac{\partial \bar{t}}{\partial q_k} + \frac{\partial \bar{\xi}_j}{\partial q_k} \right) \right] \qquad (8) \\
&= \{\bar{\xi}_i, \bar{\xi}_j\} + \sum_{k=1}^{n} \left[\frac{\partial \bar{\xi}_i}{\partial t} \left(\frac{\partial \bar{t}}{\partial q_k} \frac{\partial \bar{\xi}_j}{\partial p_k} - \frac{\partial \bar{t}}{\partial p_k} \frac{\partial \bar{\xi}_j}{\partial q_k} \right) + \frac{\partial \bar{\xi}_j}{\partial t} \left(\frac{\partial \bar{t}}{\partial q_k} \frac{\partial \bar{\xi}_i}{\partial p_k} - \frac{\partial \bar{t}}{\partial p_k} \frac{\partial \bar{\xi}_i}{\partial q_k} \right) \right] \\
&= \{\bar{\xi}_i, \bar{\xi}_j\} - \left[\left(\frac{\partial \bar{\xi}_i}{\partial t} \right) \{\bar{\xi}_1, \bar{\xi}_j\} - \left(\frac{\partial \bar{\xi}_j}{\partial t} \right) \{\bar{\xi}_1, \bar{\xi}_i\} \right] \Big/ \left(\frac{\partial \bar{\xi}_1}{\partial t} \right) = 0.
\end{aligned}
$$

The simplification from the second to last line to the last line comes from the identities $\partial \bar{t}/\partial q_k = -(\partial \bar{\xi}_1/\partial q_k)/(\partial \bar{\xi}_1/\partial t)$ and $\partial \bar{t}/\partial p_k = -(\partial \bar{\xi}_1/\partial p_k)/(\partial \bar{\xi}_1/\partial t)$. In a similar way, $\{y_i, y_j\} = 0$ when $n + 2 \leq i, j \leq 2n$ and $\{y_i, y_{n+j}\} = \delta_{ij}$ for $i = 1, \ldots, n$.

$\{y_1, y_{1+n}\} = \{t, H\} = 1$ since the time rate of change of t along a solution is 1. Because $\bar{\xi}_i$ and $\bar{\eta}_i$ are integrals and $y_{1+n} = H$, it follows that $\{y_i, y_{1+n}\} = 0$ for $i = 2, \ldots, 2n$.

Let $2 \leq i \leq n$. Then

$$
\begin{aligned}
\{y_1, y_i\} &= \sum_{k=1}^{n} \left(\frac{\partial \bar{t}}{\partial q_k} \frac{\partial y_i}{\partial p_k} - \frac{\partial \bar{t}}{\partial p_k} \frac{\partial y_i}{\partial q_k} \right) \\
&= \sum_{k=1}^{n} \left[\frac{\partial \bar{t}}{\partial q_k} \left(\frac{\partial \bar{\xi}_i}{\partial t} \frac{\partial \bar{t}}{\partial p_k} + \frac{\partial \bar{\xi}_i}{\partial p_k} \right) - \frac{\partial \bar{t}}{\partial p_k} \left(\frac{\partial \bar{\xi}_i}{\partial t} \frac{\partial \bar{t}}{\partial q_k} + \frac{\partial \bar{\xi}_i}{\partial q_k} \right) \right] \qquad (9) \\
&= \sum_{k=1}^{n} \left[\frac{\partial \bar{t}}{\partial q_k} \frac{\partial \bar{\xi}_i}{\partial p_k} - \frac{\partial \bar{t}}{\partial p_k} \frac{\partial \bar{\xi}_i}{\partial q_k} \right] = \frac{\{\bar{\xi}_1, \bar{\xi}_i\}}{\partial \bar{\xi}_1 / \partial t} = 0.
\end{aligned}
$$

In a similar manner $\{y_1, y_i\} = 0$ for $i = n + 2, \ldots, 2n$. ∎

A set of smooth functions $F_1, \ldots, F_k$ defjined on $\mathbb{O} \subset \mathbb{R}^{2n}$ are *in involution* if $\{F_i, F_j\} \equiv 0$ for $1 \leq i, j \leq k$.

Corollary 5. *Near an ordinary point the Hamiltonian system* (5) *has n independent integrals in involution.*

PROOF. In the coordinates of Theorem 3, the coordinates $\eta_1, \ldots, \eta_n$ are independent integrals in involution. ∎

Return to the ordinary equation (1) for the moment. The construction of the coordinate system of Theorem 1 requires the complete solution of the equations. In this case, $m - 1$ integrals are known. In many cases some but not all the integrals are known, in which case some simplification of the system is possible.

Theorem 6. *Assume that Equation* (1) *has* k, $1 \leq k < m$, *integrals that are independent at some point* $r \in \mathbb{O} \subset \mathbb{R}^n$. *Then locally the equations can be reduced to an* $m - k$*-dimensional system which depends on* k *parameters. Moreover, this reduction does not require the complete knowledge of the solutions as in Theorem* 1.

PROOF. Let r be the origin and $F_1, \ldots, F_k$ be the integrals. Since they are independent, the Jacobian $(\partial F_i/\partial x_j)$ has a nonzero subdeterminant of size $k \times k$; assume that it is the subdeterminant with $1 \leq i, j \leq k$. Change variables by

$$\begin{aligned} y_i &= F_i(x) \quad \text{for } i = 1, \ldots, k, \\ y_i &= x_i \qquad \text{for } i = k+1, \ldots, n. \end{aligned} \tag{10}$$

It is clear that this transformation has a nonsingular Jacobian at the origin and so defines a valid change of coordinates. Since $y_1, \ldots, y_k$ are integrals, the equations in the new coordinates are of the form

$$\begin{aligned} \dot{y}_i &= 0 \qquad\qquad\quad \text{for } i = 1, \ldots, k, \\ \dot{y}_i &= g_i(y_1, \ldots, y_m) \quad \text{for } i = k+1, \ldots, m. \end{aligned} \tag{11}$$

The first k equations integrate to give $y_i = \alpha_i$, a constant, for $i = 1, \ldots, k$. Substituting these constants into the remaining equations gives an $m - k$-dimensional system with k parameters. ∎

Consider the Hamiltonian system (5) again. If H is independent of one coordinate, say q_i, then $\dot{p}_i = \partial H/\partial q_i = 0$ or p_i is an integral. Similarly if H is independent of p_i, then q_i is an integral. If H is independent of one coordinate, then it is called an *ignorable coordinate*, and its conjugate is an integral.

Let q_1 be ignorable; so, p_1 is an integral, and $p_1 = \alpha$, a constant. When the variable p_1 is replaced by the parameter α in (5), the equations in (5) are independent of q_1 and p_1. The equations for $q_2, \ldots, q_n, p_2, \ldots, p_n$ are the equations of a Hamiltonian system of $n - 1$ degrees of freedom which depend on a parameter α. If these equations are solved explicitly in terms of t, these solutions can be substituted into the q_1 equation and q_1 can be determined by a single integration or quadrature. Thus, an ignorable coordinate reduces the equations to a system of equations of $n - 1$ degrees of freedom containing a parameter and a quadrature.

In Hamiltonian system, an integral gives rise to an ignorable coordinate and many integrals in involution give rise to many ignorable coordinates.

Theorem 7. *Let* $F_1, \ldots, F_k$, $1 \leq k \leq n$, *be smooth functions on* $\mathbb{O}$, *which are in involution and independent at a point* $r \in \mathbb{O} \subset \mathbb{R}^{2n}$. *Then there exist symplectic coordinates* (ξ, η) *at* r *such that in these coordinates* $F_i = \eta_i$ for $i = 1, \ldots, k$.

PROOF. This theorem is left as an exercise. Use induction on k, the number of functions. ∎

Theorem 8. *Assume that the Hamiltonian system* (5) *has k integrals,* $F_1, \ldots, F_k$, *in involution which are independent at some point* $r \in \mathbb{O}$. *Then there exist symplectic coordinates* ξ, η *such that* $\xi_1, \ldots, \xi_k$ *are ignorable. So the system can be reduced locally to a Hamiltonian system with* $n - k$ *degrees of freedom depending on k parameters and k quadratures.*

For the N-body problem in Jacobi coordinates (see Section III.C) the three components of the center of mass $g = (m_1 q_1 + \cdots + m_N q_N)/M$ are ignorable, and the conjugate momenta are the three components of total linear momentum, $G = p_1 + \cdots + p_N$. Jacobi coordinates effect a reduction in the number of degrees of freedom by 3.

The planar Kepler problem admits the z component of angular momentum as an integral. In polar coordinates r, θ, R, Θ of Section III.C.6, θ is an ignorable coordinate, and its conjugate momentum Θ, angular momentum, is an integral.

D. Noether's Theorem and Reduction

The last section discussed integrals in involution, but the classic three components of angular momentum are not in involution. Also all the results are local results. A complete discussion of the general case where there are global integrals that are not in involution requires a lot of symplectic geometry that would require a long premature digression. Therefore, only the classical cases will be considered here.

1. Symmetries Imply Integrals

Let ψ_t be a Hamiltonian flow on $\mathbb{R}^{2n}$; so, i) for fixed t, the map ψ_t: $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is symplectic, ii) $\psi_0 = \mathrm{id}$, the identity of $\mathbb{R}^{2n}$, and iii) $\psi_t \circ \psi_s = \psi_{t+s}$ for all $t, s \in \mathbb{R}$. By theorem IV.2, $\psi(t, \xi) = \psi_t(\xi)$ is the general solution of a Hamiltonian system $\dot{x} = J\nabla F(x)$, where F: $\mathbb{R}^{2n} \to \mathbb{R}$ is smooth.

Consider a Hamiltonian $H(x)$. The ψ_t is a symplectic symmetry for the Hamiltonian H if

$$H(x) = H(\psi(t, x)) = H(\psi_t(x)) \tag{1}$$

for all $x \in \mathbb{R}^{2n}$ and all $t \in \mathbb{R}$.

Theorem 1 (Noether's Theorem). *Let* ψ_t *be a symplectic symmetry for the Hamiltonian* (1). *Then F is an integral for the Hamiltonian system with Hamiltonian H.*

PROOF. Differentiate (1) with respect to t to get

$$0 = \frac{\partial H(\psi(t,x))}{\partial x} \cdot \frac{\partial \psi(t,x)}{\partial t} = \frac{\partial H(\psi(t,x))}{\partial x} \cdot J \cdot \frac{\partial F(\psi(t,x))}{\partial x} = \{H, F\}(\psi(t,x)).$$

■

Consider the N-body problem as discussed in Section I.C with coordinates $z = (q_1, \ldots, q_N, p_1, \ldots, p_N) \in \mathbb{R}^{6N}$. The Hamiltonian (I.C.5) is invariant under translations. That is, if $b \in \mathbb{R}^3$, then ψ_t: $(q_1, \ldots, q_N, p_1, \ldots, p_N) \to (q_1 + tb, \ldots, q_N + tb, p_1, \ldots, p_N)$ is a symplectic symmetry for the N-body problem. The Hamiltonian which generates ψ_t is $F = b^T(p_1 + \cdots + p_N)$. So by Noether's theorem, $F = b^T(p_1 + \cdots + p_N)$ is an integral for the N-body problem for all b, and so linear momentum, $L = p_1 + \cdots + p_N$ is a integral. In general, translational invariance implies the conservation of linear momentum.

Let $\mathcal{G}$ be the subgroup of $Sp(6N, \mathbb{R})$ consisting of all matrices of the form $T = (A, \ldots, A)$, where $A \in \mathrm{So}(3, \mathbb{R})$ (the special orthogonal group or group of three-dimensional rotations). Then the Hamiltonian H of the N-body problem, (I.C.5), has $\mathcal{G}$ as a symmetry group. That is, $H(Tx) = H(x)$ for all $T \in \mathcal{G}$. This simply means that the equations are invariant under a rotation of coordinates. The algebra of Hamiltonian matrices, $\mathcal{A}$, for $\mathcal{G}$ is the set of all Hamiltonian matrices of the form $B = (C, \ldots, C)$, where C is a 3×3 skew symmetric matrix. So $e^{Bt} \in \mathcal{G}$, and $H(x) = H(e^{Bt}x)$ for all $x \in \mathbb{R}^{2n}$ and $t \in \mathbb{R}$. Thus, $\psi_t(x) = e^{Bt}x$ is a symplectic symmetry for the Hamiltonian of the N-body problem. The Hamiltonian which generates $\psi_t(x) = e^{Bt}x$ is $F = \sum_{i=1}^{n} q_i^T C p_i$, and so by Noether's theorem it is an integral for the N-body problem. If we take the three choices for C as

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{2}$$

then the corresponding integrals are the three components of angular momentum. So the fact that the Hamiltonian is invariant under all rotations implies the law of conservation of angular momentum.

2. Reduction

Symmetries give rise to further reductions. In the first example, the Hamiltonian of the N-body problem is invariant under translations, and so linear momentum is invariant. Holding linear momentum fixed, say equal zero, places three linear constraints on the system; so, the space where linear momentum is fixed is a $6N - 3$-dimensional subspace of $\mathbb{R}^{6N}$. But two configurations of the N bodies which are translations of one another can be identified, i.e., $(q_1, \ldots, q_N, p_1, \ldots, p_N)$ and $(q_1 + b, \ldots, q_N + b, p_1, \ldots, p_N)$ can be identified, where b is any vector in $\mathbb{R}^3$. Making this identification reduces the dimension

by another three dimensions, making the total space $6N - 6$ dimensional. This space is called the *reduced space.*

The easiest way to do the reduction just discussed is to use the Jacobi coordinates given in (IV.B.2). The variable g_N is the center of mass, and all the other position coordinates $u_2, \ldots, u_N$ are relative coordinates; so, the identification given above gives $(g_N + b, u_2, \ldots, u_N, G_N, v_2, \ldots, v_N)$ and $(g_N, u_2, \ldots, u_N, G_N, v_2, \ldots, v_N)$ are equivalent. A representative of the equivalence class is $(0, u_2, \ldots, u_N, G_N, v_2, \ldots, v_N)$, i.e., set the center of mass at the origin. Linear momentum, G_N, is an integral. So the reduction discussed above is accomplished by setting $g_N = 0$ and fixing G_N, say zero. The problem is described by a Hamiltonian on an even-dimensional space, the *reduced space.* The Hamiltonian on the reduced space is (IV.B.13) with $g_N = G_N = 0$. Note that the problem is not Hamiltonian when just the integrals of linear momentum are fixed, but it is when these integrals are fixed and points are identified by the translational symmetry.

Now consider the So(3, $\mathbb{R}$) symmetry which gives rise to the angular momentum integrals. There are three angular momentum integrals which are independent except at the origin. Consider the subset, $M \subset \mathbb{R}^{6N}$, of phase space where angular momentum is some fixed, nonzero vector V. This is a $6N - 3$-dimensional space (submanifold) which is invariant under the flow defined by the N-body problem. Not all rotations leave M fixed, only those that are rotations about V. That is, let $\mathcal{G}'$ be the subgroup of $\mathcal{G}$, and hence of Sp($6N$, $\mathbb{R}$), consisting of all matrices of the form $T = (A, \ldots, A)$ where $A \in$ So(3, $\mathbb{R}$) and $AV = V$ (rotations about V). $\mathcal{G}'$ is the same as (isomorphic to) So(2, $\mathbb{R}$), rotations of the plane. This can be seen by changing coordinates so that V is along one of the coordinate axes. $\mathcal{G}'$ and So(2, $\mathbb{R}$) are one dimensional because they can be parameterized by the angle of rotation.

Clearly, if $T \in \mathcal{G}'$, then T leaves M invariant; so, two points $z, z' \in M$ can be identified if $Tz = z'$, i.e., if one configuration can be rotated into the other by a rotation about V. Let B be the identification space $M/_\sim$, where $z \sim z'$ when $z \in Tz'$ for some $T \in \mathcal{G}'$. It turns out that M is $6N - 3$ dimensional, and B is $6N - 4$ dimensional. The interesting fact is that B is "symplectic," and the flow of the N-body problem is Hamiltonian on B, i.e., there are local coordinate on B which are symplectic, and the equations of motion of the N-body problem are Hamiltonian. (Technically, B is a symplectic manifold.)

The two can be done together. The N-body problem is a first-order system of differential equations in a $6N$-dimensional space $\mathbb{R}^{6N}$. The first reduction of placing the center of mass at the origin and fixing linear momentum reduces the problem to a linear subspace of dimension $6N - 6$. Fixing angular momentum reduces the problem to a $6N - 9$-dimensional space, M. Identifying configurations which differ by a rotation about the angular momentum reduces the problem to the reduced space B of dimension $6N - 10$. This last operation is classically called "the elimination of the node." The N-body problem of B is a time-independent Hamiltonian system on the symplectic space B. Two further reductions can be accomplished by holding the Hamil-

tonian (energy) fixed and eliminating time to get a nonautonomous sytem of differential equations of order $6N - 12$. The reduction of the N-body problem is classical, with the elimination of the node due to Jacobi. The general results about the symplectic nature of the reduced space is in Meyer (1973) and Marsden and Weinstein (1974).

The proofs of all the facts about the reductions, symplectic manifolds, etc., would require a long digression, but a treating of a special case will whet the reader's interest. The reader is referred to the original articles or to Abraham and Marsden (1978) for the complete discussion. Let us consider the planar 3-body problem. Recall from Section IV.B that the Hamiltonian of the 3-body problem in Jacobi coordinates with the center of mass at the origin and linear momentum equal to zero is

$$H = \frac{\|v_1\|^2}{2M_1} + \frac{\|v_2\|^2}{2M_2} - \frac{m_1 m_2}{\|u_1\|} - \frac{m_1 m_2}{\|u_2 + \alpha_0 u_1\|} - \frac{m_2 m_3}{\|u_2 - \alpha_{11}\|}, \tag{3}$$

where

$$\begin{aligned} M_1 &= \frac{m_1 m_2}{m_1 + m_2}, & M_2 &= \frac{m_3(m_1 + m_2)}{m_1 + m_2 + m_3}, \\ \alpha_0 &= \frac{m_2}{m_2 + m_1}, & \alpha_1 &= \frac{m_1}{m_2 + m_1}. \end{aligned} \tag{4}$$

This effect the first reduction. Putting this Hamiltonian in polar coordinates as in (IV.C.8) gives

$$\begin{aligned} H = {} & \frac{1}{2M_1}\left\{R_1^2 + \left(\frac{\Theta_1^2}{r_1^2}\right)\right\} + \frac{1}{2M_2}\left\{R_2^2 + \left(\frac{\Theta_2^2}{r_2^2}\right)\right\} + \frac{m_2 m_1}{r_1} \\ & + \frac{m_1 m_3}{\sqrt{r_2^2 + \alpha_0^2 r_1^2 - 2\alpha_0 r_1 r_2 \cos(\theta_2 - \theta_1)}} \\ & + \frac{m_2 m_3}{\sqrt{r_2^2 + \alpha_1^2 r_1^2 - 2\alpha_1 r_1 r_2 \cos(\theta_2 - \theta_1)}}. \end{aligned} \tag{5}$$

Since the Hamiltonian depends only on the difference of the two polar angles, make the symplectic change of coordinates

$$\begin{aligned} \phi_1 &= \theta_2 - \theta_1, & \phi_2 &= -\theta_2 + 2\theta_1, \\ \Phi_1 &= 2\Theta_2 + \Theta_1, & \Phi_2 &= \Theta_2 + \Theta_1, \end{aligned} \tag{6}$$

so that (5) becomes

$$\begin{aligned} H = {} & \frac{1}{2M_1}\left\{R_1^2 + \frac{(2\Phi_1 - \Phi_2)^2}{r_1^2}\right\} + \frac{1}{2M_2}\left\{R_2^2 + \frac{(\Phi_2 - \Phi_1)^2}{r_2^2}\right\} - \frac{m_0 m_1}{r_1} \\ & - \frac{m_0 m_2}{\sqrt{r_2^2 + \alpha_0^2 r_1^2 - 2\alpha_0 r_1 r_2 \cos(\phi_1)}} - \frac{m_1 m_2}{\sqrt{r_2^2 + \alpha_1^2 r_1^2 - 2\alpha_1 r_1 r_2 \cos(\phi_1)}}. \end{aligned} \tag{7}$$

Note that the Hamiltonian is independent of ϕ_2; so ϕ_2 is an ignorable coordinate. Therefore, its conjugate momentum $\Phi_2 = \Theta_2 + \Theta_1$, total angular momentum, is an integral. The reduction to the reduced space is done by holding Φ_2 fixed and ignoring ϕ_2. The Hamiltonian (7) has three degrees of freedom, (r_1, r_2, ϕ_1), and one parameter Φ_2.

E. Periodic Solutions, Fixed Points, and Cross Sections

In view of the results of the previous sections, it would seem that the next question to be considered is the nature of the flow near an equilibrium point. This is one of the central and difficult questions in the local geometric theory. It turns out that many of the questions about equilibrium points are very similar to questions about periodic solutions. This section will introduce this similarity.

1. Equilibrium Points

Consider first a general system

$$\dot{x} = f(x), \tag{1}$$

where $f: \mathbb{O} \to \mathbb{R}^m$ is smooth, and $\mathbb{O}$ is open in $\mathbb{R}^m$. Let the general solution be $\phi(t, \xi)$. An equilibrium solution $\phi(t, \xi')$ is such that $\phi(t, \xi') \equiv \xi'$ for all t. Obviously $\phi(t, \xi')$ is an equilibrium solution if and only if $f(\xi') = 0$. So questions about the existence and uniqueness of equilibrium solutions are finite-dimensional questions. The eigenvalues of $\partial f(\xi')/\partial x$ are called the (*characteristic*) *exponents* of the equilibrium point. If $\partial f(\xi')/\partial x$ is nonsingular, or equivalently the exponents are all nonzero, then the equilibrium point is called *elementary*.

Proposition 1. *Elementary equilibrium points are isolated.*

PROOF. $f(\xi') = 0$ and $\partial f(\xi')/\partial x$ is nonsingular; so, the implicit function theorem applies to f; and there is neighborhood of ξ' with no other zeros of f. ∎

The analysis of stability, bifurcations, etc., of equilibrium points starts with an analysis of the linearized equations. For this reason, one shifts the equilibrium point to the origin, and (1) is rewritten

$$\dot{x} = Ax + g(x), \tag{2}$$

where $A = \partial f(0)/\partial x$, $g(x) = f(x) - Ax$; so, $g(0) = 0$ and $\partial g(0)/\partial x = 0$. The eigenvalues of A are the exponents. The reason the eigenvalues of A are called

exponents is that the linearized equations [e.g., $g(x) \equiv 0$ in (2)] have solutions which contain terms like $\exp(\lambda t)$, where λ is an eigenvalue of A.

2. Periodic Solutions

A periodic solution is a solution, $\phi(t, \xi')$, such that $\phi(t + T, \xi') \equiv \phi(t, \xi')$ for all t, where T is some nonzero constant. T is called *a period*, and the least positive T which satisfies that relation is called *the period* or *the least period.* In general, an equilibrium solution will not be considered as a periodic solution; however, some statements—like Lemma 2—have a simpler statement if equilibrium solutions are considered as periodic solutions. It is easy to see that if the solution is periodic and not an equilibrium solution, then the least period exists, and all periods are integer multiples of it.

Lemma 2. *A necessary and sufficient condition for $\phi(t, \xi')$ to be periodic with a period T is*

$$\phi(T, \xi') = \xi', \tag{3}$$

where T is nonzero.

PROOF. This is a restatement of Lemma A.4. ∎

This lemma shows that questions about the existence and uniqueness of periodic solutions are ultimately finite-dimensional questions. The analysis and topology of finite-dimensional spaces should be enough to answer all such questions.

Let $\phi(t, \xi')$ be periodic with least period T. The matrix $\partial\phi(T, \xi')/\partial\xi$ is called the *monodromy* matrix, and its eigenvalues are called the (*characteristic*) *multipliers* of the periodic solution. It is tempting to use the implicit function theorem on (3) to find a condition for local uniqueness of a periodic solution. To apply the implicit function theorem to (3), the matrix $\partial\phi(T, \xi')/\partial\xi - I$ would have to be nonsingular, or $+1$ would not be a multiplier. But this will never happen.

Lemma 3. *Periodic solutions of* (1) *are never isolated, and $+1$ is always a multiplier. In fact, $f(\xi')$ is an eigenvector of the monodromy matrix corresponding to the eigenvalue $+1$.*

PROOF. Since (1) is autonomous, it defines a local dynamical system; so, a translate of a solution is a solution. Therefore, the periodic solution is not isolated. Differentiating the group relation $\phi(\tau, \phi(t, \xi')) = \phi(t + \tau, \xi')$ with respect to ξ and setting $t = 0$ and $\tau = T$ gives

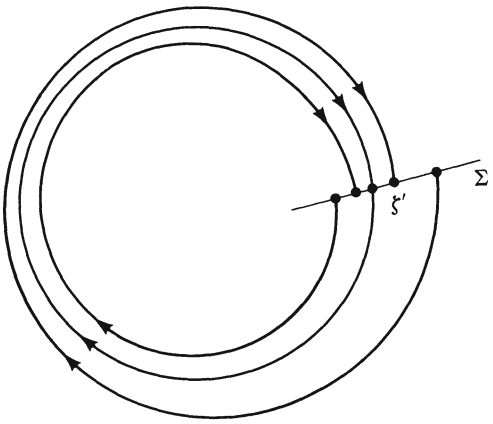

Figure E.1. The cross section.

$$\frac{\partial \phi}{\partial \zeta}(T, \xi')\dot{\phi}(0, \xi') = \dot{\phi}(T, \xi'),$$

$$\frac{\partial \phi}{\partial \zeta}(T, \xi')f(\xi') = f(\xi').$$

Since the periodic solution is not an equilibrium point, $f(\xi') \neq 0$. ∎

Because of this lemma, the correct concept is "isolated periodic orbit." In order to overcome the difficulties implicit in Lemma 3, one introduces a cross section. Let $\phi(t, \xi')$ be a periodic solution. A *cross section to the periodic solution*, or simply a *section*, is a hyperplane, Σ, of codimension 1 through ξ' and transverse to $f(\xi')$. For example, Σ would be the hyperplane $\{x: a^T(x - \xi') = 0\}$, where a is a constant vector with $a^T f(\xi') \neq 0$. The periodic solution starts on the section and, after a time T, returns to it. By the continuity of solutions with respect to initial conditions, nearby solutions do the same. See Figure E.1. So if ξ is close to ξ' on Σ, there is a time $\mathscr{T}(\xi)$ close to T such that $\phi(\mathscr{T}(\xi), \xi)$ is on Σ. $\mathscr{T}(\xi)$ is called the *first return time.* The *section map*, or *Poincaré map*, is defined as the map P: $\xi \to \phi(\mathscr{T}(\xi), \xi)$ which is a map from a neighborhood N of ξ' in Σ into Σ.

Lemma 4. *If the neighborhood N of ξ' in Σ is sufficiently small, then the first retun time, $\mathscr{T}: N \to \mathbb{R}$, and the Poincaré map, $P: N \to \Sigma$, are smooth.*

PROOF. Let $\Sigma = \{x: a^T(x - \xi') = 0\}$, where $a^T f(\xi') \neq 0$. Consider the function $g(t, \xi) = a^T(\phi(t, \xi) - \xi')$. Since $g(T, \xi') = 0$ and $\partial g(T, \xi')/\partial \xi = a^T \dot{\phi}(T, \xi') = a^T f(\xi') \neq 0$, the implicit function theorem gives a smooth function $\mathscr{T}(\xi)$ such that $g(\mathscr{T}(\xi), \xi) = 0$. g being zero defines Σ so that the first return time, $\mathscr{T}$, is

smooth. The Poincaré map is smooth because it is the composition of two smooth maps. ∎

The periodic solution now appears as a fixed point of P; indeed, any fixed point, ξ'', of P is the initial condition for a periodic solution of period $\mathcal{T}(\xi'')$, since $(\mathcal{T}(\xi''), \xi'')$ would satisfy (3). A point $\xi'' \in N$ such that $p^k(\xi'') = \xi''$ for some integer $k > 0$ is called a periodic point of P of period k. The solution of (1) through such a periodic point will be periodic with period nearly $k\mathcal{T}(\xi'')$.

The analysis of stability, bifurcations, etc., of fixed points starts with an analysis of the linearized equations. For this reason, one shifts the fixed point to the origin and writes the Poincaré map

$$P(y) = Ay + g(y), \tag{4}$$

where $A = \partial P(0)/\partial y$, $g(y) = P(y) - Ay$, so $g(0) = 0$, and $\partial g(0)/\partial y = 0$. The eigenvalues of A are the mulitpliers of the fixed point. The reason the eigenvalues, λ_i, of A are called multipliers is that the linearized map [*e.g.*, $g(x) \equiv 0$ in (4)] takes an eigenvector to λ_i times itself. A fixed point is called *elementary* if none of its multipliers are equal to $+1$.

Lemma 5. *If the multipliers of the periodic solution are* $1, \lambda_2, \ldots, \lambda_m$, *then the multipliers of the corresponding fixed point of the Poincaré map are* $\lambda_2, \ldots, \lambda_m$.

PROOF. Rotate and translate coordinates so that $\xi' = 0$ and $f(\xi') = (1, 0, \ldots, 0)$; so, Σ is the hyperplane $x_1 = 0$. Let $B = \partial\phi(T, \xi')/\partial\xi$, the monodromy matrix. By Lemma 2, $f(\xi')$ is an eigenvector of B corresponding to the eigenvalue $+1$. In these coordinates,

$$B = \begin{bmatrix} 1 & x & x & x & x \\ 0 & & & & \\ \vdots & & & A & \\ 0 & & & & \end{bmatrix}. \tag{5}$$

Clearly, the eigenvalues of B are $+1$ along with the eigenvalues of A. ∎

We also call the eigenvalues $\lambda_2, \ldots, \lambda_n$ the *nontrivial multipliers* of the periodic orbit. Recall that an orbit is the solution considered as a curve in $\mathbb{R}^n$, and so is unaffected by reparameterization. A periodic orbit of period T is *isolated* if there is a neighborhood L of it with no other periodic orbits in L with period near to T. There may be periodic solutions of much higher period near an isolated periodic orbit. A periodic orbit is isolated if and only if the corresponding fixed point of the Poincaré map is an isolated fixed point. A periodic orbit is called *elementary* if none of its nontrivial multipliers are $+1$.

Proposition 6. *Elementary fixed points and elementary periodic orbits are isolated.*

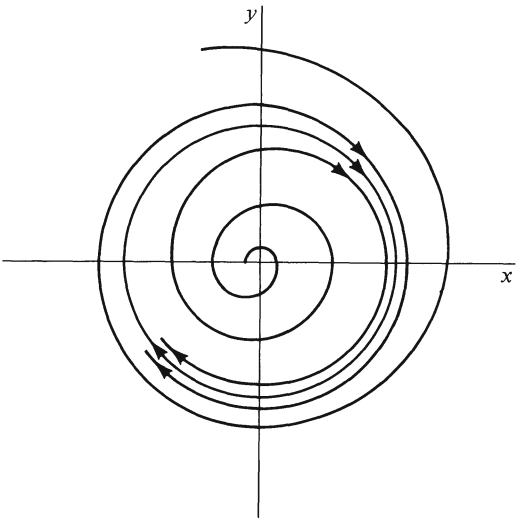

Figure E.2. Phase portrait of the example.

PROOF. Apply the implicit function theorem to the Poincaré map. ∎

3. A Trivial Example

Consider the system

$$\begin{aligned} \dot{x} &= x + y + x(1 - x^2 - y^2), \\ \dot{y} &= -x + y + y(1 - x^2 - y^2), \end{aligned} \tag{6}$$

which in polar coordinates is

$$\begin{aligned} \dot{r} &= r + r(1 - r^2), \\ \dot{\theta} &= -1; \end{aligned} \tag{7}$$

see Figure E.2. The origin is an elementary equilibrium point, and the unit circle is an elementary periodic orbit. To see the latter claim, consider the cross section $\theta \equiv 0 \bmod 2\pi$. The first return time is 2π. The linearized equation about $r = 1$ is $\dot{r} = -1$, and so the linearized Poincaré map is $r \to r \exp(-2\pi)$. The multiplier of the fixed point is $\exp(-2\pi)$.

4. Systems with Integrals

By Lemma 3, the monodromy matrix of a periodic solution has $+1$ as a multiplier. If Equation (1) were Hamiltonian, the monodromy matrix would be symplectic by Theorem III.A.2, and so by Proposition II.C.1, the algebraic

multiplicity of the eigenvalue $+1$ would be even and so at least 2. Actually, this is simply due to the fact that an autonomous Hamiltonian system has an integral.

Throughout this subsection assume that Equation (1) admits an integral F, where F is a smooth map from $\mathbb{O}$ to $\mathbb{R}$, and assume that $\phi(t, \xi')$ is a periodic solution of period T. Furthermore, assume that the integral F is nondegenerate on this periodic solution, i.e., $\nabla F(\xi')$ is nonzero. For a Hamiltonian system, the Hamiltonian, H, is always nondegenerate on a nonequilibrium solution because $\nabla H(\xi') = 0$ would imply an equilibrium.

Lemma 7. *If F is nondegenerate on the periodic solution $\phi(t, \xi')$, then the multiplier $+1$ has algebraic multiplicity at least 2. Moreover, the row vector $\partial F(\xi')/\partial x$ is a left eigenvector of the monodromy matrix corresponding to the eigenvalue $+1$.*

PROOF. Differentiating $F(\phi(t, \xi)) \equiv F(\xi)$ with respect to ξ and setting $\xi = \xi'$ and $t = T$ yields

$$\frac{\partial F(\xi')}{\partial x} \frac{\partial \phi(T, \xi')}{\partial \zeta} = \frac{\partial F(\xi')}{\partial x}, \tag{8}$$

which implies the second part of the lemma. Choose coordinates so that $f(\xi')$ is the column vector $(1, 0, \ldots, 0)^T$ and $\partial F(\xi')/\partial x$ is the row vector $(0, 1, 0, \ldots, 0)$. Since $f(\xi')$ is a right eigenvector and $\partial F(\xi')/\partial x$ is a left eigenvector, the monodromy matrix $B = \partial \phi(T, \xi')/\partial \xi$ has the form

$$B = \begin{bmatrix} 1 & x & x & x & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & x & x & x & \cdots & x \\ 0 & x & x & x & \cdots & x \\ \vdots & & & & & \vdots \\ 0 & x & x & x & \cdots & x \end{bmatrix}. \tag{9}$$

Expand by minors $p(\lambda) = \det(B - \lambda I)$. First expand along the first column to get $p(\lambda) = (1 - \lambda) \det(B' - \lambda I)$, where B' is the $(m - 1) \times (m - 1)$ matrix obtained from B by deleting the first row and column. Expand $\det(B' - \lambda I)$ along the first row to get $p(\lambda) = (1 - \lambda)^2 \det(B'' - \lambda I) = (1 - \lambda)^2 q(\lambda)$, where B'' is the $(m - 2) \times (m - 2)$ matrix obtained from B by deleting the first two rows and columns. ∎

Again there is a good geometric reason for the degeneracy implied by this lemma. The periodic solution lies in an $m - 1$-dimensional level set of the integral, and typically in nearby level sets of the integral, there is a periodic orbit. So periodic orbits are not isolated.

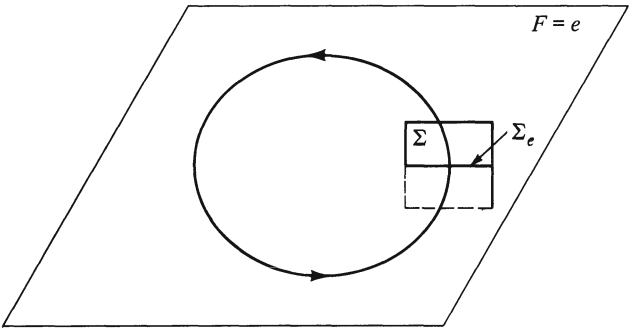

Figure E.3. Poincaré map in an integral surface.

Consider the Poincaré map $P: N \to \Sigma$, where N is a neighborhood of ξ' in Σ. Let u be flow box coordinates given by Theorem IV.B.2 so that ξ' corresponds to $u = 0$; Equations (1) are $\dot{u}_1 = 1$, $\dot{u}_2 = 0$, ..., $\dot{u}_m = 0$, and $F(u) = u_2$. In these coordinates, we may take Σ to be $u_1 = 0$. Since u_2 is the integral in these coordinates, P maps the level sets $u_2 =$ constant into themselves; so, we can ignore the u_2 component of P. Let $e = u_2$; let Σ_e be the intersection of Σ and the level set $F = e$; and let $y_1 = u_3, \ldots, y_{m-2} = u_m$ be coordinates in Σ_e. Here e is considered as a parameter (the value of the integral). In these coordinates, the Poincaré map P is a function of y and the parameter e. So $P(e, y) = (e, Q(e, y))$, where for fixed e, $Q(e, \cdot)$ is a mapping of a neighborhood N_e of the origin in Σ_e into Σ_e. Q is called the Poincaré map in an integral surface, see Figure E.3. The eigenvalues of $\partial Q(0, 0)/\partial y$ are called the *multipliers of the fixed point in the integral surface* or *the nontrivial multipliers*. By the same argument as above, we have the following lemma.

Lemma 8. *If the multipliers of the periodic solution of a system with nondegenerate integral are* $1, 1, \lambda_3, \ldots, \lambda_m$, *then the multipliers of the fixed point in the integral surface are* $\lambda_3, \ldots, \lambda_m$.

Lemma 9. *If the system is Hamiltonian, then the Poincaré map in an integral surface is symplectic.*

PROOF. In this case, use the Hamiltonian flow box theorem (Theorem IV.A.2). In this case, $H = \eta_1$, and the equations are $\dot{\xi}_1 = 1$, $\dot{\xi}_i = 0$ for $i = 2, \ldots, n$ and $\dot{\eta}_i = 0$ for $i = 1, \ldots, n$. The cross section is $\xi_1 = 0$, and the integral parameter is $\eta_1 = e$. The Poincaré map in an integral surface in these coordinates is in terms of the symplectic coordinates $\xi_2, \ldots, \xi_n, \eta_2, \ldots, \eta_n$ on Σ_e. Since the total map $x \to \phi(T, x)$ is symplectic (Theorem III.A.2), the map $y \to Q(e, y)$ is symplectic. ∎

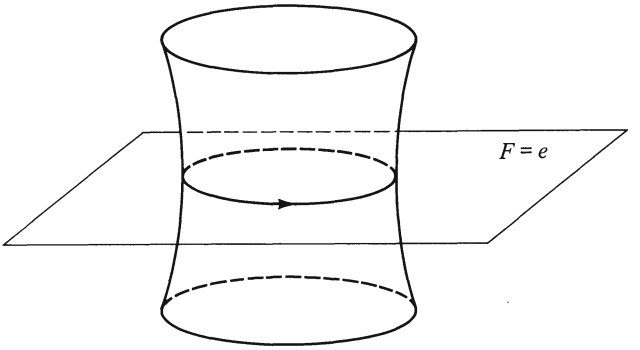

Figure E.4. Cylinder of periodic solutions.

If none of the nontrivial multipliers are 1, and the integral is nondegenerate on the periodic solution, then we will say that the periodic solution (or fixed point) is *elementary*.

Theorem 10 (The Cylinder Theorem). *An elementary periodic orbit of a system with integral lies in a smooth cylinder of periodic solutions parameterized by the integral F.* (See Figure E.4.)

PROOF. Apply the implicit function theorem to $Q(e, y) - y = 0$ to get a one-parameter family of fixed points $y^*(e)$ in each integral surface $F = e$. ∎

F. The Stable Manifold Theorem

In this section we will discuss some important theorems about the local structure of differential equations near equilibrium and diffeomorphisms near fixed points by introducing the concept of a hyperbolic point. These theorems are classical, and their proofs appear in many standard textbooks; so, we shall not prove them here. In the next section we carry forth the generalization to hyperbolic sets. In the last 20 years, the subject of hyperbolic dynamical systems has become a subject of its own. See the monographs by Nitecki (1971), Szlenk (1981), and Palis and de Melo (1980). Since Hamiltonian dynamics is more the study of elliptic points than hyperbolic points, we shall concentrate on the elliptic case and refer the reader to the literature for some of the proofs for the hyperbolic theorems.

Let the equation

$$\dot{x} = f(x) \tag{1}$$

have an equilibrium point at $x = p$; so, $f(p) = 0$. Let $A = \partial f(p)/\partial x$; so, A is an $m \times m$ constant matrix. The eigenvalues of A are called the *exponents* of equilibrium point p. The linearization of (1) about $x = p$ is $\dot{y} = Ay$, where $y =$

$x - p$. We say that p *is a hyperbolic equilibrium point for* (1) if A has no eigenvalues with zero real part; so, all the eigenvalues have either positive real parts or negative real parts. Thus, the solutions of the linearized equation are sums of exponentially increasing and decreasing terms. The set of all solutions tending to zero is a linear subspace, as is the set of all solutions tending away from zero. The full nonlinear equations have similar sets, which is the subject of the following theorems.

At first the results are local; so, one can shift the equilibrium point to the origin and write (1) in the form

$$\dot{x} = Ax + g(x), \tag{2}$$

where $g(x) = f(x) - Ax$; so, $g(0) = Dg(0) = 0$, and A is an $m \times m$ real constant matrix with no eigenvalue with zero real part. Let $\phi(t, \xi)$ be the general solution of (2); so, $\phi(0, \xi) = \xi$. Let $\varepsilon > 0$ be given; then the *local stable manifold* is

$$\mathscr{W}^s(\varepsilon) = \{\xi \in \mathbb{R}^m : |\phi(t, \xi)| < \varepsilon \text{ for all } t \geq 0\}, \tag{3}$$

and the *local unstable manifold* is

$$\mathscr{W}^u(\varepsilon) = \{\xi \in \mathbb{R}^m : |\phi(t, \xi)| < \varepsilon \text{ for all } t \leq 0\}. \tag{4}$$

If the reader is not familiar with the definition of a manifold, simply read the remark following the statement of the theorem.

Theorem 1 (The Stable Manifold Theorem for Differential Equations). *Let A have d eigenvalues with negative real parts and $m - d$ eigenvalues with positive real parts. Let g be as above. Then for ε sufficiently small, $\mathscr{W}^s(\varepsilon)$ and $\mathscr{W}^u(\varepsilon)$ are smooth manifolds of dimensions d and $m - d$, respectively. If $\xi \in \mathscr{W}^s(\varepsilon)$ [respectively $\xi \in \mathscr{W}^u(\varepsilon)$], then $\phi(t, \xi) \to 0$ as $t \to +\infty$ [respectively $\phi(t, \xi) \to 0$ as $t \to -\infty$]. Actually, there is a smooth, near identity change of coordinates which takes the stable and unstable manifolds to (different) coordinate planes.*

PROOF. See Coddington and Levinson (1955), Hale (1972), or Hartman (1964). ∎

Remarks. By a linear change of coordinates, if necessary, we may assume that

$$A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix},$$

where B is a $d \times d$ matrix with eigenvalues with negative real parts, and C is an $(m - d) \times (m - d)$ matrix with positive real parts. Writing $\mathbb{R}^m = \mathbb{R}^d \times \mathbb{R}^{m-d}$, $(z, w) \in \mathbb{R}^m = \mathbb{R}^{m-d}$; so, Equation (2) becomes

$$\begin{aligned} \dot{z} &= Bz + h(z, w), \\ \dot{w} &= Cw + k(z, w), \end{aligned} \tag{5}$$

where h, k, and their first partials vanish at the origin. One proof of the

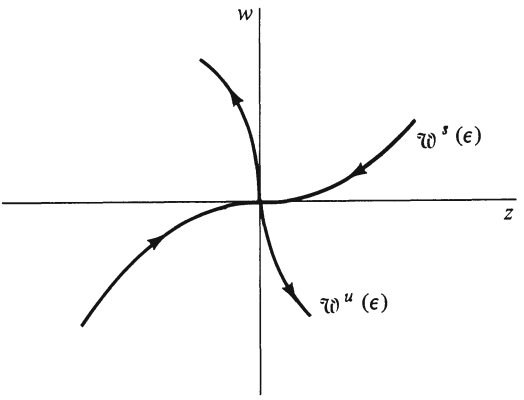

Figure F.1. Local stable and unstable manifolds.

existence of the stable manifold establishes the existence of a change of coordinates of the form $\xi = z$, $\eta = w - u(z)$, which makes the ξ coordinate hyperplane invariant or at least locally invariant. The function u is shown to be smooth and small with $u(0) = Du(0) = 0$. Thus, in the new coordinates, the local stable manifold is a piece of the d-dimensional linear space $\eta = 0$. In the original coordinates, the local stable manifold is the graph of the function u. Since $u(0) = Du(0) = 0$, the stable manifold is tangent to the d-dimensional linear space $z = 0$. See Figure F.1.

This is a local result; so, a natural question to ask is what happens to these manifolds. Now we shall show how to continue these manifolds. Assume f in (1) is globally defined and let the general solution $\phi(t, \xi)$ of (1) be globally defined; so, ϕ defines a dynamical system. The (*global*) *stable manifold* is

$$\mathcal{W}^s = \mathcal{W}^s(p) = \{\xi \in \mathbb{R}^m \colon \phi(t, \xi) \to p \text{ at } t \to +\infty\}, \tag{6}$$

and the (*global*) *unstable manifold* is

$$\mathcal{W}^u = \mathcal{W}^u(p) = \{\xi \in \mathbb{R}^m \colon \phi(t, \xi) \to p \text{ as } t \to -\infty\}. \tag{7}$$

Theorem 2 (The Global Stable Manifold Theorem for Differential Equations). *Let p by a hyperbolic equilibrium point with d exponents with negative real parts and $m - d$ exponents with positive real parts. Then the stable manifold is an immersed d-dimensional submanifold. That is, there exists a smooth function $\Gamma \colon \mathbb{R}^d \to \mathbb{R}^n$ which is globally one-to-one, and $D\Gamma$ has rank d everywhere. Similarly, the unstable manifold is an immersed $m - d$ submanifold.*

PROOF. See Palis and de Melo (1980). ∎

There are similar theorems for diffeomorphisms. Consider a diffeomorphism

$$\psi: \mathbb{R}^m \to \mathbb{R}^m \tag{8}$$

with a fixed point p. Let $A = \partial\psi(p)/\partial x$; so, A is an $m \times m$ constant matrix. The eigenvalues of A are called the *multipliers of* p. The linearization of (8) about $x = p$ is $y \to Ay$, where $y = x - p$. We say that p *is a hyperbolic fixed point* if A has no eigenvalues with absolute value equal to 1. The set of all trajectories tending to zero is a linear subspace, as is the set of all solutions tending away from zero.

The first theorem is local; so, shift the fixed point to the origin and consider

$$x \to \Phi(x) = Ax + g(x), \tag{9}$$

where g is defined and smooth in a neighborhood of the origin in $\mathbb{R}^m$ with $g(0) = 0, Dg(0) = 0$. Define the stable manifold as

$$\mathcal{W}^s(\varepsilon) = \{x \in \mathbb{R}^m: |\Phi^k(x)| < \varepsilon \text{ for all } k \geq 0\} \tag{10}$$

and the unstable manifold similarly.

Theorem 3 (The Stable Manifold Theorem for Diffeomorphisms). *Let A have d eigenvalues with absolute value less than 1 and $m - d$ eigenvalues with absolute value greater than 1. Let g be as above. Then for ε sufficiently small, $\mathcal{W}^s(\varepsilon)$ and $\mathcal{W}^u(\varepsilon)$ are smooth manifolds of dimensions d and $m - d$, respectively. If $\xi \in \mathcal{W}^s(\varepsilon)$ [respectively $\xi \in \mathcal{W}^u(\varepsilon)$], then $\Phi^k(\xi) \to 0$ as $k \to +\infty$ [respectively $\Phi^k(\xi) \to 0$ as $k \to -\infty$]. Actually there is a smooth, near identity change of coordinates which takes the stable and unstable manifolds to (different) coordinate planes.*

Assume ψ in (8) is a global diffeomorphism; so, it defines a dynamical system. The (*global*) *stable manifold* is

$$\mathcal{W}^s = \mathcal{W}^s(p) = \{\xi \in \mathbb{R}^m: \psi^k(\xi) \to p \text{ as } k \to +\infty\}, \tag{11}$$

and the (*global*) *unstable manifold* is similarly defined.

Theorem 4 (The Global Stable Manifold Theorem for Diffeomorphism). *Let p be a hyperbolic fixed point for ψ with d multipliers with absolute value less than 1 and $m - d$ multiplies with absolute value greater than 1. Then the stable manifold is an immersed d-dimensional submanifold. That is, there exists a smooth function $\Gamma: \mathbb{R}^d \to \mathbb{R}^n$ which is globally one-to-one, and $D\Gamma$ has rank d everywhere. Similarly, the unstable manifold is an immersed $m - d$ submanifold.*

For the rest of this section let ψ in (8) be diffeomorphism of $\mathbb{R}^2$, p be a hyperbolic fixed point with one multiplier greater than 1 and one multiplier less than 1 so that the stable and unstable manifolds of p are smooth curves. A point $q \in \mathcal{W}^s(p) \cap \mathcal{W}^u(p)$, $q \neq p$, is called a *homoclinic point* (*homoclinic to* p). Since q is in both the stable and unstable manifolds of p, $\psi^k(q) \to p$ as

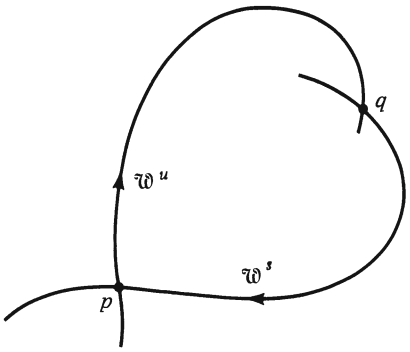

Figure F.2. Transverse intersection of stable and unstable manifolds.

$k \to \pm\infty$, the orbit of q is said to be *doubly asymptotic to p*. If the curves $\mathscr{W}^s(p)$ and $\mathscr{W}^u(p)$ are not tangent at q, the intersection is said to be *transversal*, and q is said to be a *transversal homoclinic point*. Henceforth, let q be a transversal homoclinic point, homoclinic to p. See Figure F.2.

The stable and unstable manifolds are invariant, and so $\psi^k(q) \in \mathscr{W}^s(p) \cap \mathscr{W}^u(p)$ for all k, or the whole orbit of a homoclinic point consists of homoclinic points. In a neighborhood of the hyperbolic point p, the diffeomorphism ψ is well approximated by its linear part, and so it contracts in the stable manifold direction and expands in the unstable direction. This results in the following:

Theorem 5 (Palis's Lambda Lemma). *Let Λ be any interval in the unstable manifold with p in its interior. Let λ be a small segment of the unstable manifold with q in its interior. Then for any $\varepsilon > 0$, there is a K such that for $k \geq K$, Λ is within the ε neighborhood of $\psi^k(\lambda)$. Moreover, if $a \in \Lambda$, $b \in \psi^k(\lambda)$ and* $\operatorname{dist}(a, b) < \varepsilon$, *then the tangents to Λ and $\psi(\lambda)^k$ are within ε.*

PROOF. See Palis and de Melo (1980) and Figure F.3. ∎

This theorem says that a small segment, λ, of the unstable manifold is stretched out and that C^1 approximates the whole unstable manifold by iterates of ψ.

Images of this small segment λ are stretched out along the unstable manifold until the image again intersects the stable manifold at r as shown in Figure F.4. So a homoclinic point begets another homoclinic point near the first. Repeating the argument you get:

Theorem 6 (Poincaré's Homoclinic Theorem). *A transversal homoclinic point is the limit of transversal homoclinic points.*

PROOF. See Poincaré (1899) and Figure F.4. ∎

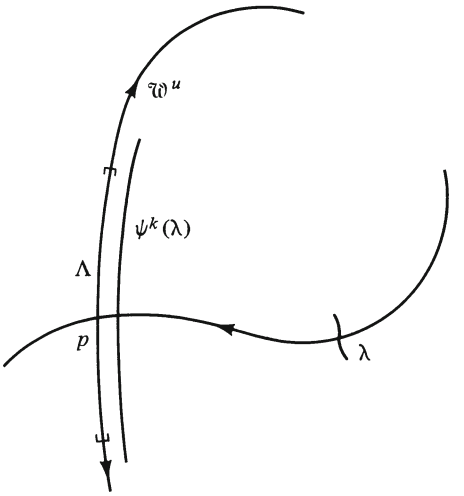

Figure F.3. The lambda lemma.

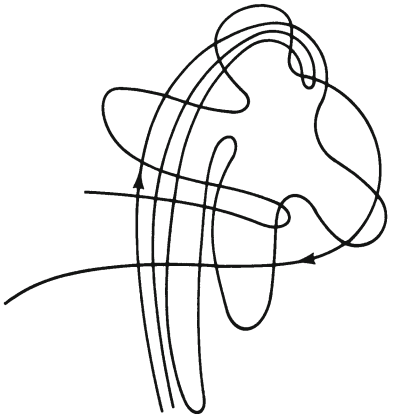

Figure F.4. The Poincaré tangle.

G. Hyperbolic Systems

This section, like the last, contains an introduction to some general topics in dynamical systems that are well documented in the literature. Therefore, very few proofs will be given.

1. The Shift Automorphism and Subshifts of Finite Type

Let $I_n = \{1, 2, \ldots, n\}$ and $\Sigma = \Sigma(n) = \times_{-\infty}^{\infty} I_n$, that is, Σ is the collection of all infinite bisequences on the symbols 1, 2, ..., n. Σ is called the *sequence space*. Thus, if $s \in \Sigma$, then $s = (\ldots, s_{-1}, s_0, s_1, s_2, \ldots)$, or more simply written,

$s = \ldots s_{-1}s_0 \cdot s_1 s_2 \ldots$, where the zeroth position is to the left of the decimal point $\cdot$. Define a distance function on Σ by $d(s, r) = 1/k$, where $k = 1 + \max\{|j|: s_i = r_i \text{ for } |i| \leq j\}$; so, two elements of Σ are close if they agree in a lot of positions around the decimal point. It can be shown that Σ is homeomorphic to the Cantor set.

Let $\sigma: \Sigma \to \Sigma$ be the *shift map* or *shift automorphism* defined by $\sigma(s)_i = s_{i+1}$, i.e, σ shifts the decimal point one place to the right. Clearly, σ is a homeomorphism. The map $\sigma: \Sigma \to \Sigma$ defines a dynamical system, called the *full shift on n symbols.* (It is not differentiable, but it is continuous.) As a dynamical system the map has many interesting properties, among which are those given in the Proposition 1.

The shift on 2 symbols can be considered as flipping a fair coin and taking 1 to mean heads and 2 to mean tails. Any infinite sequence of flipping of a coin is represented by an element of Σ and σ can be thought of as the action of flipping the coin.

Proposition 1. (i) $\sigma: \Sigma \to \Sigma$ *has periodic points of all periods.* (ii) *The periodic points are dense.* (iii) $\sigma: \Sigma \to \Sigma$ *has a dense orbit.* (iv) *Given any two periodic points* $p, q \in \Sigma$, *there is a point* $r \in \Sigma$ *with* $\sigma^k(r) \to p$ *as* $k \to \infty$ *and* $\sigma^k(r) \to q$ *as* $k \to -\infty$. *Moreover, the set of such orbits is dense.*

Remarks. A dynamical system which has a dense orbit is called *transitive.* In (iv) if $p \neq q$, then r is called a *heteroclinic point* and if $p = q$, then r is called a *homoclinic point.* One speaks of heteroclinic and homoclinic orbits also.

PROOF. The proof of all these properties uses the same idea; so, all but one [part (ii)] will be an exercise. We show that there is a periodic point arbitrary close to any give point. To that end, let $q \in \Sigma$ and $\varepsilon > 0$ be given. Let N be so large that $1/N < \varepsilon$ and let s be the finite sequence $q_{-N}q_{-N+1} \cdots q_0, q_1 \cdots q_N$. Construct a bi-infinite sequence r by concatenating an infinite number of times; so, $r = \ldots sss \ldots$ (the decimal point is placed in one of the segments s to the right of q_0). The points r and q agree on a block of length at least N about the decimal point; so, $d(r, q) \leq 1/N < \varepsilon$. Shifting the decimal point $2N + 1$ places brings the sequence r back to itself; so, r is periodic with period $2N + 1$. Thus, there is a periodic point, r, arbitrarily close to the arbitrary point q. ∎

The shift automorphism has many interesting invariant sets, and one type of invariant set which has many nice properties is a subshift of finite type. A *transition matrix* is an $n \times n$ matrix $K = \{k_{ij}\}$ with entries which are either 0 or 1. For any transition matrix K, define a subset of Σ by $\Sigma(K) = \{q \in \Sigma: k_{q_i q_{i+1}} = 1 \text{ for all } i\}$. In other words, adjacent pairs of entries in a sequence $q \in \Sigma(K)$ determine a nonzero entry in K. The transition matrix K serves as a litany of which values can follow which in a sequence $q \in \Sigma(K)$ in the sense that q_{i+1} can follow q_i if and only if $k_{q_i q_{i+1}} = 1$. In the case $k_{q_i q_{i+1}} = 1$, we write as $q_i \to q_{i+1}$ for short. Alternately, the zeros in the transition matrix K rule out certain adjacent pairs in a sequence. For example, if $n = 2$ and

$K = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, then $\Sigma(K)$ consists of all bi-infinite sequences which do not have two adjacent 1's. In general, $\Sigma(K)$ is a closed invariant subset of Σ, and $\sigma\colon \Sigma(K) \to \Sigma(K)$ is called a *subshift of finite* type. If all the entries of K are 1, then $\Sigma(K) = \Sigma$, and for emphasis, this is called the *full shift*.

Subsequently we will need one particular subshift. Henceforth, let $L = L_n$ be the transition matrix

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ & & & & & \ddots & & & \\ 0 & 0 & 0 & 0 & 0 & & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & & 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

That is, L has 1's on the first superdiagonal as well as at the n, 1 and n, n position. Thus $1 \to 2$, $2 \to 3, \ldots, n-1 \to n$ and $n \to n$, $n \to 1$. Let an overbar on a symbol, or sequence of symbols, mean that it is to be repeated infinitely often; so, $\overline{1}.234\overline{5} = \ldots 111.234555\ldots$.

Proposition 2. *Consider the dynamical system* $\sigma\colon \Sigma(L) \to \Sigma(L)$. (i) *It has a unique fixed point.* (ii) *It has periodic points of all periods greater than or equal to n.* (iii) *The periodic points and the points homoclinic to them are dense.* (iv) *There is an invariant subspace* $\Sigma^* \subset \Sigma(L)$ *for* σ^n *such that* $\sigma^n\colon \Sigma^* \to \Sigma^*$ *is equivalent to* $\sigma\colon \Sigma(2) \to \Sigma(2)$, *the full shift on 2 symbols.*

Proof. The fixed point is $f = \overline{n}.\overline{n}$. The point $\overline{n}.123\ldots\overline{n}$ is homoclinic to f. A periodic point of period n is $\overline{12\ldots n} \cdot \overline{12\ldots n}$, a periodic point of period $n+1$ is $\overline{12\ldots nn} \cdot \overline{12\ldots nn}$, etc.

Define Σ^* as those sequences $r \in \Sigma(L)$ of the form $\ldots r_{-1}r_0 \cdot r_1 r_2 \ldots$, where r_i is either the sequence $\alpha = 12\ldots n$ or $\beta = nn\ldots n$. σ^n shifts the decimal point by n positions and so moves the decimal point over a complete block. Define an map h from Σ^* to $\Sigma(2)$ by $h\colon \ldots r_{-1}r_0 \cdot r_1 r_2 \ldots \to \ldots s_{-1}s_0 \cdot s_1 s_2 \ldots$, where $s_i = 1$ if $r_i = \alpha$, and $s_i = 2$ if $r_i = \beta$. This map is a homeomorphism and takes orbits of $\sigma^k\colon \Sigma^* \to \Sigma^*$ to orbits of $\sigma\colon \Sigma(2) \to \Sigma(2)$. This is the equivalence referred to in part (iv). ∎

2. Hyperbolic Structures

The main result discussed in this section is the Smale–Conley theorem which says that a homoclinic point begets an invariant set which is equivalent to the subshift of finite type $\sigma\colon \Sigma(L) \to \Sigma(L)$—the precise statement is given in a sub-

sequent subsection. But before this theorem is discussed, some preliminaries are necessary. Consider a diffeomorphism

$$\psi\colon \mathbb{R}^m \to \mathbb{R}^m. \tag{2}$$

To this dynamical system is associated a linear dynamical system, $\Psi\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ (*the variational system*) defined as follows:

$$\Psi(p, u) = (\psi(p), D\psi(p)u). \tag{3}$$

Here $\mathbb{R}^m$ and its tangent space $T_p\mathbb{R}^m$ are identified—really $\Psi\colon \mathbb{R}^m \times T\mathbb{R}^m \to \mathbb{R}^m \times T\mathbb{R}^m$. Write

$$\Psi^k(p, u) = (\psi^k(p), Y(p, k)u). \tag{4}$$

Recall that $\mathscr{L}(\mathbb{R}^m, \mathbb{R}^m)$ is the space of linear maps from $\mathbb{R}^m$ into $\mathbb{R}^m$ (i.e., $m \times m$ matrices). Let $\Lambda \subset \mathbb{R}^m$ be a compact invariant set for ψ. A compact invariant set $\Lambda \subset \mathbb{R}^m$ has a hyperbolic structure or is a hyperbolic set or ψ admits an *exponential dichotomy over* Λ, if there are constants K and μ, $0 < \mu < 1$, and a continuous mapping $P\colon \Lambda \to \mathscr{L}(\mathbb{R}^m, \mathbb{R}^m)$ such that $P(p)$ is a linear projection operator which satisfies

$$P(\Psi^k(p))\,Y(p, k) = Y(p, k)P(p) \tag{5}$$

and

$$\begin{aligned} \|Y(p, k)P(p)u\| &\le K\mu^k\|u\|, \quad p \in \Lambda, k \ge 0, \\ \|Y(p, k)[I - P(p)]u\| &\le K\mu^{-k}\|u\|, \quad p \in \Lambda, k \le 0. \end{aligned} \tag{6}$$

Define $\mathbb{E}^s_p = \text{range}(P(p))$ and $\mathbb{E}^u_p = \text{kernel}(P(p))$; then since $P(p)$ is a projection, $\mathbb{R}^m = \mathbb{E}^s_p \oplus \mathbb{E}^u_p$, $\text{range}(P(p)) = \text{kernel}(I - P(p))$, and $\text{kernel}(P(p)) = \text{range}(I - P(p))$. The splitting of the tangent space given by $T_p\mathbb{R}^m = \mathbb{R}^m = \mathbb{E}^s_p \oplus \mathbb{E}^u_p$ is continuous in $p \in \Lambda$. Formula (5) states that the linear map $Y(p, k)$ preserves this splitting in that

$$Y(p, k)\colon \mathbb{E}^s_p \to \mathbb{E}^s_q, \qquad Y(p, k)\colon \mathbb{E}^u_p \to \mathbb{E}^u_q, \tag{7}$$

where $q = \psi^k(p)$, and Formulas (6) states that the linear map ultimately contracts vectors in $\mathbb{E}^s_p$ and expands vectors in $\mathbb{E}^u_p$.

3. Examples of Hyprbolic Sets

a. A hyperbolic fixed point, p, of $\psi\colon \mathbb{R}^m \to \mathbb{R}^m$ is a hyperbolic set. Let $A = D\psi(p)$; so, $\mathbb{R}^m = \mathbb{E}^s \oplus \mathbb{E}^u$ where $A \mid \mathbb{E}^s$ has eigenvalues with modulus less than 1, and $A \mid \mathbb{E}^u$ has eigenvalues with modulus greater than 1. The operator $P(p)$ is the projection onto $\mathbb{E}^s$, and $(I - P(p))$ is the projection onto $\mathbb{E}^u$. The (unstable) stable manifold is tangent to ($\mathbb{E}^u$) $\mathbb{E}^s$ at p. $\Psi(p, u) = (p, Au)$.

b. Thom's toral example (see Figure G.1). Let $A = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right)$, and so $A^{-1} = \left(\begin{smallmatrix} 2 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. A has eigenvalues $(3 + \sqrt{5})/2 > 1$ and $(3 - \sqrt{5})/2 < 1$ and so is hy-

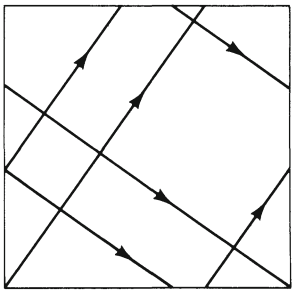

Figure G.1. Thom's torus example.

perbolic with a one-dimensional stable direction and a one-dimensional unstable direction. Since these eigenvalues are irrational, the slopes of the stable and unstable directions are irrational.

Since A and A^{-1} have integer entries, they map the integer lattice $\mathbb{Z} \times \mathbb{Z}$ into itself and so A and A^{-1} can be considered as maps of the 2-torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ into itself. Call this map of the torus A. The fixed point at the origin becomes a hyperbolic fixed point for A, which by the same argument as found in Section I.B has a stable and unstable manifold, both of which are dense in the torus. These stable and unstable manifolds cross in a dense set; so, the homoclinic points are dense also. Let q be a fixed positive integer and $Q = \{(\alpha/q, \beta/q)\colon 0 \leq \alpha, \beta \leq q\}$. A maps Q into itself and so is a permutation of this finite set; so, some power of A fixes Q. Thus, all points with rational coordinates are periodic points, and the set of periodic points is dense.

The whole manifold $\mathbb{T}^2$ has a hyperbolic structure under A. The projection $P(p) = P$ is the projection onto the eigenspace of A corresponding to the eigenvalue $(3 - \sqrt{5})/2 < 1$. A diffeomorphism of a manifold which has a hyperbolic structure everywhere is called an *Anosov system*, after the Russian mathematician who did much of the early studies of such systems. One of his main theorems is that a small perturbation of an Anosov system is equivalent to the original system, and so the geometry of the orbits is not affected much by small perturbations. This property is know as *structural stability*. See Palis and de Melo (1980) for more discussion of these ideas.

c. A transversal homoclinic point (see Figure G.2). Let $\psi\colon \mathbb{R}^m \to \mathbb{R}^m$ have a hyperbolic fixed point at p, and let q be a transversal homoclinic point; so, the stable and unstable manifolds (curves) of p interesect in a nontrivial way at q. Let Λ be the closure of the orbit of q; so, $\Lambda = \{p\} \cup \{\psi^k(q)\colon k \in \mathbb{Z}\}$. Λ has a hyperbolic structure. At each point of the orbit of q, the space $\mathbb{E}^s$ is the tangent space to the stable manifold of p, and $\mathbb{E}^u$ is the tangent space to the unstable manifold of p. At p itself, $\mathbb{E}^s$ and $\mathbb{E}^u$ are as in example **a**. Under positive and negative iterations, the orbit of q gets close to the hyperbolic fixed point p. Careful estimation yields the inequalities in (5).

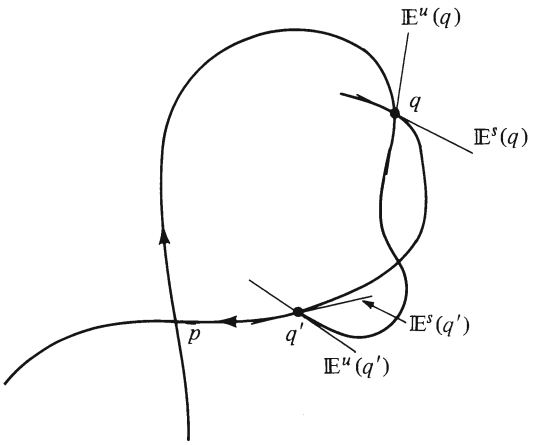

Figure G.2. Transversal homoclinic point.

4. The Shadowing Lemma

A bi-infinite sequence $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$, $x \in \mathbb{R}^m$ is an α-pseudo orbit (for ψ) if $\|x_{k+1} - \psi(x_k)\| \leq \alpha$ for all k. That is, an α-pseudo orbit differs from an actual orbit by at most a jump of distance α at each point. A pseudo orbit $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ is β-shadowed by an orbit $\{\psi^k(y)\}$ if $\|x_k - \psi^k(y)\| \leq \beta$. One of the most striking theorems in the theory of hyperbolic systems is the shadowing lemma of Bowen and Anosov. (Some say the result was known to Lamount Cranston and Margo Lane.)

Theorem 3 (The Shadowing Lemma). *Let $\Lambda \subset \mathbb{R}^m$ be a compact invariant set for ψ: $\mathbb{R}^m \to \mathbb{R}^m$ which has a hyperbolic structure. Then for every $\beta > 0$, there is an $\alpha > 0$ such that every α-pseudo orbit in Λ is β-shadowed by an orbit $\{\psi^k(y)\}$. Moreover, there is a $\beta_0 > 0$ such that if $0 < \beta < \beta_0$, then the ψ-orbit is uniquely determined by the α-pseudo orbit.*

Remarks. An outline of a proof of the shadowing lemma will be given in the appendix to this chapter. If the reader knows a little about the implicit function theorem in a Banach space, the proof is quite simple. The proof is included because the proof in Meyer and Sell (1987) contains a gap which was filled in a more complicated setting in Meyer and Sell (1989).

5. The Conley–Smale Theorem

The existence of a transverse homoclinic point for a planar diffeomorphism implies a certain amount of chaos in the dynamical system as is seen in Theorem 4.

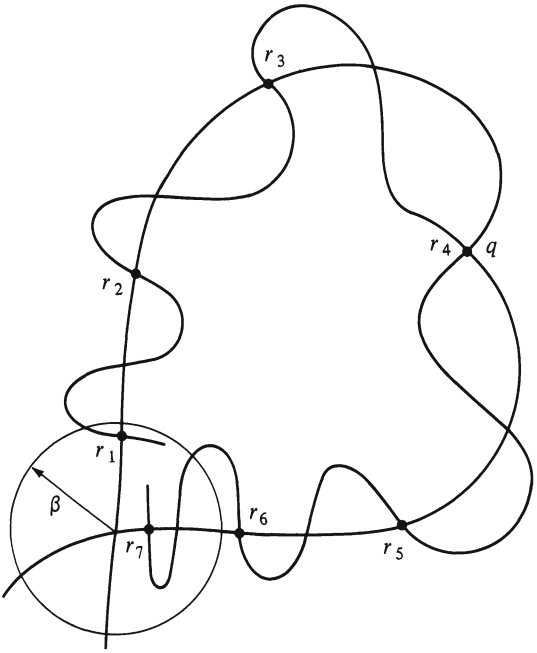

Figure G.3. The Smale–Conley theorem.

Theorem 4 (Conley–Smale). *Let $\psi\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a diffeomorphism with a hyperbolic fixed point at p and let q be a transverse homoclinic point which is homoclinic to p. Then there is an invariant set $\Gamma \subset \mathbb{R}^2$ and an n such that $\psi\colon \Gamma \to \Gamma$ is equivalent to $\sigma\colon \Sigma(L_n) \to \Sigma(L_n)$.*

Proof in Outline. Refer to Figure G.3. As discussed above, the closure of the orbit of q, $\Lambda = \{p\} \cup \{\psi^k(q)\colon k \in \mathbb{Z}\}$, has a hyperbolic structure. Let β_0 be as given in the shadowing lemma. Let β be less than β_0 and also $4\beta < \operatorname{dist}(q, \psi^k(q))$ for all $k \neq 0$. Since $\psi^k(q) \to p$ and $q \neq p$, β can be taken positive. The disks of radius β about p and q do not intersect. Let α be given by the shadowing lemma corresponding to this β, but further require that α is less than β. Draw disks of radius $\alpha/2$ about p and call it D. Draw disks of radius $\alpha/2$ about all points of the orbit of q which lie outside D. α may have to be contracted slightly so that all the disks are disjoint.

Let r be the first point on the backward orbit of q which lies in D; so, $r = \psi^{-k}(q)$, $k > 0$, $r \in D$, and $\psi(r) \notin D$. Let the forward orbit of r be denoted by $r_i = \psi^{i-1}(r)$; so, $r_1 = r$. Let n be such that r_n is the first point on the forward orbit of r in D; so, $r_1, r_n \in D$, and $r_i \notin D$ for $i = 2, 3, \ldots, n-1$. See Figure G.3.

Let $L = L_n$ be the $n \times n$ matrix given in (1) and $\Sigma(L)$ the subshift defined by it. An element $s \in \Sigma(L)$ will be used to define an α-pseudo-orbit, and so by the shadowing lemma, a ψ-orbit. The pseudo-orbit corresponding to $s =$

$\ldots s_{-1}s_0 \cdot s_1 \ldots$ is $p(s) = \{r_{s_i}\} = \ldots, r_{s_{-1}}, r_{s_0}, r_{s_1}, \ldots$. This encoding gives only one option: if the pseudo-orbit is at r_n, it can either jump to r_1 or skip in place, i.e., r_n may be followed by r_n itself. Since r_1, r_n, and r_{n+1} are all within the disk D which has radius $\alpha/2$, this jump is at most a jump by a distance α. Every other point r_i, $i \neq n$, on the pseudo-orbit must be followed by $r_{i+1} = \psi(r_i)$, and so there is no jump there.

By the shadowing lemma, there is a unique orbit $\{\psi(y)\}$ which β-shadows the pseudo-orbit $p(s)$. Let $h(s) = y$; so, h: $\Sigma(L) \to \mathbb{R}^2$. It follows from the proof of the shadowing lemma that h is continuous. To see that h is one-to-one, let $h(s) = y$ and $h(s') = y'$, where $s \neq s'$. Since s and s' are different in some entry, and in particular one must be at q for some j, say $p(s)_j = q$, and the other is not, $p(s')_j \neq q$. By construction, $\text{dist}(p(s)_j, p(s')_j) > 4\beta$, and so $\text{dist}(h(s)_j, h(s')_j) > 2\beta$, and so $h(s) \neq h(s')$. Thus, h is a continuous, one-to-one mapping of a compact Hausdorff space and, thus, is a homeomorphism.

Let $s \in \Sigma(L)$ and $h(s) = y$. By the above construction, the ψ-orbit of $\psi(y)$ and $h(\sigma(s))$ are β-shadows of each other, and so by uniqueness are equal, thus $\psi \circ h = h \circ \sigma$, and ψ and σ are equivalent. ∎

H. Further Reading

The subject of hyperbolic dynamical systems has been the object of a great deal of study, especially in the last 20 years. Some standard references for the local theory (the stable manifold theorem, etc.) are Coddington and Levinson (1955), Hale (1972), or Hartman (1964). Some basic references for the global theory are Nitecki (1972), Szlenk (1981), and Palis and de Melo (1980). Palis and de Melo (1980) is by far the easiest to read since it gives a good presentation of the basic geometric ideas behind the proofs. Szlenk (1981) is well written and contains a wealth of material. Nitecki (1972) is dated, and sometimes the proofs are a little sketchy.

Siegel and Moser (1971) has some of the material on the local theory of Hamiltonian systems. Their approach is very analytic and the text is a well written. Arnold (1978) is a lively introduction to the geometric theory with lots of intuition. However, Arnold's book lacks details at critical points in the development. Abraham and Marsden (1978) has a great deal on symplectic geometry, which makes it hard to read. (The book is well written—the material is just hard.) It does not cover the more basic material as given in this chapter. Those who like to read tough books should read Wintner (1941). It still contains a lot of information that cannot be found anywhere else.

The material on the shadowing lemma and the Smale–Conley theorem is taken from Meyer and Sell (1989). The complete theorem on reduction can be found in Meyer (1973).

Appendix. Proof of Shadowing Lemma

Use the notation of Section G. It can be shown that there is no loss of generality in assuming that $K = 1$ because this can be accomplished by renorming. Henceforth, assume that $K = 1$. Moreover, assume that all orbits of ψ are bounded. There is no loss of generality in these two assumptions—see Problems Section.

First, reformulate the theorem using some ideas from functional analysis —those unfamiliar with functional analysis may wish to skip to the next chapter. Let $\mathscr{L}$ be the space of all bounded bi-infinite sequences $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$, $x \in \mathbb{R}^m$, with norm $\|x\| = \sup_k |x_k|$, where $|\cdot|$ is the norm in $\mathbb{R}^m$. $(\mathscr{L}, \|\cdot\|)$ is a Banach space—it is essentially ℓ_∞. Define $\mathscr{F}\mathscr{L} \to \mathscr{L}$ by $\mathscr{F}(x_k)_k = \psi(x_{k-1})$. A fixed point of $\mathscr{F}$ is a ψ-orbit since $\mathscr{F}(x) = x$ means $x_k = \psi(x_{k-1})$ for all k. $x \in \mathscr{L}$ is an α-pseudo-orbit, if $\|\mathscr{F}(x) - x\| \leq \alpha$ and $x \in \mathscr{L}$ β-shadows an orbit $y \in \mathscr{L}$ if $\|x - y\| \leq \beta$. Thus, the shadowing lemma says that an almost fixed point for $\mathscr{F}$ in a hyperbolic set is close to a true fixed point.

Let $\mathscr{B}(\alpha)$ denote the set of α-pseudo-orbits of ψ in Λ. The reformulated shadowing lemma is: For every $\beta > 0$, there is an $\alpha > 0$ such that if $x \in \mathscr{B}(\alpha)$, then $\mathscr{F}$ has a fixed point y in a β neighborhood of x. This formulation suggests the use of the implicit function theorem in a Banach space with estimates. Thus, some estimates on the derivative of $\mathscr{F}$ are necessary.

Since $\mathscr{L}$ is a linear space, it is its own tangent space, and so its tangent bundle is $\mathscr{L} \times \mathscr{L}$. By using the splitting of the tangent spaces over Λ, there follows a splitting of the tangent bundle of $\mathscr{B}(\alpha)$ in the form

$$\mathscr{L} \times \mathscr{B}(\alpha) = \mathscr{S} \oplus \mathscr{U}, \tag{1}$$

where

$$\begin{aligned} \mathscr{S} &= \bigcup_x \mathscr{S}(x), \qquad \mathscr{S}(x) = \bigoplus_i \mathbb{E}^s(x_i), \\ \mathscr{U} &= \bigcup_x \mathscr{U}(x), \qquad \mathscr{U}(x) = \bigoplus_i \mathbb{E}^u(x_i). \end{aligned} \tag{2}$$

What this means is that if $x \in \mathscr{B}(\alpha)$ and $w \in \mathscr{L}$, then there is a unique decomposition $w = u \oplus v$ with $u \in \mathscr{S}(x)$ and $u \in \mathscr{U}(x)$. Indeed, one has $u_i = P(x_i)w_i$ and $v_i = (I - P(x_i))w_i$. Since the projector P is uniformly bounded on the compact set Λ, the norm $\|\cdot\|'$ defined by

$$\|w\|' = \|u\| + \|v\| \tag{3}$$

is equivalent to the original norm $\|\cdot\|$. The equivalence of the two norms means there is a constant $K_1 > 0$ such that

$$K_1^{-1}\|w\|' < \|w\| < K_1\|w\|'. \tag{4}$$

For simplicity, define $\mathscr{G}(x) = \mathscr{F}(x) - x$; so, seek a zero of $\mathscr{G}$. The function $\mathscr{G}$ is differentiable with derivative $D\mathscr{G}(x)$: $\mathscr{L} \to \mathscr{L}$, which is a linear map for each $x \in \mathscr{L}$ and is given by

$$(D\mathscr{G}(x)w)_i = D\psi(x_{i-1})w_{i-1} - w_i = Y(x_{i-1}, 1)w_{i-1} - w_i. \tag{5}$$

Lemma 1. *There is an* $\alpha' > 0$ *and a* K_2 *such that*

$$\|D\mathcal{G}(x)\| < K_2 \quad \text{and} \quad \|D\mathcal{G}(x)^{-1}\| < K_2 \tag{6}$$

for all $x \in \mathcal{B}(\alpha')$.

PROOF. The first estimate is trivial because $D\psi$ is uniformly bounded on the compact set Λ. In order to verify the second inequality, it suffices to show that $D\mathcal{G}$ has a uniform lower bound for all $x \in \mathcal{B}(\alpha')$. This will be verified by using the equivalent norm $\|w\|' = \|u\| + \|v\|$.

The first step is to show that there is an $a > 0$ such that

$$\begin{aligned} a\|u\| &\le \|D\mathcal{G}(x)u\|, \quad u \in \mathcal{S}(x), \\ a\|v\| &\le \|D\mathcal{G}(x)v\|, \quad v \in \mathcal{U}(x), \end{aligned} \tag{7}$$

provided $x \in \mathcal{B}(\alpha')$ and α' is sufficiently small. Using (G.6) with $K = 1$ and (5) and $x \in \mathcal{B}(\alpha)$ and $u \in \mathcal{S}(x)$, it follows that

$$\begin{aligned} |(D\mathcal{G}(x)u)_i| &= |D\psi(x_{i-1})u_{i-1} - u_i| \ge |u_i| - |D\mathcal{G}(x)u_{i-1}| \\ &\ge |u_i| - \mu|u_{i-1}| \ge |u_i| - \mu\|u\|. \end{aligned} \tag{8}$$

Hence, $|(D\mathcal{G}(x)u)_i| + \mu\|u\| \ge |u_i|$, which implies $\|D\mathcal{G}(x)u\| \ge (1 - \mu)\|u\|$. Similarly if $v \in \mathcal{U}(x)$, then

$$\begin{aligned} |(D\mathcal{G}(x)v)_i| &= |D\psi(x_{i-1})v_{i-1} - v_i| \ge |D\mathcal{G}(x)v_{i-1}| - |v_i| \\ &\ge \mu^{-1}|v_{i-1}| - |v_i| \ge \mu^{-1}|v_{i-1}| - \|v\|. \end{aligned} \tag{9}$$

Hence, $|(D\mathcal{G}(x)u)_i| + \|v\| \ge \mu^{-1}|v_{i-1}|$, which implies $\|D\mathcal{G}(x)v\| \ge (\mu^{-1} - 1)\|v\|$. Thus, (7) is established.

The next step is to show that there is a constant $K_3 > 0$ such that for every $\varepsilon > 0$ there is an $\alpha > 0$ satisfying

$$\|D\mathcal{G}(x)u\| + \|D\mathcal{G}(x)v\| \le \|D\mathcal{G}(x)w\|' + K_3\varepsilon\|w\|' \tag{10}$$

for all $x \in \mathcal{B}(\alpha)$. Since the projector P is continuous for $\varepsilon > 0$, there is an $\alpha = \alpha(\varepsilon)$ so that $\|P(x_1) - P(x_2)\| \le \varepsilon$ whenever $|x_1 - x_2| \le \alpha$. If $x \in \mathcal{B}(\alpha)$, then $|\psi(x_{i-1}) - x_i| \le \alpha$ for all i and, consequently,

$$\|P(\psi(x_{i-1})) - P(x_i)\| \le \varepsilon \tag{11}$$

for all i.

Now using (5) and (G.5)

$$\begin{aligned} (D\mathcal{G}(x)u)_i &= Y(x_{i-1}, 1)u_{i-1} - u_i \\ &= Y(x_{i-1}, 1)P(x_{i-1})w_{i-1} - P(x_i)w_i \\ &= P(\psi(x_{i-1}))\,Y(x_{i-1}, 1)w_{i-1} - P(x_i)w_i \\ &= P(x_i)[Y(x_{i-1}, 1)w_{i-1} - w_i] \\ &\quad + [P(\psi(x_{i-1})) - P(x_i)]\,Y(x_{i-1}, 1)w_{i-1} \\ &= P(x_i)(D\mathcal{G}(x)w)_i + [P(\psi(x_{i-1})) - P(x_i)]\,Y(x_{i-1}, 1)w_{i-1}. \end{aligned} \tag{12}$$

This with (7) gives

$$\|D\mathcal{G}(x)u\| \leq \sup_i |P(x_i)(D\mathcal{G}(x)w)_i| + K_4\varepsilon\|w\|, \tag{13}$$

where $K_4 = \sup\{\|Y(x_0, 1)\|: x_0 \in \Lambda\}$. Similarly, one obtains

$$\|D\mathcal{G}(x)v\| \leq \sup_i |[I - P(x_i)](D\mathcal{G}(x)w)_i| + K_4\varepsilon\|w\|. \tag{14}$$

By adding these last two inequalities and using (3) and (4), one gets (10) with $K_3 = 2K_4K_1$. The inequalities (7) and (11) imply

$$a\|w\|' \leq \|D\mathcal{G}(x)w\|' + K_3\varepsilon\|w\|'. \tag{15}$$

By fixing ε so that $0 < K_3\varepsilon < a$ and setting $\alpha' = \alpha(\varepsilon)$, one obtains

$$(a - K_3\varepsilon)\|w\|' \leq \|D\mathcal{G}(x)w\|' \quad \text{for } x \in \mathcal{B}(\alpha'), \tag{16}$$

which completes the proof of the lemma. ∎

The hyperbolic structure is stable under perturbation and so can be extended to a neighborhood of Λ. So by the above, the inequalities in (7) can be extended to a neighborhood of $\mathcal{B}(\alpha')$. That is, there is a $\delta' > 0$ and a K_4 such that

$$\|D\mathcal{G}(x)\| < K_4 \quad \text{and} \quad \|D\mathcal{G}(x)^{-1}\| < K_4 \tag{17}$$

for all $x \in \mathcal{L}$ with $\operatorname{dist}(x, \mathcal{B}(\alpha')) \leq \delta'$.

The following version of the inverse function theorem with estimate can be found in Hartman (1964).

Theorem 2 (The Inverse Function Theorem). *Let $\mathcal{X}$ be a Banach space, $B_\delta(x_0)$ the ball of radius δ about x_0, and $\mathcal{H}: B_\delta(x_0) \to \mathcal{X}$ a C^1 function with $y_0 = \mathcal{H}(x_0)$. Assume that $D\mathcal{H}(x)$ has a bounded inverse with*

$$\|D\mathcal{H}(x)\| \leq K \quad \text{and} \quad \|D\mathcal{H}^1(x)\| \leq K \tag{18}$$

for all $x \in B_\delta(x_0)$. where K is a constant. Let $\rho = \delta/K^2$ and $\sigma = \delta/K$. There exists a domain Ω, with $B_\rho(x_0) \subset \Omega \subset B_\delta(x_0)$, such that $\mathcal{H}$ is one-to-one on Ω. Moreover, $B_\sigma(y_0) \subset \mathcal{H}(\Omega)$. In particular, for every $y \in B_\sigma(y_0)$, there is a unique $x \in \Omega$ with $y = \mathcal{H}(x)$, and the mapping $x = \mathcal{H}^{-1}(y)$ is continuous on $B_\sigma(y_0)$.

Apply this version of the inverse function theorem to the function $\mathcal{G}$. For $\beta > 0$ define

$$\alpha = \min(\beta/2K_4, \alpha'), \quad \delta = \min(\beta, \delta'). \tag{19}$$

Let x be any fixed α-pseudo-orbit and $y_0 = \mathcal{G}(x)$. Then $\|y_0\| = \|\mathcal{G}(x)\| \leq \alpha$. If $x' \in B_\delta(x)$, then $\|x - x'\| < \delta \leq \delta_0$. Hence, (17) holds for all $x' \in B_\delta(x)$. Let $\sigma = \delta/K_4$ be given by Theorem 2. Then $\sigma = 2\alpha$, and consequently, $0 \in B_\sigma(y_0)$. By Theorem 2, there is a $y \in B_\delta(x)$ with $\mathcal{G}(y) = 0$, i.e,, y is a ψ-orbit. Furthermore, $\|y - x\| < \delta \leq \beta$; so, y is a β-shadow of x.

If $\beta_0 = \delta_0/K_4^2$, then Theorem 2 says that $\mathcal{G}$ is one-to-one on $B_{\delta_0}(x_0)$ for $x \in \mathcal{B}(\alpha')$. Thus, the distance between zeros of $\mathcal{G}$ is at least β_0, and this proves the uniqueness part of the shadowing lemma. ∎

A dynamical system $\phi: X \to X$, where (X, d) is a metric space, is *expansive*, if there is a constant $\eta > 0$ such that if $u \neq v$, then for some $n \in \mathbb{Z}$, $d(\phi^n(u), \phi^n(v)) \geq \eta$. By the uniformity of the above estimate on the distance between zeros of $\mathcal{G}$, one has that if x and x' are distinct ψ-orbits in Λ, then $\|x - x'\| \geq \beta_0$. This implies:

Corollary 3. *A dynamical system on a hyperbolic set is expansive.*

Problems

1. Let $\{\phi_t\}$ be a smooth dynamical system, i.e., $\{\phi_t\}$ satisfies (A.4). Prove that $\phi(t, \xi) = \phi_t(\xi)$ is the general solution of an autonomous differential equation.

2. Let ψ be a diffeomorphism of $\mathbb{R}^m$; so, it defines a discrete dynamical system. A non-fixed-point is called an *ordinary point*. So $p \in \mathbb{R}^m$ is an ordinary point if $\psi(p) \neq p$. Prove that there are local coordinates x at an ordinary point p and coordinates y at $q = \psi(p)$ such that in these local coordinates $y_1 = x_1, \ldots, y_m = x_m$. (This is the analog of the flow box theorem for discrete systems.)

3. Let ψ be as in Problem 2. Let p be a fixed point p of ψ. The eigenvalues of $\partial\psi(p)/\partial x$ are called the (*characteristic*) *multipliers* of p. If all the multipliers are different from $+1$, then p is called an *elementary fixed point of* ψ. Prove that elementary fixed points are isolated.

4. **a.** Let $0 < a < b$ and $\xi \in \mathbb{R}^m$ be given. Show that there is a smooth non-negative function $\gamma: \mathbb{R}^m \to \mathbb{R}$ which is identically $+1$ on the ball $\|x - \zeta\| \leq a$ and identically zero for $\|x - \xi\| \geq b$.
   b. Let $\mathbb{O}$ be any closed set in $\mathbb{R}^m$. Show that there exists a smooth, non-negative function $\delta: \mathbb{R}^m \to \mathbb{R}$ which is zero exactly on $\mathbb{O}$.

5. Let $H(q_1, \ldots, q_N, p_1, \ldots, p_N)$, q_i, $p_i \in \mathbb{R}^3$, be invariant under translation; so, $H(q_1 + s, \ldots, q_N + s, p_1, \ldots, p_N) = H(q_1, \ldots, q_N, p_1, \ldots, p_N)$ for all $s \in \mathbb{R}^3$. Show that total linear momentum, $L = \sum p_i$, is an integral. This is another consequence of the general Noether theorem.

6. An $m \times m$ nonsingular matrix T such that $T^2 = I$ is a *discrete symmetry of* (or a *reflection for*) $\dot{x} = f(x)$ if and only if $(Tx) = -Tf(x)$ for all $x \in \mathbb{R}^m$. The equation is also called a *reversible system* in this case.
   a. Prove: If T is a discrete symmetry of (1), then $\phi(t, T\xi) \equiv T\phi(-t, \xi)$ where $\phi(t, \xi)$ is the general solution of $\dot{x} = f(x)$.
   b. Consider the 2×2 case and let $T = \text{diag}(1, -1)$. What does $f(Tx) = -Tf(x)$ mean about that the parity of f_1 and f_2? Show that Part **a** means that a reflection of a solution in the x_1 axis is a solution.

7. Let T be a discrete symmetry of Equation (D.1). Let $FIX = \{x \in \mathbb{R}^m: Tx = x\}$.

Show that if $\phi(t)$ is a solution of (1) with $\phi(0) \in FIX$ and $\phi(\tau) \in FIX$ for some $\tau \neq 0$, then ϕ is 2τ-periodic.

8. Let $\mathcal{G}$ be a matrix Lie group, i.e., $\mathcal{G}$ is a closed subgroup of the general group $Gl(m, \mathbb{R})$. (See the problems at the end of Chapter II.) $\mathcal{G}$ is a *symmetry group for* $\dot{x} = f(x)$ if $Tf(x) = f(Tx)$ for al $T \in \mathcal{G}$ and $x \in \mathbb{R}^m$.

a. Prove: If $\mathcal{G}$ is a symmetry group for (1), then $\phi(t, T\xi) \equiv T\phi(t, \xi)$, where ϕ is the general solution of $\dot{x} = f(x)$.

b. Consider the 2×2 case where $\mathcal{G}$ is $SO(2, \mathbb{R})$ the group of rotations of the plane (i.e., orthogonal matrices with determinant $+1$). In polar coordinates (r, θ), $\dot{x} = f(x)$ becomes $\dot{r} = R(r, \theta)$, $\dot{\theta} = \Theta(r, \theta)$. Prove that the symmetry condition implies R and Θ are independent of θ.

9. Now let $\dot{x} = f(x)$ be Hamiltonian with Hamiltonian $H: \mathbb{R}^{2n} \to \mathbb{R}$; so, $f(x) = J\nabla H(x)$. A matrix T is *antisymplectic* if $T^TJT = -J$. An antisymplectic matrix T such that $T^2 = I$ is a *discrete symplectic symmetry for* H if $H(Tx) \equiv H(x)$.

a. Prove: A discrete symplectic symmetry of the Hamiltonian is a discrete symmetry of the equation $\dot{x} = f(x)$.

b. Consider a general Newtonian system as discussed in (I.B.4) of the form $H(x, p) = \frac{1}{2}p^TM^{-1}p + F(x)$ where $x, p \in \mathbb{R}^m$ and M is a nonsingular, symmetric matrix. Define $T = \text{diag}(I, -I)$; so, $T^2 = I$, show that T is antisymplectic and $H(T(x, p)) = H(x, p)$.

c. Consider the restricted 3-body problem as discussed in (I.C.4). Let $T = \text{diag}(1, -1, -1, 1)$; show $H(T(x, y)) = H(x, y)$ where H is the Hamiltonian of the restricted 3-body problem (I.C.9).

d. What is FIX of Problem 7 for these two examples?

10. Use Problems 7 and 9.

a. Show that a solution of the restricted problem which crosses the x_1 axis (the line of syzygy) at a time t_1 and later at a time t_2 is of period $2(t_2 - t_1)$.

b. Show that the above criterion is $x_2 = y_1 = 0$ at times t_1 and t_2 in rectangular coordinates, and $\theta = n\pi$ (n an integer), $R = 0$ in polar coordinates.

11. Let $\mathcal{G}$ be a matrix Lie group of symplectic matrices, i.e., $\mathcal{G}$ is a subgroup of the symplectic group $Sp(n, \mathbb{R})$. Let $\dot{x} = f(x)$ be Hamiltonian with Hamiltonian H. $\mathcal{G}$ is a symmetry group for the Hamiltonian H if $H(Tx) = H(x)$ for all $T \in \mathcal{G}$. Prove: A symmetry group for the Hamiltonian H is a symmetry group for the equations of motion.

12. Prove that the tangent spaces to points in a hyperbolic set can be renormed so that the constant K in (G.9) can be taken as 1. [Hint: If $w \in T_p\Lambda$, Λ hyperbolic, then $w = u + v$, $u \in \mathbb{E}^s_p$ and $v \in \mathbb{E}^u_p$. Define the norm $\|\cdot\|'_p$ in $T_p\Lambda$ by $\|w\|'_p = \max(\|u\|_p, \|v\|_p)$, where $\|u\|_p = \sup\{K^{-1}\mu^k \| Y(p, k)u\| : k \geq 0\}$ and $\|v\|_p = \sup\{K^{-1}\mu^k \| Y(p, k)v\| : k \leq 0\}$.]

13. Prove that the closure of the orbit of a transverse homoclinic point is a hyperbolic set. (For differential equationists only.)

14. Prove that the inequalities (G.5) and (G.6) defining a hyperbolic structure on a compact invariant set Λ can be extended to a neighborhood of Λ. [This requires a contracting mapping proof. Model your proof on the proof for a hyperbolic structure being stable under small perturbations. See Szlenk (1981).]

CHAPTER VI

Continuation of Periodic Solutions

In the last chapter, some local results about periodic solutions of Hamiltonian systems were presented. The systems contain a parameter, and the conditions under which a periodic solution can be continued in the parameter were discussed. Since Poincaré used these ideas extensively, it has become known as Poincaré's continuation method. By Lemma V.E.2, a solution $\phi(t, \xi')$ of an autonomous differential equation is T-periodic if and only if $\phi(T, \xi') = \xi'$, where ϕ is the general solution. This is a finite-dimensional problem since ϕ is a function defined in a domain of $\mathbb{R}^{m+1}$ into $\mathbb{R}^m$. Thus, periodic solutions can be found by the finite-dimensional methods, i.e., the finite-dimensional implicit function theorem, the finite-dimensional fixed point theorems, the finite-dimensional degree theory, etc. This chapter will present results which depend only on the finite-dimensional implicit function theorem. Chapter X will present a treatment of fixed point methods as they apply to Hamiltonian systems. In this chapter the periodic solutions vary continuously with the parameter ("can be continued"), but Chapter VII will discuss the bifurcations of periodic solutions.

After some elementary general results a variety of families of periodic solutions are given in the 3-body problem and the restricted problem. The first result is a simple proof of the Lyapunov Center Theorem with applications to the five libration points in the restricted problem. Then in the next three sections, the circular orbits of Kepler's problem are continued into the restricted problem when one mass is small (Poincaré solutions), when the infinitesimal is near a primary (Hill's solution), and when the infinitesimal is near infinity (comet solutions). Lastly, a general theorem on the continuation of periodic solutions from the restricted problem to the full 3-body problem is given.

A. Continuation of Equilibrium Points and Periodic Solutions

Assume that the differential equations depend on some parameters; so, consider

$$\dot{x} = f(x, v), \tag{1}$$

where $f\colon O \times Q \to \mathbb{R}^m$ is smooth, O is open in $\mathbb{R}^m$, and Q is open in $\mathbb{R}^k$. Let ξ' be an equilibrium point when $v = v'$ $[f(\xi', v') = 0]$, a *continuation* of this equilibrium point is a smooth function $u(v)$ defined for v near v' such that $u(v') = \xi'$, and $u(v)$ is an equilibrium point for all v $[f(u(v), v) = 0]$. The general solution $\phi(t, \xi, v)$ is smooth in the parameter v also. Let the solution $\phi(t, \xi', v')$ be T-periodic. A *continuation* of this periodic solution is a pair of smooth functions, $u(v)$, $\tau(v)$, defined for v near v' such that $u(v') = \xi'$, $\tau(v') = T$, and $\phi(t, u(v), v)$ is $\tau(v)$-periodic. One also says that the periodic solution *can be continued.* This means that the solution persists when the parameters are varied, and the periodic solution does not change very much with the parameters.

Recall (see Section V.E) that an equilibrium point ξ' for (1) when $v = v'$ $[f(\xi', v') = 0]$ is *elementary if* $\partial f(\xi', v')/\partial \xi$ is nonsingular, i.e., if zero is not an exponent. The solution $\phi(t, \xi', v')$ is T-periodic if and only if $\phi(T, \xi', v') = \xi'$. This periodic solution is *elementary* if $+1$ is an eigenvalue of the monodromy matrix $\partial \phi(T, \xi', v')/\partial \xi$ with multiplicity 1 for a general autonomous differential equation and of multiplicity 2 for a system with a nondegenerate integral (e.g., a Hamiltonian system). Recall that the eigenvalues of $\partial \phi(T, \xi', v')/\partial \xi$ are called the *multipliers* (*of the periodic solution*). Drop one $+1$ multiplier for a general equation, and drop two $+1$ multipliers from the list of multipliers to get the *nontrivial multipliers.*

Proposition 1. *An elementary equilibrium point or an elementary periodic solution or an elementary periodic solution in a system with a nondegenerate integral can be continued.*

Proof. For equilibrium points, apply the implicit function theorem to $f(x, v) = 0$. By assumption, $f(\xi', v') = 0$, and $\partial f(\xi', v')/\partial x$ is nonsingular; so, the implicit function theorem asserts the existence of the function $u(v)$ such that $u(v') = \xi'$ and $f(u(v), v) \equiv 0$.

Since the existence of the first return time, and the Poincaré map depended on the implicit function theorem these functions depend smoothly on the parameter v. For the rest of the proposition, apply the implicit function theorem to $P(x, v) - x = 0$, where $P(x, v)$ is the Poincaré map of the cross section to the periodic solution when $v = v'$.

Similarly, if the system has an integral, $I(x, v)$, then the construction of the Poincaré map in an integral surface depends smoothly on v. Again apply the implicit function theorem to the map $Q(x, v) - x = 0$, where $Q(x, v)$ is the

Poincaré map in the integral surface of the cross section to the periodic solution when $v = v'$. ∎

There is a similar definition of continuation and a similar lemma for fixed points.

Corollary 2. *The exponents of an elementary equilibrium point, and the multipliers of an elementary periodic solution (with or without nondegenerate integral) vary continuously with the parameter v.*

PROOF. For equilibrium points, the implicit function theorem was applied of $f(x, v) = 0$ to get a function $u(v)$ such that $u(v') = \xi'$ and $f(u(v), v) \equiv 0$. The exponents of the equilibrium $u(v)$ are the eigenvalues of $\partial f(u(v), v)/x$. This matrix varies smoothly with the parameter v, and so its eigenvalues vary continuously with the parameter v. (See Problems Section for an example where the eigenvalues are not smooth in a parameter.) The other parts of the theorem are proved using the same idea applied to the Poincaré map. ∎

Corollary 3. *A small perturbation of an elliptic (respectively a hyperbolic) periodic orbit of a Hamiltonian system of two degrees of freedom is elliptic (respectively hyperbolic).*

PROOF. If the system has two degrees of freedom, then a periodic solution has as multipliers $+1, +1, \lambda, \bar{\lambda} = \lambda^{-1}$, and so the multipliers lie either on the real axis or the unit circle. If the periodic solution is hyperbolic, then λ and λ^{-1} lie on the real axis and are not 0 or ± 1. A small change cannot make these eigenvalues lie on the unit circle or take the value 0 or ± 1. Thus, a small change in a hyperbolic periodic solution is hyperbolic. A similar argument holds for elliptic periodic solutions. ∎

B. Lyapunov's Center Theorem

An immediate consequence of the discussion in the previous section is the following celebrated theorem.

Theorem 1 (The Lyapunov Center Theorem). *Assume that a system with a nondegenerate integral has an equilibrium point with exponents $\pm\omega i, \lambda_3, \ldots, \lambda_m$, where $i\omega \neq 0$ is pure imaginary. If $\lambda_j/i\omega$ is never an integer for $j = 3, \ldots, m$, then there exists a one-parameter family of periodic orbits emanating from the equilibrium point. Moreover, when approaching the equilibrium point along the family, the periods tend to $2\pi/\omega$ and the nontrivial multipliers tend to $\exp(2\pi\lambda_j/\omega), j = 3, \ldots, m$.*

Remark. The Hamiltonian is always a nondegenerate integral for a nonconstant periodic solution.

PROOF. Say that $x = 0$ is the equilibrium point, and the equation is

$$\dot{x} = Ax + g(x), \tag{1}$$

where $g(0) = \partial g(0)/\partial x = 0$. Since we seek periodic solution near the origin, scale by $x \to \varepsilon x$ where ε is to be considered as a small parameter. The equations become

$$\dot{x} = Ax + O(\varepsilon), \tag{2}$$

and when $\varepsilon = 0$, the system is linear. Since this linear system has exponents $\pm\omega i$, it has a periodic solution of period $2\pi/\omega$ of the form $\exp(At)a$, where a is a fixed nonzero vector. The multipliers of this periodic solution are the eigenvalues of $\exp(A2\pi/\omega)$, or 1, 1, $\exp(2\pi\lambda_j/\omega)$. By assumption, the nontrivial multipliers are not $+1$, and so this periodic solution is elementary. From Proposition A.1, there is a periodic solution of the form $\exp(At)a + O(\varepsilon)$. In the unscaled coordinates, the solution is of the form $\varepsilon \exp(At)a + O(\varepsilon^2)$, and the result follows. ∎

1. Applications to the Euler and Lagrange Libration Points

In Section II.G, the linearized equations at the five libration (equilibrium) points of the restricted 3-body problem were analyzed. The eigenvalues at the three collinear libration points of Euler were shown to be a pair of real eigenvalues and a pair of pure imaginary eigenvalues. Thus, Lyapunov's theorem implies that there is a one-parameter family of periodic solutions emanating from each of these libration points.

By symmetry we may assume that $0 < \mu \leq \frac{1}{2}$. At the equilateral triangle libration points of Lagrange, the characteristic equation of the linearized system was found to be

$$\lambda^4 + \lambda^2 + \frac{27}{4}\mu(1 - \mu), \tag{3}$$

where μ is the mass ratio parameter. The roots of (3) satisfy

$$\lambda^2 = \tfrac{1}{2}(-1 \pm \sqrt{1 - 27\mu(1 - \mu)}), \tag{4}$$

which implies that for $0 < \mu < \mu_1 = (1 - \sqrt{69}/9)/2 \simeq 0.0385$, the eigenvalues are distinct pure imaginary numbers $\pm\omega_1 i$, $\pm\omega_2 i$, with $0 < \omega_2 < \omega_1$. Since $i\omega_2/i\omega_1$ is less than 1 in modulus, Lyapunov's theorem implies that there is a family of periodic orbits emanating from $\mathscr{L}_4$ with period approaching $2\pi/\omega_1$ for all μ, $0 < \mu < \mu_1$. This family is called the *short period family*.

Define μ_r to be the value of μ for which $\omega_1/\omega_2 = r$. If $0 < \mu < \mu_1$ and $\mu \neq \mu_n$, $n = 1, 2, \ldots$, then Lyapunov's theorem implies that there is a family

of periodic orbits emanating from $\mathscr{L}_4$ with period approaching $2\pi/\omega_2$. This family is called the *long period family.*

The mass ratios μ_r satisfies

$$\mu_r = \frac{1}{2} - \frac{1}{2}\left\{1 - \frac{16r^2}{27(r^2+1)^2}\right\}^{1/2}, \tag{5}$$

so, $0 \cdots < \mu_3 < \mu_2 < \mu_1$.

C. Poincaré's Orbits

The essence of the continuation method is that the problem contains a parameter, and for one value of the parameter, there is a periodic solution whose multipliers can be computed. The restricted 3-body problem has a parameter, μ, the mass ratio parameter, and when $\mu = 0$, the problem is just the Kepler problem in rotating coordinates. The Kepler problem has many periodic solutions, but they all have their multipliers equal to $+1$ in fixed coordinates, whereas the circular orbits have nontrivial multipliers in rotating coordinates. Thus, the circular solutions of the Kepler problem can be continued into the restricted problem for small values of μ.

The reason that all the multipliers are $+1$ for the Kepler problem in fixed coordinates is that all the periodic solutions in any energy level have the same period and so are not isolated in an energy level (see Proposition V.E.6).

The Hamiltonian of the restricted problem (I.C.9) is

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{\mu}{d_1} - \frac{1-\mu}{d_2}, \tag{1}$$

where $d_1^2 = (x_1 - 1 + \mu)^2 + x_2^2$, $d_2^2 = (x_1 + \mu)^2 + x_2^2$, and $K = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Recall that $x^T K y$ is just angular momentum. Consider μ as a small parameter; so, the Hamiltonian is of the form

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{1}{\|x\|} + O(\mu). \tag{2}$$

Be careful of the $O(\mu)$ term because it has terms which go to infinity near the primaries; therefore, a neighborhood of the primaries must be excluded. When $\mu = 0$, this is the Kepler problem in rotating coordinates. Put this problem in polar coordinates (see Section IV.C) to get (when $\mu = 0$)

$$H = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \Theta - \frac{1}{r}, \tag{3}$$

$$\begin{aligned} \dot{r} &= R, & \dot{R} &= \frac{\Theta^2}{r^3} - \frac{1}{r^2}, \\ \dot{\theta} &= \frac{\Theta}{r^2} - 1, & \dot{\Theta} &= 0. \end{aligned} \tag{4}$$

Θ, angular momentum, is an integral; so, let $\Theta = c$ be a fixed constant. For $c \neq 1$, the circular orbit $R = 0$, $r = c^2$ is a periodic solution with period $|2\pi c^3/(1 - c^3)|$ (this is the time for θ to increase by 2π). Linearizing the r and R equations about this solution gives

$$\dot{r} = R, \qquad \dot{R} = -c^{-6}r, \tag{5}$$

which has solutions of the form $\exp(\pm it/c^3)$, and so the nontrivial multipliers of the circular orbits are $\exp(\pm i2\pi/(1 - c^3))$ which are not $+1$, provided $1/(1 - c^3)$ is not an integer. Thus, we have proved:

Theorem 1 (Poincaré). *If $c \neq 1$ and $1/(1 - c^3)$ is not an integer, then the circular orbits of the Kepler problem in rotating coordinates with angular momentum c can be continued into the restricted problem for small values of μ. These orbits are elliptic.*

The rotating coordinates used here rotate counterclockwise, and so in fixed coordinates the primaries rotate clockwise. If $c < 0$, then $\dot{\theta} < 0$, and $1/(1 - c^3)$ is never an integer. Orbits with these angular momentum rotate clockwise in either coordinate system and so are called *retrograde orbits.*

If $c > 0$, $c \neq 1$, and $1/(1 - c^3)$ is not an integer, then in the fixed coordinates, these orbits rotate counterclockwise and so are called *direct orbits.* The circular orbits of the Kepler problem when $1/(1 - c^3)$ is an integer, say k, undergo a bifurcation when $\mu \neq 0$, but this is too lengthy a problem to be discussed here.

D. Hill's Orbits

Another way to introduce a small parameter is to consider the case when the infinitesimal particle is very near one of the primaries. This is usually referred to as Hill's problem because he extensively investigated the motion of the moon, which to a first approximation is like this problem.

Consider the restricted problem where one primary is at the origin, i.e., replace x_1 by $x_1 - \mu$ and y_2 by $y_2 + \mu$; so, the Hamiltonian (C.1) becomes

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{\mu}{d_1} - \frac{1 - \mu}{d_2} - \mu^2, \tag{1}$$

where $d_1^2 = (x_1 - 1)^2 + x_2^2$, $d_2^2 = x_1^2 + x_2^2$. Introduce a scale parameter ε by changing coordinates by $x = \varepsilon^2 \xi$, $y = \varepsilon^{-1}\eta$, which is a symplectic change of coordinates with multiplier ε^{-1}. In the scaling, all constant terms will be dropped from the Hamiltonian because they do not affect the equations of motion. Note that if $\|\xi\|$ is approximately 1, then $\|x\|$ is about ε^2, or $\|x\|$ is very small when ε is small. Thus, ε is a measure of the distance of the infinitesimal particle from the primary at the origin and so will be consid-

ered as the small parameter. We fix the mass ratio parameter, μ, as arbitrary (i.e., not small), and for simplicity we set $c^2 = 1 - \mu$, $c > 0$. The Hamiltonian becomes

$$H = -\Theta + \varepsilon^3\left[\frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \frac{1}{r}\right] + O(\varepsilon^5), \tag{2}$$

The dominant term is the Hamiltonian of the Kepler problem, and the next most important term is the rotational term; so, this formula says that when the infinitesimal is close to the primary that has mass $c^2 = 1 - \mu$ the main force on it is the gravitational force of the primary that has mass $c^2 = 1 - \mu$. The next most important term is the Coriolis term.

Kepler's third law says that the period of a circular orbit varies with the radius to the 3/2 power; so, time should be scaled by $t \to \varepsilon^{-3}t$ and $H \to \varepsilon^3 H$, and the Hamiltonian is

$$H = \left\{\frac{\|\eta\|^2}{2} - \frac{c^2}{\|\xi\|}\right\} - \varepsilon^3 \xi^T K \eta + O(\varepsilon^4). \tag{3}$$

Introduce polar coordinates as before; so,

$$H = \frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \frac{c^2}{r} - \varepsilon^3\Theta + O(\varepsilon^4), \tag{4}$$

$$\begin{aligned} \dot{r} &= R, & \dot{R} &= \frac{\Theta^2}{r^3} - \frac{c^2}{r^2}, \\ \dot{\theta} &= \frac{\Theta}{r^2} - \varepsilon^3, & \dot{\Theta} &= 0. \end{aligned} \tag{5}$$

In the equations in (5) the terms of order ε^4 have been omitted. Omitting these terms gives a system where Θ is an integral. The two solutions, $\Theta = \pm c$, $R = 0$, $r = 1$, are periodic solutions of (5) of period $2\pi/(c \mp \varepsilon^3)$. Linearizing the r and R equations about this solution gives

$$\dot{r} = R, \qquad \dot{R} = -c^2 r. \tag{6}$$

Equations (6) have solutions of the form $\exp(\pm ict)$, and so the nontrivial multipliers of the circular orbits of (5) are $\exp(\pm ic2\pi/(c \mp \varepsilon^3)) = +1 \pm \varepsilon^3 2\pi i/c + O(\varepsilon^6)$.

Consider the period map in a level surface of the Hamiltonian about this circular orbit. Let u be the coordinate in this surface, with $u = 0$, corresponding to the circular orbit when $\varepsilon = 0$. From the above, the period map has a fixed point at the origin up to terms of order ε^3 and is the identity up to terms of order ε^2, and at ε^3 there is a term whose Jacobian has eigenvalues $\pm 2\pi i/c$. That is, the period map is of the form $P(u) = u + \varepsilon^3 p(u) + O(\varepsilon^4)$, where $p(0) = 0$, and $\partial p(0)/\partial u$ has eigenvalues $\pm 2\pi i/c$; so, in particular, $\partial p(0)/\partial u$ is nonsingular. Apply the implicit function theorem to $G(u, \varepsilon) = (P(u) - u)/\varepsilon^3 = p(u) + O(\varepsilon)$. Since $G(0, 0) = 0$ and $\partial G(0, 0)/\partial u = \partial p(0)/\partial u$, there is a smooth

function $\bar{u}(\varepsilon)$ such that $G(\bar{u}(\varepsilon), \varepsilon) = 0$ for all ε sufficiently small. Thus, the two solutions can be continued from the equations in (5) to the full equations, where the $O(\varepsilon^4)$ terms are included. These solutions are elliptic also.

In the scaled variables, these solutions have $r \simeq 1$ and period $T \simeq 2\pi$. In the original unscaled variables, the periodic solution has $\|x\| \simeq \varepsilon^3$ with period $T \simeq \varepsilon^{-3}$.

Theorem 1. *There exist two one-parameter families of nearly circular elliptic periodic solutions of the restricted* 3*-body problem which encircle a primary for all values of the mass ratio parameter. These orbits tend to the primary.*

E. Comets

Another way to introduce a small parameter is to consider orbits that are close to infinity. In the Hamiltonian of the restricted problem (C.1), scale the variables by $x \to \varepsilon^{-2}x$, $y \to \varepsilon y$; this is symplectic with multiplier ε. The Hamiltonian becomes

$$H = -x^T K y + \varepsilon^3 \left\{ \frac{\|y\|^2}{2} - \frac{1}{\|x\|} \right\} + O(\varepsilon^5). \tag{1}$$

Now ε small means that the infinitesimal is near infinity, and (1) says that near infinity the Coriolis force dominates, and the next most important force looks like a Kepler problem with both primaries at the origin. Again change to polar coordinates to get

$$H = -\Theta + \varepsilon^3 \left[\frac{1}{2}\left(R^2 + \frac{\Theta^2}{r^2} \right) - \frac{1}{r} \right] + O(\varepsilon^5), \tag{2}$$

$$\begin{aligned} \dot{r} &= \varepsilon^3 R, & \dot{R} &= \varepsilon^3 \left(\frac{\Theta^2}{r^3} - \frac{1}{r^2} \right), \\ \dot{\theta} &= -1 + \varepsilon^3 \frac{\Theta}{r^2}, & \dot{\Theta} &= 0. \end{aligned} \tag{3}$$

As before, the terms of order ε^5 have been dropped from the equations in (3), and to this order of approximation, Θ is a integral. A pair of circular periodic solutions of (3) are $\Theta = \pm 1$, $R = 0$, $r = 1$, which are periodic of period $2\pi/(1 \mp \varepsilon^3)$. Linearizing the r and R equations about these solutions gives

$$\dot{r} = \varepsilon^3 R, \qquad \dot{R} = -\varepsilon^3 r. \tag{4}$$

Equations (4) have solutions of the form $\exp(\pm i\varepsilon^3 t)$, and so the nontrivial multipliers of the circular orbits of (3) are $\exp(\pm i\varepsilon^3 2\pi/(1 \mp \varepsilon^3)) = +1 \pm \varepsilon^3 2\pi i + O(\varepsilon^6)$. Repeat the argument given in the last section to continue these solutions into the restricted problem.

Theorem 1. *There exist two one-parameter families of nearly circular large elliptic periodic solutions of the restricted 3-body problem for all values of the mass ratio parameter. These orbits tend to infinity.*

F. Continuation from the Restricted to the Full Problem

In this chapter, four classes of periodic solutions of the restricted problem have been established: Lyapunov solutions at the libration points, Poincaré's orbits of the first kind, Hill lunar orbits, and comet orbits. All of these families are elementary, and most are elliptic. In this section, these solutions and more will be continued into the full 3-body problem, where one of the three particles has small mass.

Periodic solutions of the N-body problem are never elementary because the N-body problem has many symmetries and integrals. As was shown in Section V.E, an integral for the system implies $+1$ as a multiplier of a periodic solution. In fact, the multiplicity of $+1$ as a multiplier of a periodic solution is at least 8 in the planar N-body problem and at least 12 in space. The only way around this problem is to exploit the symmetries and integrals themselves and to go directly to the reduced space as discussed in Section V.D.

A solution of the N-body problem will be called a *periodic solution* if its projection on the reduced space is periodic. Note that it need not be periodic in phase space; in fact, it is not usually. A periodic solution of the N-body problem will be called an *elementary periodic solution*, if its projection on the reduced space is periodic and the multiplicity of the multiplier $+1$ of the periodic solution on the reduced space is exactly 2, i.e., it is elementary on the reduced space.

The main result of this section is the following general theorem.

Theorem 1 (Hadjidemetriou). *Any elementary periodic solution of the planar restricted 3-body problem whose period is not a multiple of 2π can be continued into the full 3-body problem with one small mass.*

The proof is an easy consequence of two procedures that have previously been discussed: the scaling of Section IV.D.3 and the reduction in Section V.D. These facts will be recalled now before the formal proof of this theorem is given.

Recall the scaling given in Section IV.D.3, consider the planar 3-body problem in rotating coordinates with one small particle, $m_3 = \varepsilon^2$. The Hamiltonian is then of the form

$$H_3 = \frac{\|v_3\|^2}{2\varepsilon^2} - u_3^T K v_3 - \sum_{i=1}^{2} \frac{\varepsilon^2 m}{\|u_i - u_3\|} + H_2, \tag{1}$$

where H_2 is the Hamiltonian of the 2-body problem in rotating coordinates. ε

is a small parameter which measures the smallness of one mass. A small mass should make a small perturbation on the other particles; thus, ε should measure the smallness of the mass and how close the two finite particles' orbits are to circular. To accomplish this, use one variable that represents the deviation from a circular orbit.

Let $Z = (u_1, u_2, v_1, v_2)$; so, H_2 is a function of the 8-vector Z. A circular solution of the 2-body problem is a critical point of the Hamiltonian of the 2-body problem in rotating coordinates, i.e., H_2. Let $Z^* = (a_1, a_2, b_1, b_2)$ be such a critical point (later we will specify Z^*). By Taylor's theorem

$$H_2(Z) = H_2(Z^*) + \tfrac{1}{2}(Z - Z^*)^T S(Z - Z^*) + O(\|Z - Z^*\|^3), \tag{2}$$

where S is the Hessian of H_2 at Z^*. Since the equations of motion do not depend on constants, drop the constant term in (2). Change variables by $Z - Z^* = \varepsilon U$, $u_3 = \xi$, $v_3 = \varepsilon^2 \eta$, which gives a symplectic change of variables with multiplier ε^{-2}. The Hamiltonian becomes

$$\begin{aligned} H_3 &= G + \tfrac{1}{2} U^T S U + O(\varepsilon), \\ G &= \frac{\|\eta\|^2}{2} - \xi^T K \eta - \sum_{i=1}^{2} \frac{m_i}{\|\xi - a_i\|}. \end{aligned} \tag{3}$$

G in (3) is the Hamiltonian of the restricted 3-body problem if we take $m_1 = \mu$, $m_2 = 1 - \mu$, $a_1 = (1 - \mu, 0)$, $a_2 = (-\mu, 0)$. (Since it is necessary to discuss several different Hamiltonians in the same section, our usual convention of naming all Hamiltonians H will lead to mass confusion.) The quadratic term above is simply the linearized equations about the circular solutions of the 2-body problem in rotating coordinates. Thus, to first order in ε, the Hamiltonian of the full 3-body problem decouples into the sum of the Hamiltonian for the restricted problem and the Hamiltonian of the linearized equations about the circular solution.

Now look at this problem on the reduced space. Let $U = (q_1, q_2, p_1, p_2)$ and $M = \varepsilon^2 + m_1 + m_2 = \varepsilon^2 + 1$ (total mass); so, $u_i = a_i - \varepsilon q_i$, and $v_i = -m_i K a_i - \varepsilon p_i$. The center of mass C, linear momentum L, and angular momentum A in these coordinates are

$$\begin{aligned} C &= \{\varepsilon^2 \xi - \varepsilon(m_1 q_1 + m_2 q_2)\}/M, \\ L &= \varepsilon^2 \eta - \varepsilon(p_1 + p_2), \\ A &= \varepsilon^2 \xi^T K \eta - (a_1 - \varepsilon q_1)^T K(m_1 K a_1 + \varepsilon p_1) - (a_2 - \varepsilon q_2)^T K(m_2 K a_2 + \varepsilon p_2). \end{aligned} \tag{4}$$

Note that when $\varepsilon = 0$, these three qualities depend only on the variables of the 2-body problem, $U = (q_1, q_2, p_1, p_2)$, and are independent of the variables of the restricted problem, ξ, η. So when $\varepsilon = 0$, the reduction is on the 2-body problem alone.

Look at the reduction of the 2-body problem in rotating coordinates with masses μ and $1 - \mu$, and let $\nu = \mu(1 - \mu)$. Fixing the center of mass at the origin and ignoring linear momentum is done by moving to Jacobi coordi-

nates which will be denoted by (α, β); see Section IV.B.16 but replace (u, v) with (α, β) to give

$$T = \frac{\|\beta\|^2}{2\nu} - \alpha^T K\beta - \frac{\nu}{\|\alpha\|}. \tag{5}$$

Put this problem in polar coordinates (see Section IV.C) to get

$$T = \frac{1}{2\nu}\left(R^2 + \frac{\Theta^2}{r^2}\right) - \Theta - \frac{\nu}{r}, \tag{6}$$

$$\dot{r} = \frac{R}{\nu}, \qquad \dot{R} = \frac{\Theta^2}{\nu r^3} - \frac{\nu}{r^2},$$
$$\dot{\theta} = \frac{\Theta}{\nu r^2} - 1, \qquad \dot{\Theta} = 0. \tag{7}$$

The reduction to the reduced space is done by holding the angular momentum, Θ, fixed and ignoring the angle θ (mod out the rotational symmetry). The distance between the primaries has been chosen as 1; so, the relative equilibrium must have $r = 1$; therefore, $\Theta = \nu$. The linearization about this critical point is

$$\dot{r} = R/\nu, \qquad \dot{R} = -\nu r. \tag{8}$$

This linear equation is a harmonic oscillator with frequency 1 and comes from the Hamiltonian $S = R^2/\nu + \nu r^2$.

In summary, the Hamiltonian of the 3-body problem on the reduced space when $\varepsilon = 0$ is

$$H_R = G + S = G + R^2/\nu + \nu r^2. \tag{9}$$

PROOF OF THEOREM 1. Let $\xi = \phi(t)$, $\eta = \psi(t)$ be a T-periodic solution of the restricted problem with multipliers 1, 1, τ, τ^{-1}. By assumption, $\tau \neq 1$ and $T \neq k2\pi$, where k is an integer. Now $\xi = \phi(t)$, $\eta = \psi(t)$, $r = 0$, $R = 0$ is a T-periodic solution of the system whose Hamiltonian is H_R, i.e., the Hamiltonian of the 3-body problem on the reduced space with $\varepsilon = 0$. The multipliers of this periodic solution are 1, 1, τ, τ^{-1}, $\exp(iT)$, and $\exp(-iT)$. Since T is not a multiple of 2π, $\exp(\pm iT) \neq 1$, and so this solution is elementary. By Proposition A.1, this solution can be continued into the full problem with $\varepsilon \neq 0$, but small. ∎

G. Some Elliptic Orbits

All of the multipliers of the elliptic solutions of the Kepler problem in either fixed or rotating coordinates are $+1$ because they are not isolated in an energy level; see Proposition V.E.6. Thus, there is no hope of using the

methods used previously; however, the restricted problem has a symmetry which when exploited properly proves that some elliptic orbits can be continued from the Kepler problem into the restricted problem. The main idea is given in the following lemma.

Lemma 1. *A solution of the restricted* 3*-body problem* (C.1) *which crosses the line of syzygy (the* x_1 *axis) orthogonally at a time* $t = 0$ *and later at a time* $t = T/2 > 0$ *is T-periodic and symmetric with respect to the line of syzygy.*

PROOF. This is an easy consequence of the exercises following Chapter V. ∎

That is, if $x = \phi(t)$, $y = \psi(t)$ is a solution of the restricted problem such that $x_2(0) = y_1(0) = x_2(T/2) = y_1(T/2) = 0$, where $T > 0$, then this solution is T-periodic and symmetric in the x_1 axis.

In Delaunay coordinates (ℓ, g, L, G) (see Section IV.E), an orthogonal crossing of the line of syzygy at a time t_0 is

$$\ell(t_0) = n\pi, \qquad g(t_0) = m\pi, \quad n, m \text{ integers.} \tag{1}$$

These equations will be solved using the implicit function theorem to yield the following theorem.

Theorem 2. *Let* m, k *be relatively prime integers and* $T = 2\pi m$. *Then the elliptic T-periodic solution of the Kepler problem in rotating coordinates* (IV.E.3) *which satisfies*

$$\ell(0) = \pi, \qquad g(0) = \pi, \qquad L^3(0) = m/k \tag{2}$$

and does not go through $x = (1, 0)$ *can be continued into the restricted problem for* μ *small. This periodic solution is symmetric with respect to the line of syzygy.*

PROOF. The Hamiltonian of the restriced 3-body problem in Delaunay elements for small μ is

$$H = -\frac{1}{2L^2} - G + O(\mu), \tag{3}$$

and the equations of motion are

$$\begin{aligned} \dot{\ell} &= 1/L^3 + O(\mu), & \dot{L} &= 0 + O(\mu), \\ \dot{g} &= -1 + O(\mu), & \dot{G} &= 0 + O(\mu). \end{aligned} \tag{4}$$

Let $L_0^3 = m/k$, and let $\ell(t, \Lambda, \mu)$, $g(t, \Lambda, \mu)$, $L(t, \Lambda, \mu)$, and $G(t, \Lambda, \mu)$ be the solution which goes through $\ell = \pi$, $g = \pi$, $L = \Lambda$, $G =$ anything at $t = 0$; so, it is a solution with an orthogonal crossing of the line of syzygy at $t = 0$.

From (4) $\ell(t, \Lambda, 0) = t/\Lambda^3 + \pi$, $g(t, \Lambda, 0) = -t + \pi$. Thus, $\ell(T/2, L_0, 0) = (1 + k)\pi$ and $g(T/2, L_0, 0) = (1 - m)\pi$, and so when $\mu = 0$, this solution has

another orthogonal crossing at time $T/2 = m\pi$. Also

$$\det \begin{pmatrix} \partial \ell/\partial t & \partial \ell/\partial \Lambda \\ \partial g/\partial t & \partial g/\partial \Lambda \end{pmatrix}_{t=T/2, L=L_0, \mu=0} = \begin{pmatrix} k/m & -3\pi(k^4/m)^{1/3} \\ -1 & 0 \end{pmatrix} \neq 0. \quad (5)$$

Thus, the theorem follows by the implicit function theorem. ∎

It is not too hard to show that for a fixed m and k, only a finite number of such elliptic orbits pass through the singularity at the other primary, $x = (1, 0)$. This rules out a finite number of collision orbits and a finite number of G's.

It is only a little more difficult to establish the existence of symmetric elliptic periodic solutions near a primary as in Section D [see Arenstorf (1968)]. It is also easy to show the existence of symmetric elliptic periodic solutions near infinity as in Section E [see Meyer (1981a)].

H. Further Reading

Poincaré was interested in the existence of periodic solutions in celestial mechanics. One of the reasons he gave for his interest was that he felt that, typically, periodic solutions were dense in Hamiltonian systems. The first volume of Poincaré (1899) contains much of his work on the subject and the foundations of his continuation method. Poincare's work was carried on by Moulton and his students. Moulton (1920) contains most of the work of this school.

Most of Poincare's and Moulton's work is concerned with continuing circular solutions. Arensdorf (1963, 1968) showed how to continue symmetric periodic solutions. Barrar (1965) uses Delaunay elements to greatly simplify much of the earlier work. Barrar's paper is a little gem.

The simple proof of Lyapunov's center theorem was taken from Markus and Meyer (1980). The proof that periodic solutions can be continued from the restricted to the full problem was taken from Meyer (1981b). This last paper has many more theorems proven by the continuation method on the existence of periodic solutions in the N-body problem.

Problems

1. Consider a periodic system of equations of the form $\dot{x} = f(t, x, v)$ where v is a parameter, and f is T-periodic in t, $f(t + T, x, v) = f(t, x, v)$. Let $\phi(t, \xi, v)$ be the general solution, $\phi(0, \xi, v) = \xi$.

a. Show that $\phi(t, \xi', v')$ is T-periodic if and only if $\phi(T, \xi', v') = \xi'$.

b. A T-periodic solution $\phi(t, \xi', v')$ *can be continued* if there is a smooth function $\bar{\xi}(v)$ such that $\bar{\xi}(v') = \xi'$, and $\phi(T, \bar{\xi}(v'), v')$ is T-periodic. The *multipliers* of the

T-periodic solution $\phi(t, \xi', \nu')$ are the eigenvalues of $\partial\phi(T, \xi', \nu')/\partial\xi$. Show that a T-periodic solution can be continued if all of its multipliers are different from $+1$.

2. Consider the classical Duffing's equation $\ddot{x} + x + \gamma x^3 = A \cos \omega t$ or $\dot{x} = y = \partial H/\partial y$, $\dot{y} = -x - \gamma x^3 + A \cos \omega t = -\partial H/\partial x$, where $H = \frac{1}{2}(y^2 + x^2) + \gamma x^4/4 - Ax \cos \omega t$. Show that if $1/\omega \neq 0, \pm 1, \pm 2, \pm 3, \ldots$, then for small forcing, A, and small nonlinearity, γ, there is a small periodic solution of the forced Duffing's equation with the same period as the external forcing, $T = 2\pi/\omega$. In the classical literature this solution is sometimes referred to as the *harmonic*. (Hint: Set the parameters γ and A to zero, then the equation is linear and solvable. Note that zero is a T-periodic solution.)

3. Show that the eigenvalues of $\begin{pmatrix} 0 & -1 \\ \mu & 0 \end{pmatrix}$ are continuous in μ but not smooth in μ.

4. Hill's lunar problem is defined by the Hamiltonian

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{1}{\|x\|} + (3x_1^2 - \|x\|^2),$$

where $x, y \in \mathbb{R}^2$. Show that it has two equilibrium points on the x_2 axis. Linearize the equations of motion about these equilibrium points, and discuss how Lyapunov's center and the stable manifold theorem apply.

5. Show that the scaling used in Section D to obtain Hill's orbits for the restricted problem works for Hill's lunar problem (Problem 4) also. Why doesn't the scaling for comets work?

6. Prove Lemma G.1, and verify that (G.1) is the condition for an orthogonal crossing of the line of syzygy in Delaunay elements.

CHAPTER VII

Perturbation Theory and Normal Forms

Perturbation theory is one of the few ways that one can bridge the gap between the behavior of a real nonlinear system and its linear approximation. Because the theory of linear systems is so much simpler, investigators are tempted to fit the problem at hand to a linear model without proper justification. Such a linear model may lead to quantitative as well as qualitative errors. On the other hand, so little is known about the general behavior of a nonlinear system that some sort of approximation has to be made.

Many interesting problems can be formulated as a system of equations which depend on a small parameter, ε, with the property that when $\varepsilon = 0$ the system is linear, or at least integrable. This chapter develops a very powerful and general method for handling the formal aspects of perturbations of linear and integrable systems, and the next three chapters contain several rigorous results which depend on these formal considerations.

A. The Method of Lie Transforms

One of the most general methods of mathematics is to simplify a problem by a change of variables. The method of Lie transforms developed by Deprit (1969) and extended by Kamel (1970) and Henrard (1970a, b, c) is a general procedure to change variables in a system of equations which depend on a small parameter. Deprit's original method was for Hamiltonian systems only, but the extensions by Kamel and Henrard handle non-Hamiltonian equations. Only the Hamiltonian case will be treated here.

1. Generating a Near Identity Symplectic Change of Variables

The general idea of this method is to generate a symplectic change of variables depending on a small parameter as the general solution of a Hamiltonian system of differential equations; see Theorem IV.A.2. $X(\varepsilon, y)$ is said to be a *near identity symplectic change of variables* (or *transformation*) if X is symplectic for each fixed ε and is of the form $X(\varepsilon, y) = y + O(\varepsilon)$, i.e., $X(0, y) = y$. Since $X(0, y) = y$, $\partial X(\varepsilon, y)/\partial y$ is nonsingular for small ε so by the inverse function theorem, the map $y \to X(\varepsilon, y)$ has a differentiable inverse for small ε. Both X and its inverse are symplectic for fixed ε.

Consider the nonautonomous Hamiltonian system

$$\frac{dx}{d\varepsilon} = J\nabla W(\varepsilon, x) \tag{1}$$

and the initial condition

$$x(0) = y, \tag{2}$$

where W is smooth. The basic theory of differential equations asserts that the general solution of this problem is a smooth function $X(\varepsilon, y)$ such that $X(0, y) \equiv y$, and by Theorem IV.A.2, the function X is symplectic for fixed ε. That is, the differential equation (1) and the initial condition (2) define a near identity symplectic change of variables.

Conversely, let $X(\varepsilon, y)$ be a near identity symplectic change of variables; so, there is an inverse function $Y(\varepsilon, x)$ such that $X(\varepsilon, Y(\varepsilon, x)) \equiv x$ and $Y(\varepsilon, X(\varepsilon, y)) \equiv y$ where defined—Y is symplectic too. Differentiating $Y(\varepsilon, X(\varepsilon, y)) \equiv y$ with repect to ε yields $(\partial Y(\varepsilon, X(\varepsilon, y))/\partial x)(\partial X(\varepsilon, y)/\partial \varepsilon) + \partial Y(\varepsilon, X(\varepsilon, y))/\partial \varepsilon \equiv 0$ or $\partial X(\varepsilon, y)/\partial \varepsilon \equiv -[Y(\varepsilon, X(\varepsilon, y))/\partial x]^{-1}\, \partial Y(\varepsilon, X(\varepsilon, y))/\partial \varepsilon$. This means that $X(\varepsilon, y)$ is the general solution of $dx/d\varepsilon = U(\varepsilon, x)$, where $U(\varepsilon, x) = -[Y(\varepsilon, x)/\partial x]^{-1}\, \partial Y(\varepsilon, x)/\partial \varepsilon$. By Theorem IV.A.2, this equation is Hamiltonian; so, there is a function $W(\varepsilon, x)$ such that $U(\varepsilon, x) = J\nabla W(\varepsilon, x)$. This proves:

Lemma 1. *$X(\varepsilon, y)$ is a near identity symplectic change of variables if and only if it is the general solution of a Hamiltonian differential equation of the form* (1) *satisfying initial condition* (2).

A Hamiltonian system of equations generates symplectic transformations directly, which is in contrast to the symplectic transformations given by the generating functions in Theorem IV.C.2, where the new and old variables are mixed in a strange way.

2. The Forward Algorithm

Let $X(\varepsilon, y)$, $Y(\varepsilon, x)$, and $W(\varepsilon, x)$ be as above; so, $X(\varepsilon, y)$ is the solution of (1) satisfying (2). Think of $x = X(\varepsilon, y)$ as a change of variables $x \to y$ which depends on a parameter. Throughout this chapter, when we change variables, we will not change the parameter ε.

Let $H(\varepsilon, x)$ be a Hamiltonian and $G(\varepsilon, y) \equiv H(\varepsilon, X(\varepsilon, y))$; so, G is the Hamiltonian H in the new coordinates. We will call G the *Lie transform* of H (generated by W). Sometimes H will be denoted by H_* and G by H^*, and sometimes G will be denoted by $\mathscr{L}(W)H$ to show that G is the Lie transform of H generated by W. Let the function $H = H_*$, $G = H^*$, and W all have series expansions in the small parameter ε. The forward algorithm of the method of Lie transforms is a recursive set of formulas that relate the terms in these various series expansions.

In particular let

$$H(\varepsilon, x) = H_*(\varepsilon, x) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_i^0(x), \tag{3}$$

$$G(\varepsilon, y) = H^*(\varepsilon, y) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_0^i(y), \tag{4}$$

$$W(\varepsilon, x) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) W_{i+1}(x). \tag{5}$$

The method of Lie transforms introduces a double indexed array $\{H_j^i\}$, i, $j = 0, 1, \ldots$ which agree with the definitions given in (3) and (4) when either i or j is zero. The other terms are intermediary terms introduced to facilitate the computation.

Theorem 2. *Using the notation given above, the functions* $\{H_j^i\}$, $i = 1, 2, \ldots$, $j = 0, 1, \ldots$ *satisfy the recursive identities*

$$H_j^i = H_{j+1}^{i-1} + \sum_{k=0}^{j} \binom{j}{k} \{H_{j-k}^{i-1}, W_{k+1}\}. \tag{6}$$

Remarks. The above formulas contain the standard binomial coefficient

$$\binom{j}{k} = \frac{j!}{k!(j-k)!}.$$

Note that since the transformation generated by W is a near identity transformation, the first term in H_* and H^* is the same, namely, H_0^0. Also note that the first term in the expansion for W starts with W_1. This convention imparts some nice properties to the formulas in (6). Each term in (6) has indices summing to $i + j$, and each term on the right-hand side has upper index $i - 1$ (i.e., is in one column to the left).

In order to construct the change of variables $X(\varepsilon, y)$, note that X is the transform of the identity function or $X(\varepsilon, y) = \mathscr{L}(W)(\mathrm{id})$, where $\mathrm{id}(x) = x$.

The interdependence of the functions $\{H_j^i\}$ can easily be understood by considering the Lie triangle

$$\begin{array}{ccccc} H_0^0 & & & & \\ \downarrow & & & & \\ H_1^0 & \to & H_0^1 & & \\ \downarrow & & \downarrow & & \\ H_2^0 & \to & H_1^1 & \to & H_0^2. \\ \downarrow & & \downarrow & & \downarrow \end{array} \tag{7}$$

The coefficients of the expansion of the old function H_* are in the left column, and those of the new function H^* are on the diagonal. Formula (6) states that to calculate any element in the Lie triangle, you need the entries in the column one step to the left and up.

For example, to compute the series expansion for H^* through terms of order ε^2, you first compute H_0^1 by the formula

$$H_0^1 = H_1^0 + \{H_0^0, W_1\}, \tag{8}$$

which gives the term of order ε, and then you compute

$$\begin{aligned} H_1^1 &= H_1^0 + \{H_1^0, W_1\} + \{H_0^0, W_2\}, \\ H_0^2 &= H_1^1 + \{H_0^1, W_1\}. \end{aligned} \tag{9}$$

Then $H^*(\varepsilon, x) = H_0^0(x) + H_0^1(x)\varepsilon + H_0^2(x)(\varepsilon^2/2) + \cdots$.

The computer age has made long formulas obsolete by favoring algorithms. Here is the simple algorithm to compute H^* through terms in $\varepsilon^{\text{maxorder}}$. Here $H_j^i = H(j, i)$.

```
input: maxorder, (H(row, 0) do row = 0 to maxorder),
                 (W(k) do k = 1 to maxorder);
output: (H(0, col) for col = 1 to maxorder);
algorithm: lie-transform;
  do row = 1 to maxorder;
    do col = 1 to row;
      H(row - col, col) = H(row - col + 1, col - 1) +
          sum_{k=0}^{row-col} binomial(row - col, k) {H(row - col - k, col - 1), W(k + 1)};
    end do;
  end do;
end algorithm;
```

PROOF OF THEOREM 2. Recall that $H^*(\varepsilon, y) = G(\varepsilon, y) = H(\varepsilon, X(\varepsilon, y))$, where $X(\varepsilon, y)$ is the general solution of (3). Define the differential operator $\mathscr{D} = \mathscr{D}_W$ by

$$\mathscr{D}F(\varepsilon, x) = \frac{\partial F}{\partial \varepsilon}(\varepsilon, x) + \{F, W\}(\varepsilon, x), \tag{10}$$

so that

$$\frac{d}{d\varepsilon}\left(F(\varepsilon, x)\Big|_{x=X(\varepsilon,y)}\right) = \mathscr{D}F(\varepsilon, x)\Big|_{x=X(\varepsilon,y)}. \tag{11}$$

Define new functions by $H^0 = H$, $H^i = \mathscr{D}H^{i-1}$, $i \geq 1$. Let these functions have series expansions

$$H^i(\varepsilon, x) = \sum_{k=0}^{\infty}\left(\frac{\varepsilon^k}{k!}\right)H_k^i(x), \tag{12}$$

so,

$$\begin{aligned} H^i(\varepsilon, x) &= \mathscr{D}\sum_{k=0}^{\infty}\left(\frac{\varepsilon^k}{k!}\right)H_k^{i-1}(x) \\ &= \sum_{k=1}^{\infty}\left(\frac{\varepsilon^{k-1}}{(k-1)!}\right)H_k^{i-1}(x) + \left\{\sum_{k=0}^{\infty}\left(\frac{\varepsilon^k}{k!}\right)H_k^{i-1}, \sum_{s=0}^{\infty}\left(\frac{\varepsilon^s}{s!}\right)W_{s+1}\right\} \\ &= \sum_{j=0}^{\infty}\left(\frac{\varepsilon^j}{j!}\right)\left(H_{j+1}^{i-1} + \sum_{k=0}^{j}\binom{j}{k}\{H_{j-k}^{i-1}, W_{k+1}\}\right). \end{aligned} \tag{13}$$

So the functions H_j^i are related by (6). It remains to show that $H_* = G$ has the expansion (4). By Taylor's theorem and (11)

$$\begin{aligned} G(\varepsilon, y) &= \sum_{n=0}^{\infty}\left(\frac{\varepsilon^n}{n!}\right)\frac{d^n}{d\varepsilon^n}G(\varepsilon, y)\Big|_{\varepsilon=0} \\ &= \sum_{n=0}^{\infty}\left(\frac{\varepsilon^n}{n!}\right)\frac{d}{d\varepsilon^n}\left(H(\varepsilon, x)\Big|_{x=X(\varepsilon,y)}\right)_{\varepsilon=0} \\ &= \sum_{n=0}^{\infty}\left(\frac{\varepsilon^n}{n!}\right)\left(\mathscr{D}^nH(\varepsilon, x)\Big|_{x=X(\varepsilon,y)}\right)_{\varepsilon=0} \\ &= \sum_{n=0}^{\infty}\left(\frac{\varepsilon^n}{i!}\right)H_0^n(x). \end{aligned}$$

■

3. The Remainder Function

Assume now that the Hamiltonian, and hence the equations, is a time-dependent Hamiltonian, i.e., consider

$$\dot{x} = J\nabla H(\varepsilon, t, x), \tag{14}$$

where H has an expansion

$$H(\varepsilon, t, x) = H_*(\varepsilon, t, x) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_i^0(t, x). \tag{15}$$

Make a symplectic change of coordinates, $x = X(\varepsilon, t, y)$, which transforms (14) to the Hamiltonian differential equation

$$\dot{y} = J\nabla G(\varepsilon, t, y) + J\nabla R(\varepsilon, t, y) = J\nabla K(\varepsilon, t, y), \tag{16}$$

where $G(\varepsilon, t, y) = H^*(\varepsilon, t, y) = H(\varepsilon, t, X(\varepsilon, t, y))$ is the Lie transform of H, R is the remainder function, and $K = G + R$ is the new Hamiltonian. Let G, R, and K have series expansions of the form

$$\begin{aligned} G(\varepsilon, t, y) &= \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_0^i(t, y), \qquad R(\varepsilon, t, y) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) R_0^i(t, y), \\ K(\varepsilon, t, y) &= \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) K_0^i(t, y). \end{aligned} \tag{17}$$

Let the symplectic change of variables $X(\varepsilon, t, y)$ be the general solution of the Hamiltonian system of equations

$$dx/d\varepsilon = J\nabla W(\varepsilon, t, x), \qquad x(0) = y, \tag{18}$$

where $W(\varepsilon, x)$ is a Hamiltonian function with a series expansion of the form

$$W(\varepsilon, t, x) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) W_{i+1}(t, x). \tag{19}$$

The variable t is simply a parameter in (19), and so the function $G = H^*$ can be computed by Formulas (6) in Theorem 2 using the Lie triangle as a guide. The remainder term, R, needs further consideration.

Theorem 3. *Using the notation of this subsection, the remainder function is given by*

$$R(\varepsilon, t, y) = -\int_0^{\varepsilon} \mathscr{L}_W \left(\frac{\partial W}{\partial t}\right)(s, t, y)\, ds. \tag{20}$$

PROOF. Making the symplectic change of variable $x = X(\varepsilon, t, y)$ in (14) directly gives

$$\dot{y} = \left(\frac{\partial X}{\partial y}(\varepsilon, t, y)\right)^{-1} J\nabla_x H(\varepsilon, t, X(\varepsilon, t, y) - \left(\frac{\partial X}{\partial y}(\varepsilon, t, y)\right)^{-1} \frac{\partial X}{\partial t}(\varepsilon, t, y). \tag{21}$$

By the discussion in Section IV.A the first term on the right-hand side of (21) is $J\nabla G$, and so

$$J\nabla R(\varepsilon, t, y) = -\left(\frac{\partial X}{\partial y}(\varepsilon, t, y)\right)^{-1} \frac{\partial X}{\partial t}(\varepsilon, t, y). \tag{22}$$

$A(\varepsilon) = \partial X(\varepsilon, t, y)/\partial y$ is the fundamental matrix solution of the variational equation; i.e., it is the matrix solution of

$$\frac{dA}{d\varepsilon} = \left(J\frac{\partial^2 W}{\partial x^2}(\varepsilon, t, X(\varepsilon, t, y))\right)A, \qquad A(0) = I. \tag{23}$$

Differentiating $\partial X(\varepsilon, t, y)/\partial\varepsilon = J\nabla W(\varepsilon, t, X(\varepsilon, t, y))$ with respect to t shows that $B(\varepsilon) = \partial X(\varepsilon, t, y)/\partial t$ satisfies

$$\frac{dB}{d\varepsilon} = \left(J\frac{\partial^2 W}{\partial x^2}(\varepsilon, t, X(\varepsilon, t, y))\right)B + J\frac{\partial^2 W}{\partial x \partial t}(\varepsilon, t, X(\varepsilon, t, y)). \tag{24}$$

Since $X(0, t, y) \equiv y$, $B(0) = 0$, and so, by the variation of constants formula,

$$B(\varepsilon) = \int_0^\varepsilon A(\varepsilon)A(s)^{-1}J\frac{\partial^2 W}{\partial x \partial t}(\varepsilon, t, X(\varepsilon, t, y))\, ds; \tag{25}$$

therefore,

$$\begin{aligned} J\nabla R(\varepsilon, t, y) &= -\left(\frac{\partial X}{\partial y}(\varepsilon, t, y)\right)^{-1}\frac{\partial X}{\partial t}(\varepsilon, t, y) = -A(\varepsilon)^{-1}B(\varepsilon) \\ &= -\int_0^\varepsilon A(s)^{-1}J\frac{\partial^2 W}{\partial x \partial t}(s, t, X(s, t, y))\, ds \\ &= -\int_0^\varepsilon JA(s)^T\frac{\partial^2 W}{\partial x \partial t}(s, t, X(s, t, y))\, ds \\ &= -J\frac{\partial}{\partial y}\int_0^\varepsilon \frac{\partial W}{\partial t}(s, t, X(s, t, y))\, ds \\ &= -J\frac{\partial}{\partial y}\int_0^\varepsilon \mathscr{L}_W\left(\frac{\partial W}{\partial t}\right)(s, t, y)\, ds. \end{aligned} \tag{26}$$

In the above, the fact that A is symplectic is used to make the substitution $A^{-1}J = JA^T$. ∎

Thus, to compute the remainder function, first compute the transform of $-\partial W/\partial t$, and then integrate it. That is, let $S_*(\varepsilon, t, x) = \sum(\varepsilon^i/i!)S_i^0(t, x)$, where $S_i^0(t, x) = -\partial W_{i-1}(t, x)/\partial t$. Compute the Lie transform of S_* by the previous algorithms to get $\mathscr{L}_W(S) = S^*(\varepsilon, t, x) = \sum(\varepsilon^i/i!)S_0^i(t, x)$. Then $R_0^i = S_0^{i-1}$.

For example, to compute the series expansion for $K = G + R$, the new Hamiltonian, through terms of order ε^2, set $K_0^0 = H_0^0$, then compute K_0^1 by the formulas

$$H_0^1 = H_1^0 + \{H_0^0, W_1\}, \qquad R_0^1 = -\frac{\partial W_1}{\partial t}, \qquad K_0^1 = H_0^1 + R_0^1, \tag{27}$$

which gives the term of order ε, and then compute

$$\begin{aligned} H_1^1 &= H_2^0 + \{H_1^0, W_1\} + \{H_0^0, W_2\}, & H_0^2 &= H_1^1 + \{H_0^1, W_1\}, \\ R_0^2 &= -\frac{\partial W_2}{\partial t} - \left\{\frac{\partial W_1}{\partial t}, W_1\right\}, & K_0^2 &= H_0^2 + R_0^2. \end{aligned} \tag{28}$$

Then $K^*(\varepsilon, x) = K_0^0(x) + K_0^1(x)\varepsilon + K_0^2(x)(\varepsilon^2/2) + \cdots$.

B. The Perturbation Algorithm

In many of the cases of interest, the Hamiltonian is given, and the change of variables is sought to simplify it. When the Hamiltonian, and hence the equations, is in sufficiently simple form, it is said to be in "normal form," an expression whose meaning will be discussed in detail later.

1. An Example: Duffing's Equation

In Section IV.C.2 the Hamiltonian of Duffing's equation was given as

$$H = \frac{1}{2}(q^2 + p^2) + \frac{\gamma}{4}q^4 \tag{1}$$

in rectangular coordinates, (q, p); and in action-angle variables, (I, ϕ), it was given as

$$H = I + \frac{\gamma}{8}I^2(3 + 4\cos 2\phi + \cos 4\phi). \tag{2}$$

This Hamiltonian is analytic in rectangular coordinates, and so has the d'Alembert character. Consider γ as a small parameter by setting $\varepsilon = \gamma/8$; so, $H(\varepsilon, I, \phi) = H^*(\varepsilon, I, \phi) = H_0^0(I, \phi) + \varepsilon H_1^0(I, \phi)$, where

$$H_0^0 = I, \qquad H_1^0 = I^2(3 + 4\cos 2\phi + \cos 4\phi). \tag{3}$$

By Formula (A.8),

$$H_0^1 = H_1^0 + \{H_0^0, W_1\}; \tag{4}$$

so,

$$H_0^1 = I^2(3 + 4\cos 2\phi + \cos 4\phi) - \frac{\partial W_1}{\partial \phi}. \tag{5}$$

Choose $W = W_1$ so that H_0^1 contains as few terms as possible (one definition of "normal form"). For the transformation generated by W to be analytic in rectangular coordinates, W must be a Poisson series with the d'Alembert character. Thus, the simplest form for H_1^0 is

$$H_0^1 = 3I^2, \tag{6}$$

which is done by taking W to be

$$W_1 = I^2(2\sin 2\phi + (1/4)\sin(4\phi). \tag{7}$$

With this W, the Hamiltonian in the new coordinates, (J, θ) would be

$$H_*(\varepsilon, J, \theta) = J + \frac{3\gamma}{8}J^2 + O(\gamma^2). \tag{8}$$

and the equations of motion would be

$$\dot{J} = O(\gamma^2), \qquad \dot{\theta} = -1 - \frac{3\gamma}{4}J + O(\gamma^2). \tag{9}$$

In these coordinates, up to terms $O(\gamma^2)$, the solutions move on circles $J =$ constant with uniform angular frequency $-1 - (3\gamma/4)J$.

Let us do this simple example again, but this time in complex coordinates $z = q + ip$, $\bar{z} = q - ip$. This change of variables is symplectic with multiplier $2i$; so, the Hamiltonian becomes

$$H(z, \bar{z}) = iz\bar{z} + \frac{\gamma i}{32}(z^4 + 4z^3\bar{z} + 6z^2\bar{z}^2 + 4z\bar{z}^3 + \bar{z}^4). \tag{10}$$

H is real in the rectangular coordinates (q, p), because H is conjugated by interchanging z and $\bar{z}$, i.e., $H(z, \bar{z}) = \overline{H(\bar{z}, z)}$. This is the reality condition for these variables. Let $\varepsilon = \gamma/32$ and

$$H_0^0 = iz\bar{z}, \qquad H_1^0 = i(z^4 + 4z^3\bar{z} + 6z^2\bar{z} + 4z\bar{z}^3 + \bar{z}^4); \tag{11}$$

so Equation (4) with $W = W_1$ becomes

$$H_0^1 = i(z^4 + 4z^3\bar{z} + 6z^2\bar{z}^2 + 4z\bar{z}^3 + \bar{z}^4) + \frac{1}{2}\left(z\frac{\partial W}{\partial z} - \bar{z}\frac{\partial W}{\partial \bar{z}}\right). \tag{12}$$

Try $W = az^\alpha\bar{z}^\beta$, then $(z\,\partial W/\partial z + \bar{z}\,\partial W/\partial\bar{z})/2 = (\alpha - \beta)az^\alpha\bar{z}^\beta/2$; so, all the terms in H_1^0 can be eliminated except those with $\alpha = \beta$. That is, if we take

$$W = -i(z^4/2 + 4z^3\bar{z} - 4z\bar{z}^3 + \bar{z}^4/2), \tag{13}$$

then

$$H_* = H_0^0 + H_1^0 = iz\bar{z} + (3\gamma i/16)(z\bar{z})^2 + O(\varepsilon^2). \tag{14}$$

Notice that both W and H_* satisfy the reality condition and so are real functions in the original coordinates q, p. The two methods of solving the problem (action-angle variables and complex variables) give the same results when written in rectangular coordinates.

2. The General Algorithm

The main Lie transform algorithm starts with a given Hamiltonian which depends on a small parameter, ε, and constructs a change of variables so that the Hamiltonian in the new variables is simple. The algorithm is built around the following observation.

Consider first a time-independent Hamiltonian, $H_*(\varepsilon, x)$, with series expansion as given in Equation (A.3); so, all the H_i^0 are known. Assume that all the entries in the Lie triangle are known down to the Nth row; so, the H_j^i are known for $i + j \leq N$, and assume that the W_i are knwon for $i \leq N$. Let L_j^i, $i + j \leq N$, be computed from the same initial Hamiltonian, but with $U_1, \ldots, U_N$, where $U_i = W_i$ for $i = 1, 2, \ldots, N - 1$ and $U_N = 0$. Then

$$\begin{aligned} H_j^i &= L_j^i \quad \text{for} \quad i+j < N, \\ H_j^i &= L_j^i + \{H_0^0, W_N\} \quad \text{for } i+j = N. \end{aligned} \tag{15}$$

This is easily seen from the recursive formulas in Theorem A.2. Recall the remark that the sum of all the indices must add to the row number; so, W_N does not affect the terms in the first $N-1$ rows. The second equation in (15) follows from a simple induction across the Nth row.

From this observation, the algorithm is as follows. Assume all the rows in the Lie triangle have been computed down to the $(N-1)$st row, that $W_1, \ldots, W_{N-1}$ have been determined, and that the diagonal terms $H_0^1, \ldots, H_0^{N-1}$ are in "normal form," i.e., simple in some sense. Now it is time to compute W_N so that H_0^N is in normal form. *Step 1*: Compute the Nth row from the formulas in Theorem A.2 assuming that $W_N = 0$, and call these terms L_j^i, $i+j = N$. *Step 2*: Solve the equation $H_0^N = L_0^N + \{H_0^0, W_N\}$ for W_N and H_0^N, so that H_0^N is in "normal form" or simple. *Step 3*: Add $\{H_0^0, W_N\}$ to each term in the Nth row, i.e., calculate $H_j^i = L_j^i + \{H_0^0, W_N\}$ for all $i+j = N$. Repeat.

The general algorithm to put a Hamiltonian in normal form up to $\varepsilon^{\text{maxorder}}$ is given below. Here $H_j^i = H(j, i)$.

input: maxorder, (H(row, 0) for row = 0 to maxorder);
output: (H(0, col), W(col) for col = 1 to maxorder);
algorithm: normalization;
  do row = 1 to maxorder;
    W(row) = 0;
    do col = 1 to row;
      H(row − col, col) = H(row − col + 1, col − 1) +

$$\sum_{k=0}^{\text{row}-\text{col}} \binom{\text{row} - \text{col}}{k} \{H(\text{row} - \text{col} - k, \text{col} - 1), W(k+1)\};$$

      end do;
    solve Htemp = H(0, row) + {H(0, 0), W temp};
    W(row) ← W temp;
    do col = 1 to row;
      H(row − col, col) = H(row − col, col) + {H(0, 0), W temp};
      end do;
    end do;
  end algorithm;

Of course the definition of normal form and simple depend on the problem at hand. It depends on the equation $H_0^N = L_0^N + \{H_0^0, W_N\}$, which in turn depends on H_0^0. This equation is called the Lie equation.

3. The General Perturbation Theorem

The algorithm can be used to prove a general theorem which includes almost all applications. Use the notation of Section A.

Theorem 1. *Let* $\{\mathscr{P}_i\}_{i=0}^{\infty}$, $\{\mathscr{Q}_i\}_{i=1}^{\infty}$, *and* $\{\mathscr{R}_i\}_{i=1}^{\infty}$ *be sequences of linear spaces of smooth functions defined on a common domain* O *in* $\mathbb{R}^{2n}$ *with the following properties*:

(i) $\mathscr{Q}_i \subset \mathscr{P}_i$, $i = 1, 2, \ldots$
(ii) $H_i^0 \in \mathscr{P}_i$, $i = 0, 1, 2, \ldots$
(iii) $\{\mathscr{P}_i, \mathscr{R}_j\} \subset \mathscr{P}_{i+j}$, $i, j = 0, 1, 2, \ldots$
(iv) *for any* $D \in \mathscr{P}_i$, $i = 1, 2, \ldots$, *there exists* $B \in \mathscr{Q}_i$ *and* $C \in \mathscr{R}_i$ *such that*

$$B = D + \{H_0^0, C\}. \tag{16}$$

Then there exists a W *with a formal Hamiltonian of the form* (A.5) *with* $W_i \in \mathscr{R}_i$, $i = 1, 2, \ldots$, *which generates a near identity symplectic change of variables* $x \to y$ *such that the Hamiltonian in the new variables has a series expansion given by* (A.4) *with* $H_0^i \in \mathscr{Q}_i$, $i = 1, 2, \ldots$.

Remarks. The Lie equation (16) is the heart of a perturbation problem. H_0^0 defines the unperturbed system when $\varepsilon = 0$, so it is supposed to be well understood. For example, it might be the harmonic oscillator or the 2-body problem. $\mathscr{F} = \{H_0^0, \cdot\}$ is a linear operator on the functions. One must analyze this operator to determine in what linear spaces Equation (16) is solvable. Roughly speaking, the Hamiltonian (A.3) starts with terms in the $\mathscr{P}$-spaces ($H_i^0 \in \mathscr{P}_i$), and the equation in normal form has terms in the $\mathscr{Q}$-space ($H_0^i \in \mathscr{Q}_i$). The $\mathscr{Q}$-spaces are smaller than the $\mathscr{P}$-spaces ($\mathscr{Q}_i \subset \mathscr{P}_i$). So the normal form is "simpler." The transformation is generated by a Hamiltonian differential equation with Hamiltonian W in the $\mathscr{R}$-spaces ($W_i \in \mathscr{R}_i$). D is an old term, B is a new term, and C is a generator.

Proof. Use induction on the rows of the Lie triangle. Induction Hypothesis I_n: Let $H_j^i \in \mathscr{P}_{i+j}$ for $0 \le i + j \le n$ and $W_i \in \mathscr{R}_i$, $H_0^i \in \mathscr{Q}_i$ for $1 \le i \le n$.

I_0 is true by assumption, and so assume I_{n-1}. By Equation (A.6)

$$H_{n-1}^1 = H_n^0 + \sum_{k=0}^{n-2} \binom{n-1}{k} \{H_{n-1-k}^0, W_{k+1}\} + \{H_0^0, W_n\}. \tag{17}$$

The last term is singled out because it is the only term that contains an element, W_n, which is not covered by the induction hypothesis or the hypothesis of the theorem. All the other terms are in $\mathscr{P}_n$ by I_{n-1} and (iii). Thus,

$$H_{n-1}^1 = L^1 + \{H_0^0, W_n\}, \tag{18}$$

where $L^1 \in \mathscr{P}_n$ is known. A simple induction on the columns of the Lie triangle using Equation (A.6) shows that

$$H_{n-s}^s = L^s + \{H_0^0, W_n\}, \tag{19}$$

where $L^s \in \mathscr{P}_n$ for $s = 1, 2, \ldots, n$, and so

$$H_n^0 = L^n + \{H_0^0, W_n\}. \tag{20}$$

By (iv), solve (20) for $W_n \in \mathscr{R}_n$ and $H_0^n \in \mathscr{Q}_i$. Thus, I_n is true. ∎

The theorem given above is formal in the sense that the convergence of the various series is not discussed. In interesting cases the series diverge, but useful information can be obtained in the first few terms of the normal form. One can stop the process at any order N to obtain a W which is a polynomial in ε and so converges. From the proof given above, it is clear that the terms in the series for H^* up to order N are unaffected by the termination. Thus, the more useful form of theorem 1 is:

Corollary 2. *Let $N \geq 1$ be given, and let $\{\mathscr{P}_i\}_{i=0}^N$, $\{\mathscr{Q}_i\}_{i=1}^N$, and $\{\mathscr{R}_i\}_{i=1}^N$ be sequences of linear spaces of smooth functions defined on a common domain O in $\mathbb{R}^{2n}$ with the following properties*:

(i) $\mathscr{Q}_i \subset \mathscr{P}_i$, $i = 1, 2, \ldots, N$,
(ii) $H_i^0 \in \mathscr{P}_i$, $i = 0, 1, 2, \ldots, N$,
(iii) $\{\mathscr{P}_i, \mathscr{R}_j\} \subset \mathscr{P}_{i+j}$, $i + j = 0, 1, 2, \ldots, N$,
(iv) *for any $D \in \mathscr{P}_i$, $i = 1, 2, \ldots, N$, there exists $B \in \mathscr{Q}_i$ and $C \in \mathscr{R}_i$ such that*

$$B = D + \{H_0^0, C\}. \tag{21}$$

Then there exists a polynomial W,

$$W(\varepsilon, x) = \sum_{i=0}^{N-1} \left(\frac{\varepsilon^i}{i!}\right) W_{i+1}(x), \tag{22}$$

with $W_i \in \mathscr{R}_i$, $i = 1, 2, \ldots, N$, *such that the change of variables $x = X(\varepsilon, y)$, where $X(\varepsilon, y)$ is the general solution of $dx/d\varepsilon = J\nabla W(\varepsilon, x)$, $x(0) = y$, transforms the convergent Hamiltonian*

$$H(\varepsilon, x) = H_*(\varepsilon, x) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_i^0(x) \tag{23}$$

to the convergent Hamiltonian

$$H(\varepsilon, x) = H^*(\varepsilon, x) = \sum_{i=0}^{N} \left(\frac{\varepsilon^i}{i!}\right) H_0^i(x) + O(\varepsilon^{N+1}), \tag{24}$$

with $H_0^i \in \mathscr{Q}_i$, $i = 1, 2, \ldots, N$.

4. The Nonautonomous Case

In the nonautonomous case, the algorithm is slightly different. The remainder function, $\mathscr{R}(\varepsilon, t, y)$, is the indefinite integral of $S^*(\varepsilon, t, y)$, where $S^*(\varepsilon, t, y) = -\mathscr{L}_W(\partial W/\partial t)\,(s, t, y)$, the Lie transform of $-\partial W/\partial t$. One constructs two Lie triangles, one for the Hamiltonian and one for the function S_*. Since R is the indefinite integral of S^*, if you want the new Hamiltonian up to terms of order ε^N, then you need all the Lie triangle for H_* down to the Nth row, but only down to the $(N-1)$st for S_*. One simply works down the two triangles together, but with the S triangle one row behind.

Assume that all the entries in the Lie triangle for H are known down to the Nth row (H_j^i, $i+j \le N$) and that all the entries in the Lie triangle for S are known down to the $(N-1)$st row (S_j^i, $i+j \le N-1$) using the W_i for $i \le N$. Let L_j^i, $i+j \le N$, be computed from the same Hamiltonian; so, $L_i^0 = H_i^0$ for all i, but with $U_1, \ldots, U_N$, where $U_i = W_i$ for $i = 1, 2, \ldots, N-1$ and $U_N = 0$. Let $\mathscr{Q}_j^i$ be the terms in the Lie triangle for the remainder using the U_i's. Then

$$\begin{aligned} H_j^i &= L_j^i \quad \text{for } i+j < N, & S_j^i &= Q_j^i \quad \text{for } i+j < N-1, \\ H_j^i &= L_j^i + \{H_0^0, W_N\} \quad \text{for } i+j = N, & S_j^i &= Q_j^i - \frac{\partial W_n}{\partial t} \quad \text{for } i+j = N-1. \end{aligned} \tag{25}$$

This is easily seen from the recursive formulas in Theorem A.2.

From this observation, the algorithm is as follows. Assume that all the rows in the Lie triangle for H have been computed down to the $(N-1)$st row, that all the rows in the Lie triangle for S have been computed down to the $(N-2)$nd row and that $W_1, \ldots, W_{N-1}$ have been determined, and that the K_0^1, $\ldots$, K_0^{N-1} are in "normal form," i.e., simple in some sense. Now it is time to compute W_N so that H_0^N is in normal form. *Step 1*: Compute the Nth row for H and the $(N-1)$st row for the remainder assuming that $W_N = 0$, and call these terms L_j^i, $i+j = N$, and $\mathscr{Q}_j$, $i+j = N-1$, respectively. *Step 2*: Solve the equation $H_0^N = L_0^N + \mathscr{Q}_0^{N-1} + \{H_0^0, W_N\} - \partial W_N/\partial t$ for W_N and H_0^N so that H_0^N is in "normal form" or simple. *Step 3*: Add $\{H_0^0, W_N\}$ to each term in the Nth row for H, and add $\partial W_N/\partial t$ to each term in the $(N-1)$st row for $\mathscr{Q}$. Repeat.

The nonautonomous version of Theorem 1 is:

Theorem 3. *Let $\{\mathscr{P}_i\}_{i=0}^{\infty}$, $\{\mathscr{Q}_i\}_{i=1}^{\infty}$, and $\{\mathscr{R}_i\}_{i=1}^{\infty}$ be sequences of linear spaces of smooth functions defined on a common domain O in $\mathbb{R}^1 \times \mathbb{R}^{2n}$. Let $\dot{\mathscr{R}}_i$ be the space of all derivatives of functions in $\mathscr{R}_i$. Assume the following:*

(i) $\mathscr{Q}_i \subset \mathscr{P}_i$, $i = 1, 2, \ldots$,
(ii) $H_i^0 \in \mathscr{P}_i$, $i = 0, 1, 2, \ldots$,
(iii) $\{\mathscr{P}_i, \mathscr{R}_j\} \subset \mathscr{P}_{i+j}$ *and* $\{\mathscr{P}_i, \dot{\mathscr{R}}_j\} \subset \mathscr{P}_{i+j}$, *for* $i, j = 0, 1, 2, \ldots$,
(iv) *for any* $D \in \mathscr{P}_i$, $i = 1, 2, \ldots$, *there exists* $B \in \mathscr{Q}_i$ *and* $C \in \mathscr{R}_i$ *such that*

$$B = D + \{H_0^0, C\} - \frac{\partial C}{\partial t}. \tag{26}$$

Then there exists a W with a formal Hamiltonian of the form (A.5) with $W_i \in \mathscr{R}_i$, $i = 1, 2, \ldots$, *which generates a near identity symplectic change of variables $x \to y$ such that the Hamiltonian in the new variables has a series expansion given by Equation* (A.4) *with $H_0^i \in \mathscr{Q}_i$, $i = 1, 2, \ldots$.*

5. Duffing's Equation Revisited

Consider the Hamiltonian (2) of Duffing's equation as written in action-angle variables. The operator $\{H_0^0, C\} = \partial C/\partial\phi$ is very simple to understand. Equation (26) becomes

$$B = D + \frac{\partial C}{\partial \phi}. \tag{27}$$

If D is a finite Poisson series with d'Alembert character, then by taking B to be the term of D which is independent of the angle ϕ and $C = \int (B - D)\, d\phi$, B and C satisfy Equation (27). This leads us to the following definitions of the spaces.

Let $\mathscr{P}_i$ be the space of all finite Poisson series with d'Alembert character corresponding to homogeneous polynomials of degree $2i + 2$ in rectangular coordinates. So an element in $\mathscr{P}_i$ is of the form I^{i+1} times a finite Fourier series in ϕ. Let $\mathscr{Q}_i$ be the space of all polynomials of the form AI^{i+1}, where A is a constant. Let $\mathscr{R}_i$ be the subspace of $\mathscr{P}_i$ of Poisson series without a term independent of ϕ. So $\mathscr{P}_i = \mathscr{Q}_i \oplus \mathscr{R}_i$. Since the Poisson bracket of homogeneous polynomials of degree $2i + 2$ and degree $2j + 2$ is a polynomial of degree $2(i + j) + 2$, and since symplectic changes of coordinates preserve Poisson brackets, we have $\{\mathscr{P}_i, \mathscr{R}_j\} \subset \mathscr{P}_{i+j}$. Thus, by Corollary 3, there exists a formal, symplectic transformation which transforms the Hamiltonian of Duffing's equation into the form

$$H^*(\varepsilon, J) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_0^i(J) \tag{28}$$

and the equations of motion become

$$\dot{J} = 0, \qquad \dot{\phi} = -\frac{\partial H^*}{\partial \phi}(\varepsilon, J) = -\omega(\varepsilon, J). \tag{29}$$

Thus, formally, the solutions move on circles with a uniform frequency $\omega(\varepsilon, J)$, which depends on ε and J. By the theorems of Poincaré (1885) and Rüssman (1959) the series converges in this simple case.

5. Uniqueness of Normal Forms

One of the important special cases where Theorem 1 applies is when the operator $L = \{H_0^0, \cdot\}: \mathscr{P}_i \to \mathscr{P}_i$ is *simple*, i.e., when $\mathscr{P}_i = Q_i \oplus \mathscr{R}_i$, $Q_i = \text{kernel}\,(L_i)$, and $\mathscr{R}_i = \text{range}(L_i)$. In this case, the Lie equation (16) has a unique solution. This is not enough to assure uniqueness of the normal form. One needs one extra condition.

Theorem 4. *Let $\{\mathscr{P}_i\}_{i=0}^{\infty}$ be squences of linear spaces of smooth functions defined on a common domain O in $\mathbb{R}^{2n}$. Let $L = \{H_0^0, \cdot\}: \mathscr{P}_i \to \mathscr{P}_i$ be simple; so, $\mathscr{P}_i = \mathscr{Q}_i \oplus \mathscr{R}_i$, $\mathscr{Q}_i = \text{kernel}(L_i)$, and $\mathscr{R}_i = \text{range}(L_i)$. Assume*

(i) $H_i^0 \in \mathscr{P}_i$, $i = 0, 1, 2, \ldots$,
(ii) $\{\mathscr{P}_i, \mathscr{P}_j\} \subset \mathscr{P}_{i+j}$, $i, j = 0, 1, 2, \ldots$.

Then there exists a W with a formal expansion of the form (A.5) with $W_i \in \mathscr{R}_i$, $i = 1, 2, \ldots$, *such that W generates a near identity symplectic change of vari-*

ables $x \to y$ which transforms the Hamiltonian $H_*(\varepsilon, x)$ with the formal series expansion given in Equation (A.3) *to the Hamiltonian $H^*(\varepsilon, y)$ with the formal series expansion given by Equation* (A.4) *with $H_0^i \in \mathcal{Q}_i$, $i = 1, 2, \ldots$. Moreover, if*

(iii) $\{\mathcal{Q}_i, \mathcal{Q}_j\} = 0, i, j = 1, 2, \ldots,$

then the terms in the normal form are unique.

Remark. All the obvious remarks about the time-dependent cases hold here also. The normal form is unique, but the transformation taking the equation need not be unique. Clearly, this theorem applies to the Duffing example. We shall not need this theorem in our development. See Liu (1985) for a proof or see Problems Section.

C. Normal Form at an Equilibrium

Consider an analytic Hamiltonian, H, which has an equilibrium point at the origin in $\mathbb{R}^{2n}$, and assume that the Hamiltonian is zero at the origin. Then H has a Taylor series expansion of the form

$$H(x) = H_{\#}(x) = \sum_{i=0}^{\infty} H_i(x), \tag{1}$$

where H_i is a homogeneous polynomial in x of degree $i + 2$; so, $H_0(x) = \frac{1}{2}x^T Sx$, where S is a $2n \times 2n$ real symmetric matrix, and $A = JS$ is a Hamiltonian matrix. The linearized equations about the critical point $x = 0$ are

$$\dot{x} = Ax = JSx = J\nabla H_0(x), \tag{2}$$

and the general solution of (2) is $\phi(t, \xi) = \exp(At)\xi$. A traditional analysis is to solve (2) by linear algebra techniques and then hope that the solutions of the nonlinear problem are not too dissimilar from the solutions of the linear equation. In many cases this hope is unjustifiable. The next best thing is to put the equations in normal form and to study the solutions of the normal form equations. This too has its pitfalls.

1. The Classical Case

The matrix A is *simple* if it has $2n$ linearly independent eigenvectors which may be real or complex. The matrix A being simple is equivalent to A being similar to a diagonal matrix by a real or complex similarity transformation. This is why A is sometimes said to be *diagonalizable*. The classical theorem on normal forms is:

Theorem 1. *Let A be simple. Then there exists a formal, symplectic change of variables,*

$$x = X(y) = y + \cdots, \tag{3}$$

which transforms the Hamiltonian (1) to

$$H^{\#}(y) = \sum_{i=0}^{\infty} H^i(y), \tag{4}$$

where H^i is a homogeneous polynomial of degree $i + 2$ such that

$$H^i(e^{At}y) \equiv H^i(y), \tag{5}$$

for all $i = 0, 1, \ldots$, all $y \in \mathbb{R}^{2n}$, and all $t \in \mathbb{R}$.

Remark. Formula (5) is the classical definition of *normal form* for a Hamiltonian near an equilibrium point with a simple linear part. Formula (5) says that H^i is an integral for the linear system (2); so, by Theorem I.A.1, (5) is equivalent to

$$\{H^i, H^0\} = 0 \tag{6}$$

for all i.

PROOF. In order to study the solutions near the origin, scale the variables by $x \to \varepsilon x$. This is a symplectic transformation with multiplier ε^2; so, the Hamiltonian becomes

$$H(\varepsilon, x) = H_*(\varepsilon, x) = \sum_{i=0}^{\infty} \left(\frac{\varepsilon^i}{i!}\right) H_i^0(x), \tag{7}$$

where $H_i^0 = i!H_i$. Since we are working formally, we can set $\varepsilon = 1$ at the end, or we can rescale by $x \to \varepsilon^{-1}x$.

Let P_i be the linear space of all real homogeneous polynomials of degree $i + 2$; so, $H_1^0 \in P_i$. Since A is simple, A has $2n$ linearly independent eigenvectors $s_1, \ldots, s_{2n}$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_{2n}$. The s_i are row eigenvectors; so, $s_i A = \lambda_i s_i$. Let $2r$ of the eigenvalues be complex, and number them so that $\lambda_i = \bar{\lambda}_{n+i}$ for $i = 1, \ldots, r$. Choose the eigenvectors so that $s_i = \bar{s}_{n+i}$ for $i = 1, \ldots, r$. The other eigenvalues and eigenvectors are real. Let $K \in P_i$; so, K is a homogeneous polynomial of degree $i + 2$. Since the s_i's are independent, K may be written in the form

$$K = \sum \varkappa_{m_1 m_2 \ldots m_{2n}} (s_1 x)^{m_1} (s_2 x)^{m_2} \ldots (s_{2n} x)^{m_{2n}}, \tag{8}$$

where the sum is over all $m_1 + \cdots + m_{2n} = i + 2$. So the monomials in

$$B = \{(s_1 x)^{m_1} (s_2 x)^{m_2} \ldots (s_{2n} x)^{m_{2n}} : m_1 + \cdots + m_{2n} = i + 2\} \tag{9}$$

span P_i. It is also clear that they are independent; so, form a basis for P_i. The coefficients in (8) may be complex but must satisfy the reality condition that

interchanging the subscripts m_i and m_{n+i} for $i = 1, \ldots, r$ in the $\varkappa$'s is the same as conjugation.

Now let $L = L_i$: $P_i \to P_i$ be the linear operator of Theorem B.4 as it applies to Hamiltonian systems, that is, define L by $LG = \{H_0^0, G\} = -(\partial G/\partial x)Ax$; so,

$$\begin{aligned} &L((s_1x)^{m_1}(s_2x)^{m_2}\ldots(s_{2n}x)^{m_{2n}}) \\ &\quad = -(m_1\lambda_1 + \cdots + m_{2n}\lambda_{2n})^{m_1}(s_1x)^{m_1}(s_2x)^{m_2}\ldots(s_{2n}x)^{m_{2n}}. \end{aligned} \tag{10}$$

So the elements of B are eigenvectors of L and the eigenvalues are $(m_1\lambda_1 + \cdots + m_{2n}\lambda_{2n})$, $m_1 + \cdots + m_{2n} = i + 2$. Thus, we can define L-invariant subspaces

$$\begin{aligned} \mathscr{K}_i = \operatorname{span}((s_1x)^{m_1}(s_2x)^{m_2}\ldots(s_{2n}x)^{m_{2n}}&: m_1 + \cdots + m_{2n} = i + 2, \\ &m_1\lambda_1 + \cdots + m_{2n}\lambda_{2n} = 0), \end{aligned} \tag{11}$$

$$\begin{aligned} \mathscr{R}_i = \operatorname{span}((s_1x)^{m_1}(s_2x)^{m_2}\ldots(s_{2n}x)^{m_{2n}}&: m_1 + \cdots + m_{2n} = i + 2, \\ &m_1\lambda_1 + \cdots + m_{2n}\lambda_{2n} \neq 0). \end{aligned} \tag{12}$$

In summary, $\mathscr{K}_i = \operatorname{kernel}(L)$, $\mathscr{R}_i = \operatorname{range}(L)$, and $P_i = \mathscr{K}_i \oplus \mathscr{R}_i$. Thus, this classical theorem follows from the first part of Theorem B.4 because we have shown that the operators L_i: $P_i - P_i$ are simple. However, the extra condition (iii) in Theorem B.4 is not satisfied in general; so, the normal form may not be unique. ∎

Birkhoff (1927) considered a special case of the above.

Corollary 2. *Assume that the quadratic part of* (1) *is of the form*

$$H_0(x) = \sum_{i=1}^{n} \lambda_i x_i x_{n+i}, \tag{13}$$

where the λ_i's *are independent over the integers, i.e., there is no nontrivial relation of the form*

$$\sum_{i=1}^{n} k_i\lambda_i = 0, \tag{14}$$

where the k_i's *are integers. Then there exists a formal, symplectic change of variables* $x = X(y) = y + \cdots$ *which transforms the Hamiltonian* (1) *to the Hamiltonian* (4), *where* $H^i(y)$ *is a homogeneous polynomial of degree* $i + 1$ *in the n products* $y_1y_{n+1}, \ldots, y_ny_{2n}$. *So,* $H^\#(y_1, \ldots, y_{2n}) = H^\#(y_1y_{n+1}, \ldots, y_ny_{2n})$ *where* $H^\#$ *is a function of n variables. Moreover, in this case, the normal form is unique.*

Remark. Formally the equations of motion for the system in normal form are

$$\begin{aligned} \dot{y}_i &= y_{n+i}D_iH^\#(y_1y_{n+1}, \ldots, y_ny_{2n}), \\ \dot{y}_{i+n} &= -y_nD_iH^\#(y_1y_{n+1}, \ldots, y_ny_{2n}). \end{aligned} \tag{15}$$

Here D_i stands for the partial derivative with respect to the ith variable. In this form, the system of equations have n formal integrals in involution, $I_1 = y_1 y_{n+1}, \ldots, I_n = y_n y_{2n}$.

In the case when the $\lambda_j = i\omega_j$ are pure imaginary and the y_i are the complex coordinates discussed in Lemma II.D.6, then we can switch to action-angle variables by $y_i = \sqrt{I_i/2}e^{i\phi_i}$, $y_{n+i} = \sqrt{I_i/2}e^{-i\phi_i}$. The Hamiltonian in normal form is a function of the action variables only; so, the Hamiltonian is $H^\dagger(I_1, \ldots, I_n)$, and the equations of motion are

$$\dot{I} = \frac{\partial H^\dagger}{\partial \phi} = 0, \qquad \dot{\phi}_i = -\frac{\partial H^\dagger}{\partial I_i} = \omega_i(I_1, \ldots, I_n). \tag{16}$$

Here $\omega_i(I_1, \ldots, I_n) = \pm\omega_i + \cdots$, and the sign is determined by the cases in Lemma II.D.6. Setting the action variables equal to nonzero constants, $I_1 = c_1, \ldots, I_n = c_n$, defines an invariant set which is an n-torus with n angular coordinates $\phi_1, \ldots, \phi_n$. On each torus the angular frequencies $\omega_i(I_1, \ldots, I_n)$, are constant, and so, define a linear flow on the torus as discussed in I.B.5. The frequencies vary from torus to torus in general.

Notation. For this proof and subsequent discussions, some notation is useful. Let $\mathscr{Z} = \mathbb{Z}_+^{2n}$ denote the set of all $2n$-tuples of non-negative integers; so, $k \in \mathscr{Z}$ means $k = (k_1, \ldots, k_{2n})$, $k_i \geq 0$, k_i an integer. Let $|k| = k_1 + \cdots + k_{2n}$. If $x \in \mathbb{R}^{2n}$ and $k \in \mathscr{Z}$, then define $x^k = x_1^{k_1} x_2^{k_2} \ldots x_{2n}^{k_{2n}}$.

PROOF. The linear part is clearly simple. Let $H^i(y) = \sum h_k y^k$, where the sum is over $k \in \mathscr{Z}$, $|k| = i + 2$. The general solution of the linear system is $y_i = y_{i,0} \exp(\lambda_i t)$, $y_{i+n} = y_{i+n,0} \exp(-\lambda_i t)$ for $i = 1, \ldots, n$. Formula (5) implies that $\sum h_k \exp t\{(k_1 - k_{n+1})\lambda_1 + \cdots + (k_n - k_{2n})\lambda_n\} y^k$ is constant in t, and this implies that $\{(k_1 - k_{n+1})\lambda_1 + \cdots + (k_n - k_{2n})\}\lambda_n = 0$. But since the λ_i's are independent over the integers, this implies $k_1 = k_{n+1}, \ldots, k_n = k_{2n}$. That is, H^i is a function of the products $y_1 y_{n+1}, \ldots, y_n y_{2n}$ only.

By the remark above, the kernel consists of those functions which depend only on $I_1, \ldots, I_n$ and not on the angles in action-angle variables. Therefore, the extra condition (iii) of Theorem B.4 holds, and the normal form is unique. ∎

Remark. If the condition (14) only holds for $k_i s$ with $|k_1| + \cdots + |k_n| \leq N$, then the terms in the Hamiltonian up to the terms of order N can be put in normal form, and these terms are unique.

2. The General Equilibria

Recently there has been a lot of progress on normal forms in the case when A is not simple, and the research goes on. In the 1970s, the question of the stability of the Lagrange triangular point $\mathscr{L}_4$ was studied intensely. For

Hamiltonian systems, it is not enough to look at the linearized system alone because the higher-order terms in the normalized equations can effect the stability (see the discussion in Chapter VIII). The matrix of the linearization of the equations at $\mathscr{L}_4$ when $\mu = \mu_1$ is not simple as was seen in Section II.G. The normal form for this case, and other similar cases was carried out by the Russian school; see Sokol'skij (1978). First Kummer (1976, 1978) and then Cushman, Deprit, and Mosak (1983) used group representation theory. Representation theory is very helpful in understanding the general case, but there are simpler ways to understand the basic ideas and examples. In Meyer (1984b) a theorem like theorem 1 above was given for non-Hamiltonian systems but A was replaced by A^T in (5); so, the terms in the normal form are invariant under the flow $\exp(A^T t)$. A far better proof can be found in Elphick et al. (1987), which is what we will present here.

The proof of Theorem 1 rested on the fact that for a simple matrix, A, $\mathbb{R}^{2n}$ is the direct sum of the range and kernel of A, and this held true for the operator $L = \{H_0^0, \cdot\}$ defined on homogeneous polynomials as well. The method of Elphick et al. is based on the following simple lemma in linear algebra known as the Fredholm alternative and an inner product defined on homogeneous polynomials given after the lemma.

Lemma 3. *Let $\mathbb{V}$ be a finite-dimensional inner product space with inner product. Let $A: \mathbb{V} \to \mathbb{V}$ be a linear transformation, and A^* its adjoint [so $(Ax, y) = (x, A^*y)$ for all $x, y \in \mathbb{V}$]. Then $\mathbb{V} = R \oplus K^*$ where R is the range of A and K^* is the kernel of A^*.*

PROOF. Let $x \in R$; so, there is a $u \in \mathbb{V}$ such that $Au = x$. Let $y \in K^*$; so, $A^*y = 0$. Since $0 = (u, 0) = (u, A^*y) = (Au, y) = (y, x)$, it follows that R and K are orthogonal subspaces. Let K be the kernel of A. In a finite-dimensional space, $\dim \mathbb{V} = \dim R + \dim K$ and $\dim K = \dim K^*$. Since R and K^* are orthogonal, $\dim(R + K^*) = \dim R + \dim K^* = \dim \mathbb{V}$; so, $\mathbb{V} = R \oplus K^*$. ∎

Let $\mathscr{P} = \mathscr{P}_i$ be the linear space of all homogeneous polynomials of degree i in $2n$ variables $x \in \mathbb{R}^{2n}$. So if $P \in \mathscr{P}$, then

$$P(x) = \sum_{|k|=i} p_k x^k = \sum_{|k|=i} p_{k_1 k_2 \ldots k_{2n}} x_1^{k_1} x_2^{k_2} \ldots x_{2n}^{k_{2n}}. \tag{17}$$

Define $P(\partial)$ to be the differential operator

$$P(\partial) = \sum_{|k|=i} p_k \frac{\partial^k}{\partial x^k}, \tag{18}$$

where we have introduced the notation

$$\frac{\partial^k}{\partial x^k} = \frac{\partial^{k_1}}{\partial x_1^{k_1}} \frac{\partial^{k_2}}{\partial x_2^{k_2}} \cdots \frac{\partial^{k_{2n}}}{\partial x_{2n}^{k_{2n}}}. \tag{19}$$

Let $Q \in \mathscr{P}$, $Q(x) = \sum q_h x^h$ be another homogeneous polynomial, and define an inner product $\langle \cdot, \cdot \rangle$ on $\mathscr{P}$ by

$$\langle P, Q\rangle = P(\partial)Q(x). \tag{20}$$

To see that this is indeed an inner product, note that $\partial^k x^h/\partial x^k = 0$ if $k \neq h$ and $\partial^k x^h/\partial x^k = k! = k_1!k_2!\cdots k_{2n}!$ if $k = h$; so,

$$\langle P, Q\rangle = \sum_{|k|=i} k! p_k q_k. \tag{21}$$

Let $A = JS$ be a Hamiltonian matrix where S is a symmetric matrix of the quadratic Hamiltonian H^0; so, $H^0(x) = \frac{1}{2}x^T Sx$. From Corollary B.3 and the proof of Theorem C.1, the operator of importance is $L_A\colon \mathscr{P} \to \mathscr{P}$, where

$$L_A P = \{H_0^0, P\} = -\frac{\partial P}{\partial x}Ax = \frac{d}{dt}P(e^{At}x)\bigg|_{t=0}. \tag{22}$$

Lemma 4. *Let $A\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be as above and A^T its transpose (so A^T is the adjoint of A with respect to the standard inner product in $\mathbb{R}^{2n}$). Then for all P, $Q \in \mathscr{P}$,*

$$\langle P(x), Q(Ax)\rangle = \langle P(A^T x), Q(x)\rangle \tag{23}$$

and

$$\langle P, L_A Q\rangle = \langle L_{A^T} P, Q\rangle. \tag{24}$$

That is, the adjoint of L_A with respect to $\langle\cdot,\cdot\rangle$ is L_{A^T}.

Proof. Equation (23) follows from (22) because (22) implies $\langle P(x), Q(e^{At}x)\rangle = \langle P(e^{A^T t}x), Q(x)\rangle$. Differentiating this last expression with respect to t and setting $t = 0$ gives (24).

Let $y = Ax$ $(y^i = \sum_j A_j^i x^j)$ and $F(y) = F(Ax)$. Since $\partial F(y)/\partial x^j = \sum_i (\partial F(y)/\partial y^i)(\partial y^i/\partial x^j) = \sum_i (\partial F(y)/\partial y^i)A_j^i$, it follows that $\partial/\partial x = A^T \partial/\partial y$. $\langle P(x), Q(Ax)\rangle = P(\partial_x)Q(Ax) = P(A^T\partial_y)Q(y) = \langle P(A^T y), Q(y)\rangle$. ∎

Theorem 5. *Let A be a Hamiltonian matrix. Then there exists a formal, symplectic change of variables, $x = X(y) = y + \cdots$, which transforms the Hamiltonian* (1) to

$$H^{\#}(y) = \sum_{i=0}^{\infty} H^i(y), \tag{25}$$

where H^i is a homogeneous polynomial of degree $i + 2$ such that

$$H^i(e^{A^T t}y) \equiv H^i(y), \tag{26}$$

for all $i = 0, 1, \ldots$, all $y \in \mathbb{R}^{2n}$, and all $t \in \mathbb{R}$.

Remark. Let $H_0^T(x) = H_T^0(x) = \frac{1}{2}x^T Rx$ be the quadratic Hamiltonian for the adjoint linear equation; so, $A^T = JR$. Then (26) is equivalent to

$$\{H^i, H_T^0\} = 0 \tag{27}$$

for $i = 1, 2, \ldots$.

PROOF. By Corollary B.3, we must solve Equations (B.25) or $L_A C + D = B$, where $D \in P_i = \mathscr{P}$ is given, and $C \in Q_i = \mathscr{P}$, and $D \in Q_i = \text{kernel}(L_{A^T})$. By Lemma 4, we can write $D = B - G$, where $B \in \text{kernel}(L_{A^T})$; so, $\{B, H_T^0\} = 0$, and $G \in \text{range}(L_A)$; so, $G = L_A C$, $C \in \mathscr{P}$. With these choices, Equation (B.25) is solved. Verification of the rest of the hypothesis in Corollary B.3 is just as in the proof of Theorem 1. ∎

Theorem 1 is a corollary of this theorem because when A is simple, it is diagonalizable, and so, its own adjoint. We proved Theorem 1 separately because the proof is constructive.

3. Examples of Normal Forms in the Nonsimple Case

Consider the Hamiltonian system (1), where $n = 1$ and $x = (q, p)$. Let

$$\begin{aligned} H_0(q, p) &= \tfrac{1}{2}p^2, \qquad H_0^T(q, p) = \tfrac{1}{2}q^2, \\ A &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad A^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{28}$$

Since $\exp(A^T t) = \begin{pmatrix} 1 & 1+t \\ 0 & 1 \end{pmatrix}$, (24) implies that the higher-order terms in the normal form are independent of p, or $H^i = H^i(p, \cdot)$. Thus, the Hamiltonian in normal form is $\frac{1}{2}p^2 + G(q)$, which is the Hamiltonian for the second-order equation $\ddot{q} + g(q) = 0$, where $g(q) = \partial G(q)/\partial q$.

Now consider a Hamiltonian system with two degrees of freedom with a linearized system with repeated pure imaginary roots that are nonsimple. In Section II.D, the normal form for the quadratic part of such a Hamiltonian was given as

$$H_0 = \omega(\xi_2\eta_1 - \xi_1\eta_2) + \tfrac{1}{2}\delta(\xi_1^2 + \xi_2^2), \tag{29}$$

where $\omega \neq 0$ and $\delta = \pm 1$. The linearized equations are

$$\begin{bmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \\ \dot{\eta}_1 \\ \dot{\eta}_2 \end{bmatrix} = \begin{bmatrix} 0 & \omega & 0 & 0 \\ -\omega & 0 & 0 & 0 \\ -\delta & 0 & 0 & \omega \\ 0 & -\delta & -\omega & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \eta_1 \\ \eta_2 \end{bmatrix}. \tag{30}$$

The transpose of (30) is defined by the Hamiltonian

$$H_0^T = \omega(\xi_2\eta_1 - \xi_1\eta_2) - \tfrac{1}{2}\delta(\eta_1^2 + \eta_2^2), \tag{31}$$

and the transposed equations are

$$\begin{bmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \\ \dot{\eta}_1 \\ \dot{\eta}_2 \end{bmatrix} = \begin{bmatrix} 0 & -\omega & -\delta & 0 \\ \omega & 0 & 0 & -\delta \\ 0 & 0 & 0 & -\omega \\ 0 & 0 & \omega & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \eta_1 \\ \eta_2 \end{bmatrix}. \tag{32}$$

Sokol'skij (1978) suggested changing to polar coordinates (see Section IV.C.6) to make the transposed equations simple. That is, he changed coordinates by

$$\begin{aligned} \eta_1 &= r\cos\theta, & R &= (\xi_1\eta_1 + \xi_2\eta_2)/r, \\ \eta_2 &= r\sin\theta, & \Theta &= \eta_1\xi_2 - \eta_2\xi_1. \end{aligned} \tag{33}$$

In these coordinates,

$$H_0^T = -\omega\Theta + \tfrac{1}{2}\delta r^2, \qquad H_0 = \omega\Theta + \tfrac{1}{2}\delta(R^2 + \Theta^2/r^2), \tag{34}$$

and the transposed equations are

$$\dot{r} = 0, \qquad \dot{\theta} = \omega, \qquad \dot{R} = \delta r, \qquad \dot{\Theta} = 0. \tag{35}$$

By (26) and (27), the higher-order terms in the normal form are independent of θ and R and so depend only on $r^2 = \eta_1^2 + \eta_2^2$ and $\Theta = \eta_1\xi_2 - \eta_2\xi_1$.

Thus, the theory of the normal form in this case depends on three qualities

$$\Gamma_1 = \xi_2\eta_1 - \xi_1\eta_1, \qquad \Gamma_2 = \tfrac{1}{2}(\xi_1^2 + \xi_2^2), \qquad \Gamma_3 = \tfrac{1}{2}(\eta_1^2 + \eta_2^2). \tag{36}$$

The Hamiltonian $H_0 = \omega\Gamma_1 + \Gamma_2$ and the higher-order terms in the normal form are functions of Γ_1 and Γ_3 only. This is known as *Sokol'skij's* normal form.

D. Normal Form at $\mathscr{L}_4$

Recall that in Section II.G we showed that the linearization of the restricted 3-body problem at the Lagrange triangular point $\mathscr{L}_4$ had two pairs of pure imaginary eigenvalues, $\pm i\omega_1$, $\pm i\omega_2$, when $0 < \mu < \mu_1 = \frac{1}{2}(1 - \sqrt{69/9})$, and that there are symplectic coordinates so that the quadratic part of the Hamiltonian is

$$H_2 = \omega_1 I_1 - \omega_2 I_2, \tag{1}$$

where I_1, I_2, ϕ_1, ϕ_2 are action-angle variables.

Recall that in Section V.E.6 we defined μ_r to be the value of μ for which $\omega_1/\omega_2 = r$, and that $0 \cdots < \mu_3 < \mu_2 < \mu_1$. When $0 < \mu < \mu_1$, and $\mu \neq \mu_2, \mu_3$ then, by Corollary C.2, the Hamiltonian of the restricted 3-body problem can be normalized through the fourth-order terms; so, the Hamiltonian becomes

$$H = \omega_1 I_1 - \omega_2 I_2 + \tfrac{1}{2}(AI_1^2 + 2BI_1I_2 + CI_2^2) + \cdots. \tag{2}$$

After six months of hand calculations, Deprit and Deprit–Bartholome computed:

$$\begin{aligned} A &= \frac{1}{72}\omega_2^2\frac{(81 - 696\omega_1^2 + 124\omega_1^4)}{(1 - 2\omega_1^2)^2(1 - 5\omega_1^2)}, \\ B &= -\frac{1}{6}\frac{\omega_1\omega_2(43 + 64\omega_1^2\omega_2^2)}{(1 - 2\omega_1^2)(1 - 2\omega_2^2)(1 - 5\omega_1^2)(1 - 5\omega_2^2)}, \\ C(\omega_1, \omega_2) &= A(\omega_2, \omega_1). \end{aligned} \tag{3}$$

Meyer and Schmidt (1986a) computed the normal form through terms of sixth order by machine. The results are too lengthy to reproduce here. It did serve as an independent check of the calculations of Deprit and Deprit–Bartholome. In Section II.G, the quadratic part of the Hamiltonian of the restricted 3-body problem at $\mathscr{L}_4$ for $\mu = \mu_1$ was brought into normal form by a linear symplectic change of coordinates. In these coordinates, the quadratic part of the Hamiltonian is of the form

$$H_0 = \omega(\xi_2\eta_1 - \xi_1\eta_2) + \tfrac{1}{2}(\xi_1^2 + \xi_1^2) = \omega\Gamma_1 + \Gamma_2, \tag{4}$$

where $\omega = \sqrt{2}/2$ and $\delta = +1$.

The normal form for the Hamiltonian of the restricted 3-body problem at $\mathscr{L}_4$ for $\mu = \mu_1$ is of the form

$$\begin{aligned} H &= \omega\Gamma_1 + \Gamma_2 + c\Gamma_1^2 + 2d\Gamma_1\Gamma_3 + \Gamma_3^2 + \cdots \\ &= \omega(\xi_2\eta_1 - \xi_1\eta_2) + \tfrac{1}{2}(\xi_1^2 + \xi_1^2) \\ &\quad + c(\xi_2\eta_1 - \xi_1\eta_2)^2 + d(\eta_1^2 + \eta_2^2)(\xi_2\eta_1 - \xi_1\eta_2) + \tfrac{1}{4}e(\eta_1^2 + \eta_2^2)^2 + \cdots, \end{aligned} \tag{5}$$

where c, d, and e are constants. As another related problem, consider a quadratic Hamiltonian $Q(y, \varepsilon)$ which depends on a parameter ε, which for $\varepsilon = 0$ is H_0. That is, $Q(y, \varepsilon) = Q_0(y) + \varepsilon Q_1(y) + \cdots$, where $Q_0 = H_0$ in (C.29). Then this Hamiltonian can be brought into normal form to an order so that Q_1, $Q_2, \ldots$ depend only on Γ_1 and Γ_3. [see Schmidt (1990) for the calculations.]

The quadratic part of the Hamiltonian of the restricted 3-body problem at the Lagrange triangular point, $\mathscr{L}_4$, for values of the mass ratio parameter $\mu = \mu_1 + \varepsilon$ can be brought into normal form by a linear symplectic change of coordinates. The normal form up to order 4 looks like

$$\begin{aligned} Q &= \omega\Gamma_1 + \Gamma_2 + \varepsilon\{a\Gamma_1 + b\Gamma_3\} + \cdots \\ &= \omega(\xi_2\eta_1 - \xi_1\eta_2) + \tfrac{1}{2}(\xi_1^2 + \xi_1^2) \\ &\quad + \varepsilon\{a(\xi_2\eta_1 - \xi_1\eta_2) + \tfrac{1}{2}b(\eta_1 + \eta_2)\} + \cdots. \end{aligned} \tag{6}$$

Schmidt (1990) calculated that

$$a = 3\sqrt{2}\sqrt{69}/16, \qquad b = 3\sqrt{69}/8. \tag{7}$$

E. Normal Forms for Periodic Systems and Diffeomorphisms

This section contains a series of reductions which reduce the study of the normal forms for symplectomorphisms to the study of normal forms of periodic systems. Then as examples, the normal forms for symplectomorphisms of the plane are given in preparation for the study of generic bifurcations of fixed points given in the Chapter VIII.

1. The Reduction

The study of a neighborhood of a periodic solution of an autonomous, Hamiltonian system was reduced to the study of the Poincaré map in an energy surface by the discussion in Section V.E. This Poincaré map is a symplectomorphism with a fixed point corresponding to the periodic orbit.

Let the origin be a fixed point for the symplectomorphism

$$\Psi(x) = \Gamma x + \psi(x), \tag{1}$$

where Γ is a $2n \times 2n$ symplectic matrix, and ψ is higher order, i.e., $\psi(0) = \partial\psi(0)/\partial x = 0$. By Theorem V.B.1 and the discussion following that theorem, if Γ has a logarithm, then (1) is the period map of a periodic Hamiltonian system. Since $\Psi^2(x) = \Gamma^2 x + \cdots$, and Γ^2 always has a logarithm, if Ψ is not a period map, then Ψ^2 is. Except for one example given at the end of this chapter, only the case when Γ has a real logarithm will be treated here.

Given a periodic system, by the Floquet–Lyapunov theorem (see Theorem II.E.3 and the discussion following it), there is a linear, symplectic, periodic change of variables which makes the linear part of the Hamiltonian equations constant in t. Thus, the study of symplectomorphisms near a fixed point is equivalent to studying a 2π-periodic Hamiltonian system of the form

$$H(t, x) = H_{\#}(t, x) = \sum_{i=0}^{\infty} H_i(t, x), \tag{2}$$

where H_i is a homogeneous polynomial in x of degree $i + 2$ with 2π-periodic coefficients, and $H_0(x) = \frac{1}{2}x^T S x$ where S is a $2n \times 2n$ constant, real, symmetric matrix, and $A = JS$ is a constant, real, Hamiltonian matrix. The linearized equations about the critical point $x = 0$ are

$$\dot{x} = Ax = JSx = J\nabla H_0(x), \tag{3}$$

and the general solution of (3) is $\phi(t, \xi) = \exp(At)\xi$.

2. The General Periodic Case

In this subsection, the generalization of the general normal form given in Section C.2 is extended to periodic systems. As before, we consider the periodic system (2) but no longer assume that the linear system is simple. First let us consider the generalization of Theorem C.5.

Consider the 2π-periodic equations

$$\dot{x} = A(t)x + f(t), \tag{4}$$

$$\dot{x} = A(t)x, \tag{5}$$

$$\dot{y} = -A(t)^T y. \tag{6}$$

Equation (5) is the homogeneous equation corresponding to the nonhomogeneous equation (4), and (6) is the adjoint equation of (5).

Lemma 1. *The nonhomogeneous equation* (4) *has a* 2π*-periodic solution* $\phi(t)$ *if and only if*

$$\int_0^{2\pi} y^T(s) f(s)\, ds = 0, \tag{7}$$

for all 2π*-periodic solutions* $y(t)$ *of the adjoint equation* (6).

PROOF. Let $x(t, x_0)$ be the solution of (4) with $x(0, x_0) = x_0$. Then

$$x(t, x_0) = X(t)x_0 + \int_0^t X(t)Y^T(s)f(s)\, ds, \tag{8}$$

where $X(t)$ and $Y(t)$ are the fundamental matrix solutions of (5) and (6), respectively, because $X^{-1} = Y^T$. The solution is 2π-periodic if and only if $x(t, x_0) = x_0$, or

$$\begin{aligned} &Bx_0 = g, \\ &B = I - X(2\pi), \qquad g = \int_0^{2\pi} X(2\pi)Y^T(s)f(s)\, ds. \end{aligned} \tag{9}$$

By Lemma C.3, the linear equation $Bx_0 = g$ has a solution if and only if $v^T g = 0$ for all v with $B^T v = 0$. That is, there is a 2π periodic solution if and only if

$$\int_0^{2\pi} v^T X(2\pi)Y^T(s)f(s)\, ds = 0 \quad \text{for all } v \text{ with } X(2\pi)^T v = v. \tag{10}$$

But if $X(2\pi)^T v = v$, then the integral in (10) is $\int_0^{2\pi} v^T Y^T(s)f(s)\, ds = 0$. But $X(2\pi)^T v = v$ if and only if $Y(2\pi)v = v$ and if and only if $Y(s)v$ is a 2π-periodic solution of (6). ∎

Consider the periodic Hamiltonian system (2). Scale by $x \to \varepsilon x$ as in the proof of Theorem C.1, and use the same notation for the scaled Hamiltonian. By Theorem C.3 we must define spaces $\mathscr{P}_i$, $\mathscr{Q}_i$, and $\mathscr{R}_i$ with $\mathscr{Q}_i \subset \mathscr{P}_i$, $H_i^0 \in \mathscr{P}_i$, $H_0^i \in \mathscr{Q}_i$, $W_i \in \mathscr{R}_i$. The Lie equation to be solved in this case is

$$E = D + \{H_0^0, C\} - \frac{\partial C}{\partial t}, \tag{11}$$

where D is given in $\mathscr{P}_i$, and we are to find $E \in \mathscr{Q}_i$ and $C \in \mathscr{R}_i$.

Let B be the adjoint of A, i.e., the transpose in the real case. Define $K(x) = \frac{1}{2}x^T Rx$, where $B = JR$; so, K is the Hamiltonian of the adjoint linear system. Let $\mathscr{P}_i$ be the space of polynomials in x with coefficients which are smooth 2π-periodic functions of t. Let $\mathscr{L} = \{H_0^0, \cdot\}: \mathscr{P}_i \to \mathscr{P}_i$, and let $\mathscr{J} = \{K, \cdot\}: \mathscr{P}_i \to \mathscr{P}_i$. $\mathscr{F}$ is the adjoint of $\mathscr{L}$ if we use the metric defined by Elphick et al. that was used in Section C.2. Therefore, given D, (11) has a unique 2π-periodic solution, C, where E is a 2π-periodic solution of the homogeneous, adjoint equation

$$0 = \{K, E\} + \frac{\partial E}{\partial t}. \tag{12}$$

Characterizing the 2π-periodic solutions of (12) characterizes the normal form. Expand the elements of $\mathscr{P}_i$ in Fourier series. Let $E = d(x)e^{imt}$, and substitute into (12) to get

$$0 = \{K, e\} + imd. \tag{13}$$

Thus, one characterization of the normal form is in terms of the eigenvectors of $\mathscr{I} = \{K, \cdot\}: \mathscr{P}_i \to \mathscr{P}_i$. That is, $\mathscr{Q}_i$ has a basis of the form $\{d(x)e^{imt}$: d is an eigenvector of $\mathscr{I}$ corresponding to the eigenvalue $im\}$.

Theorem 2. *Let $H^0(x) = H_0(x) = (1/2)x^T Sx$, where $A = JS$ is an arbitary, constant Hamiltonian matrix, and let B be the adjoint of A. Then there exists a formal, symplectic, 2π-periodic change of variables $x = X(t, y) = y + \cdots$ which transforms the Hamiltonian* (2) *to the Hamiltonian system*

$$\dot{y} = J\nabla H^{\#}(t, y), \qquad H^{\#}(t, y) = \sum_{i=0}^{\infty} H^i(t, y), \tag{14}$$

where

$$\{H^i, K\} + \frac{\partial H^i}{\partial t} = 0 \quad \text{for } i = 1, 2, 3, \ldots, \tag{15}$$

or equivalently,

$$H^i(t, e^{Bt}x) \equiv H^i(0, x) \quad \text{for } i = 1, 2, 3, \ldots. \tag{16}$$

Corollary 3. *Let A be simple and have eigenvalues $\lambda_1, \ldots, \lambda_n, -\lambda_1, \ldots, -\lambda_n$. Assume that $\lambda_1, \ldots, \lambda_n$ and i are independent over the integers, i.e. there is no relation of the form $k_1\lambda_1 + \cdots + k_n\lambda_n = mi$, where $k_1, \ldots, k_n$ and m are integers. Then there exists a formal, symplectic, 2π-periodic change of variables $x = X(t, y) = y + \cdots$ which transforms the Hamiltonian* (2) to *an autonomous Hamiltonian system*

$$\dot{y} = J\nabla H^{\#}(y), \qquad H^{\#}(y) = \sum_{i=0}^{\infty} H^i(y), \tag{17}$$

where $H^0 = H_0$, and

$$\{H^i, H^0\} = 0, \tag{18}$$

or equvivalently,

$$H^i(e^{At}y) \equiv H^i(y) \tag{19}$$

for all $i = 0, 1, 2, \ldots, y \in \mathbb{R}^{2n}, t \in \mathbb{R}$.

PROOF. Let $A = B = \operatorname{diag}(\lambda_1, \ldots, -\lambda_n)$. A typical term in the normal form given by Theorem 2 is of the form $h(t, x) = h_k e^{imt}x^k$. Applying (16) to this term

gives $h_k \exp\{im + (k_1 - k_{n+1})\lambda_1 + \cdots + (k_n - k_{2n})\lambda_n\}t = 0$. By the assumption on the independence, this can only hold if $m = 0$, $k_1 = k_{n+1}, \ldots, k_n = k_{2n}$. Thus, the Hamiltonian is in the normal form of Birkhoff as described in Corollary C.2. ∎

Corollary 4. *Let Γ be simple and have a real logarithm. Then there exists a formal, near identity, symplectic change of variables $x \to y$ such that in the new coordinates the symplectomorphism in* (1) *is of the form*

$$\Phi(y) = \Gamma y + \phi(y), \tag{20}$$

where

$$\phi(\Gamma y) \equiv \Gamma\phi(y) \quad \text{or} \quad \Phi(\Gamma y) \equiv \Gamma\Phi(y). \tag{21}$$

PROOF. Let $\Gamma = \exp(2\pi A)$. Since Γ is simple, so is A, and therefore it can be taken as its own adjoint. Then by the reduction given above, the map (1) is the period map of a system of Hamiltonian differential equations. Assume that the symplectic change of coordinates has been made so that the Hamiltonian is in normal form, and let the equations in these coordinates be $\dot{y} = Ay + f(t, y)$. Condition (16) implies $f(t, e^{At}x) = e^{At}f(0, x)$, and this implies $f(t, \Gamma x) = \Gamma f(t, x)$. Let $\xi(t, \eta)$ be a solution of this equation with $\xi(0, \eta) = \eta$. Define $\zeta(t, \eta) = \Gamma\xi(t, \Gamma^{-1}\eta)$, so $\xi(0, \eta) = \zeta(0, \eta) = \eta$. $\dot{\zeta} = \Gamma\{A\xi + f(t, \xi)\} = A\Gamma\xi + \Gamma f(t, \xi) = A\Gamma\xi + f(t, \Gamma\xi) = A\zeta + f(t, \zeta)$. By the uniqueness theorem for ordinary differential equations, $\xi(t, \eta) = \zeta(t, \eta) = \Gamma\xi(t, \Gamma\eta)$; so, the period map satisfies (21). ∎

3. General Hyperbolic and Elliptic Points

Consider as examples the case when $n = 1$; so, Ψ in (1) is a symplectomorphism of the plane with a fixed point at the origin.

First, consider the case when Γ has eigenvalues μ, μ^{-1}, where $0 < \mu < 1$, i.e., the origin is a hyperbolic fixed point. By Lemma II.E.10, there are symplectic coordinates, say x, so that

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

Let $2\pi\alpha = \ln \mu$; so,

$$\Gamma = \exp\begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} 2\pi \quad \text{and} \quad A = \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix}.$$

By the discussion give above, the symplectomorphism Ψ is the period map of the 2π-periodic system (2) with $H_0(x) = \alpha x_1 x_2$. By Corollary 3, there is a formal, 2π-periodic, symplectic change of variables, $x \to y$, which transforms (2) to the autonomous system (17) with (19) holding. Since the solution of the

linear system is $y_1(t) = y_{10}e^{\alpha t}$, $y_2(t) = y_{20}e^{-\alpha t}$, the condition (19) implies that the Hamiltonian (17) is a function of the product y_1y_2 only. Let $H^\#(y) = K^\#(y_1y_2) = \alpha y_1y_2 + K(y_1y_2)$. By the above discussion, the normal form for (1) is the time 2π-map of the autonomous system whose Hamiltonian is $K^\#$. The equations defined by $K^\#$ are

$$\begin{aligned} \dot{y}_1 &= y_1(\alpha + k(y_1y_2)), \\ \dot{y}_2 &= -y_2(\alpha + k(y_1y_2)), \end{aligned} \tag{22}$$

where k is the derivative of K. Equations (22) have y_1y_2 as an integral, and so the equations are solvable, and the solution is

$$\begin{aligned} y_1(t) &= y_{10} \exp(\alpha + k(y_1y_2))t, \\ y_2(t) &= y_{20} \exp[-(\alpha + k(y_1y_2))t]. \end{aligned} \tag{23}$$

Thus, the normal form for (1) in this case is

$$\Psi(y) = \begin{pmatrix} y_1 g(y_1y_2) \\ y_2 g(y_1y_2)^{-1} \end{pmatrix}, \tag{24}$$

where g has a formal expansion $g(u) = \mu u + \cdots$. If a symplectomorphism is in this form, then the origin is called a *general hyperbolic point.* This map takes the hyperbolas $y_1y_2 =$ constant into themselves. In this case, the transformation to normal form converges by a classical theorem of Moser (1956).

Next consider the case when A has eigenvalues $\lambda = \alpha + \beta i$, $\bar{\lambda} = \alpha - \beta i$, where $\alpha^2 + \beta^2 = 1$, $\beta \neq 0$, i.e., the origin is an elliptic fixed point. By Corollary II.D.9 there are symplectic coordinates, say x, so that

$$\Gamma = \begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix}.$$

Let

$$\Gamma = \exp 2\pi \begin{pmatrix} \omega i & 0 \\ 0 & -\omega i \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} \omega i & 0 \\ 0 & -\omega i \end{pmatrix}$$

in either case. Assume that ω is not an integer, i.e., λ is not a root of unity. By the discussion given above, the symplectomorphism Ψ is the period map of the 2π-periodic system (2) with $H_0(x) = i\omega x_1x_2$. By Corollary 3, there is a formal, 2π-periodic, symplectic change of variables, $x \to y$, which transforms (2) to the autonomous system (17) satisfying (19). Equation (19) implies that the Hamiltonian is a function of y_1y_2 only. Let $H^\#(y) = K^\#(y_1y_2) = i\omega y_1y_2 + iK(y_1y_2)$. By the above discussion, the normal form for (1) is the time 2π-map of the autonomous system whose Hamiltonian is $K^\#$. Change to action-angle variables (I, ϕ); so, the Hamiltonian becomes $H^\#(I, \phi) = K^\#(I) = \omega I + K(I)$. The equations defined by $K^\#$ are

$$\begin{aligned} \dot{I} &= 0, \\ \dot{\phi} &= -\omega - k(I), \end{aligned} \tag{25}$$

where k is the derivative of K. Equations (25) have I as an integral, and so the equations are solvable, and the solution is

$$\begin{aligned} I(t) &= I_0, \\ \phi(t) &= \phi_0 + (-\omega + k(I_0))t. \end{aligned} \tag{26}$$

Thus, the normal form for (1) in action-angle variable in this case is

$$\Psi(I, \phi) = \begin{pmatrix} I \\ \phi + g(I) \end{pmatrix}, \tag{27}$$

where g has a formal expansion $g(I) = -\omega + \beta I \ldots$. If a symplectomorphism is in this form with $\beta \neq 0$, then the origin is called a *general elliptic point*, or Ψ is called a *twist map*. The map (27) takes circles into circles and rotates each circle by an amount $g(I)$.

4. Higher Resonance in the Planar Case

Let us consider the case when $n = 1$, and the symplectomorphism Ψ has an elliptic fixed point whose multiplier is a root of unity. Theorem 2 and Corollary 4 apply as well.

Let Γ have eigenvalues $\lambda = \alpha + \beta i$, $\bar{\lambda} = \alpha - \beta i$, where λ is a kth root of unity; so, $\lambda^k = 1$, $k > 2$, and $\lambda = \exp(h2\pi i/k)$, where h is an integer. The origin is called a *k-resonance elliptic point* in this case. By Corollary II.D.9, there are symplectic coordinates, say x, so that

$$\Gamma = \begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix};$$

then

$$\Gamma = \exp 2\pi \begin{pmatrix} hi/k & 0 \\ 0 & -hi/k \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} hi/k & 0 \\ 0 & -hi/k \end{pmatrix}.$$

Since A is diagonal, it is its own adjoint. By the discussion given above, the symplectomorphism Ψ is the period map of the 2π-periodic system (2) with $H_0(x) = (hi/k)(x_1 x_2)$, where the reality condition is $\bar{x}_1 = x_2$. The normal form for the Hamiltonian is given by Theorem 2.

Let $h(t, x)$ be a typical term in the normal form expansion; so,

$$h = e^{ist} x_1^{m_1} x_2^{m_2}. \tag{28}$$

The term h satisfies (16) if and only if

$$(hi/k)(m_1 - m_2) + si = 0; \tag{29}$$

so it is in the normal form if it is of the form

$$(x_1 x_2)^m \quad \text{or} \quad x_1^{m_1} x_2^{m_2} e^{-rit}, \tag{30}$$

where $r = (m_1 - m_2)h/k$, and m, m_1, m_2, and r are integers.

In action-angle coordinates (I, ϕ), $H^0(I, \phi) = (h/k)I$, and the solution of the linear system is $I = I_0, \phi = \phi_0 - (h/k)t$. Thus, $H^{\#}(t, I, \phi)$ is a function of I and $(k\phi + ht)$; so, let $H^{\#}(t, I, \phi) = K^{\#}(I, k\phi + ht) = (h/k)I + K(I, k\phi + ht)$.

The lowest-order terms in (28) which contain t, the new terms, are $x_1^k e^{-hit}$ and $x_2^k e^{hit}$. In action-angle coordinates these terms are like $I^{k/2}\cos(k\phi + ht)$ and $I^{k/2}\sin(k\phi + ht)$. Thus, the normalized Hamiltonian is a function of I and $(k\phi + ht)$ only, and it is of the form

$$H^{\#}(t, I, \phi) = (h/k)I + aI^2 + bI^3 + \cdots + I^{k/2}\{\alpha\cos(k\phi + ht) + \beta\sin(k\phi + ht)\} + \cdots. \tag{31}$$

The equations of motion are

$$\begin{aligned} \dot{I} &= I^{k/2}\{-\alpha\sin(k\phi + ht) + \beta\cos(k\phi + ht)\} + \cdots, \\ \dot{\phi} &= -(h/k) - 2aI - \cdots - (k/2)I^{(k-2)/2}\{\alpha\cos(k\phi + ht) + \beta\sin(k\phi + ht)\} + \cdots. \end{aligned} \tag{32}$$

By a rotation, $\phi \to \phi + \delta$; the first sin term can be absorbed into the cos term, so there is no loss in generality in assuming that $\beta = 0$ in (31) and (32). Henceforth, we will assume this rotation has been made, and so, $\beta = 0$.

Note that in the $\dot{\phi}$ equation in (32), there are two nonlinear terms. When $k > 4$, the term which contains the angle is of higher order in I, whereas for $k = 3$ it is lower order. When $k = 4$, the two terms are both of order I^1. We shall see in Chapters VIII that the cases when $k = 3$ or 4 must be treated separately.

The 2π-map is then of the form

$$\begin{aligned} I &= I_0 - \alpha I_0^{k/2}\sin(k\phi_0) + \cdots, \\ \phi &= \phi_0 - (2\pi h/k) - 4\pi a I_0 + \alpha\pi k I^{(k-2)/2}\cos(k\phi_0) + \cdots. \end{aligned} \tag{33}$$

5. Normal Forms when Multipliers Are ± 1

Consider the case where the multiplier is $+1$ first. Since in this subsection no trigonometric functions will be used, assume that the periodic systems are periodic with period 1. If Γ has the eigenvalue $+1$, then either Γ is the identity, and A is the zero matrix, or there are symplectic coordinates such that $\Gamma = \left(\begin{smallmatrix} 1 & \pm 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $A = \left(\begin{smallmatrix} 0 & \pm 1 \\ 0 & 0 \end{smallmatrix}\right)$. In the first case, when $\Gamma = I$ and $A = 0$, Theorem 2 gives no information, and this is because the situation is highly degenerate and nongeneric.

Therefore, consider the case when $\Gamma = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$; so, the adjoint of A is $B = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ and $\exp Bt = \left(\begin{smallmatrix} 1 & 0 \\ t & 1 \end{smallmatrix}\right)$. Let $x = (u, v)$. Condition (16) of Theorem 2 is $H^i(u, v + ut, t) \equiv H^i(u, v, 0)$. This condition and the fact that H^i must be periodic in t implies that $H^i(u, v, t) = K^i(u)$. Thus, the normal form is

$$H^{\#}(t, u, v) = v^2/2 + K(u) = v^2/2 + \beta u^3/3 + \cdots \tag{34}$$

and the equations of motion are

$$\begin{aligned} \dot{u} &= v + \cdots, \\ \dot{v} &= -\frac{\partial K}{\partial u}(u) = -\beta u^2 + \cdots \end{aligned} \tag{35}$$

The period map is not so easy to compute and is not so simple. Fortunately, in applications, the critical information occurs at a very low order. By using the Lie transform methods discussed in the Problems Section one finds that the period map is

$$\begin{aligned} u &= u_0 + v_0 - (\beta/12)(6u_0^2 + 4uv_{00} + v_0^2) + \cdots, \\ v &= v_0 - (\beta/3)(3u_0^2 + 3u_0v_0 + v_0^2) + \cdots. \end{aligned} \tag{36}$$

Now consider the case when Γ has eigenvalue -1. In this case $\Gamma = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} -1 & \pm 1 \\ 0 & -1 \end{smallmatrix}\right)$. Consider the case when $\Gamma = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ first since it has a real logarithm, $\Gamma = \exp 2\pi A$, $A = \left(\begin{smallmatrix} 0 & 1/2 \\ -1/2 & 0 \end{smallmatrix}\right)$. This is almost the same as the higher-order resonance considered in the previous subsection. Condition (21) of Corollary 4 implies that the normal form in this case is simply an odd function. That is, $\Phi(y) = -y + \phi(y)$ is in normal form when $\phi(-y) = -\phi(y)$.

Now consider the case when $\Gamma = \left(\begin{smallmatrix} -1 & -1 \\ 0 & -1 \end{smallmatrix}\right)$. We shall make two changes of coordinates to bring this case to normal form. First, instead of the usual uniform scaling, scale by $x_1 \to \varepsilon x_1 x_2 \to \varepsilon^2 x_2$ so that the map (1) becomes $\Psi(x) = -x + O(\varepsilon)$. This nonuniform scaling moves the off diagonal term to the higher-order terms, and now the lead term is the same as discussed in the last paragraph. Thus, there is a symplectic change of coordinates $z = R(x)$ such that in the new coordinates, z, the map (1) is odd, i.e., $R \circ \Psi \circ R^{-1}(z) = \Xi(z) = \Gamma z + \cdots$ is odd.

Write $\Xi(z) = -\Lambda(z) = -\{\Omega z + \zeta(z)\}$, where $\Omega = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Now Ω is of the form discussed above, and so, there is a symplectic change of coordinates $y = S(z)$ which puts Λ in the normal form given by the time 1-map of a Hamiltonian system of the form (34), where now $K(u)$ is even. Since Λ is odd, the transformation S can be made odd also; see the Problems Section. Thus, $S \circ \Lambda \circ S^{-1} = \Theta$ is in the normal form given by the time 1-map of a Hamiltonian system of the form

$$H^{\#}(t, u, v) = v^2/2 + K(u) = v^2/2 + \beta u^4/4 + \cdots. \tag{37}$$

Using the method discussed in the Problems Section gives Θ: $(u_0, v_0) \to (u, v)$, where

$$\begin{aligned} u &= u_0 + v_0 - (\beta/120)(10u_0^3 + 10u_0^2v_0 + 5u_0v_0^2 + v_0^3) + \cdots, \\ v &= v_0 - (\beta/3)(4u_0^3 + 6u_0^2v_0 + 4u_0v_0^2 + v_0^3) + \cdots. \end{aligned} \tag{38}$$

Combining these changes of coordinates and using the fact that S is odd, it follows that $(S \circ R) \circ \Psi \circ (S \circ R)^{-1} = -\Theta$. That is, in the new coordinates, the map is just the negative of (38), or the normal form for the map is

$$\begin{aligned} u &= -u_0 - v_0 + (\beta/120)(10u_0^3 + 10u_0^2v_0 + 5u_0v_0^2 + v_0^3) + \cdots, \\ v &= -v_0 + (\beta/3)(4u_0^3 + 6u_0^2v_0 + 4u_0v_0^2 + v_0^3) + \cdots. \end{aligned} \tag{39}$$

F. Further Reading

Volume II of Poincaré (1899) contains a great many results from the nineteenth century on perturbation theory and the beginnings of normal forms. Birkhoff (1927) contains the normal forms for equilibrium points for autonomous and periodic systems in the case when the linear system is simple. The theory of Lie trnasforms can be found in Deprit (1969) and Henrard (1970a, b, c).

Normal forms at an equilibrium when the linear system is not simple is a current research problem. It starts with the Russian school; see Sokol'skij (1978). Kummer (1976, 1978) and Cushman, Deprit, and Mosak (1983) used group representation theory to find the normal forms in this case. Theorem C.5 is found in Meyer (1984b), but the proof comes from Elphick et al. (1987).

Problems

1. a. The normal forms for a Hamiltonian system with leading term $H_0^0(q, p) = p^2/2$ is $H^*(q, p) = p^2/2 + Q(q)$; see (C.28). This normal form also appears in Section E.5 when the case of multipliers equal to $+1$ is discussed. Carefully draw the phase portrait for the system with Hamiltonian $H(q, p) = p^2/2 + \beta q^3$ when $\beta = +1$ and -1. Also see Section E.5.

b. In Section E.5 when the multiplier -1 is discussed, the normal form was $H^*(q, p) = p^2/2 + Q(q)$ with Q even. Carefully draw the phase portrait for the system with Hamiltonian $H(q, p) = p^2/2 + \beta q^4$ when $\beta = +1$ and -1.

2. a. Compute the next term in the normal form of the unforced Duffing equation (B.1) by hand. Recall that H_0^0, H_1^0, H_0^1, and W_1 are given in Section B.1. [Hint: To get the next term you do not have to compute all of H_1^1, H_0^2 and W_2. H_0^2 is the term which is independent of ϕ in $H_1^1 + \{H_0^1, W_1\}$. Show that $\{H_0^1, W_1\}$ has no term independent of ϕ. Now $H_1^1 = H_2^0 + \{H_1^0, W_1\} + \{H_0^0, W_2\}$. Since $H_0^2 = 0$, you need to compute the term independent of ϕ in $\{H_0^1, W_1\}$.]

b. Using Macsyma, Reduce, Scratchpad, etc., find the first four terms in the normal form for the unforced Duffing equation. [Answer: $H^* = I + \varepsilon(3\gamma/8)I^2 + (\varepsilon^2/2)(17\gamma^2/32)I^3 + (\varepsilon^3/6)(5725\gamma^3/1024)I^4 + \cdots$.]

3. The Hamiltonian for Duffing's equation is of the form $(q^2 + p^2)/2 + P(p)$ where P is an even polynomial.

a. Show that such a Hamiltonian in action-angle variables is a Poisson series with only cosine terms.

b. Show that the Poisson bracket of two Poisson series, one of which is a cosine series and the other of which is a sine series, is always a cosine series.

c. Let H_j^i and W_i be from the normalization of such a Hamiltonian with an even potential. Show that H_j^i can always be taken as a cosine series and W_i as a sine series. [Hint: Define the spaces $\mathscr{P}_i$, $\mathscr{Q}_i$, and $\mathscr{R}_i$ of Theorem B.1.]

4. Consider a Hamiltonian differential equation of the form $\dot{x} = \varepsilon F_{\#}(\varepsilon, t, x) = \varepsilon F_1(t, x) + \varepsilon^2 F_2(t, x) + \cdots$, where F is T-periodic in t. Show that there is a formal symplectic series expansion $x = X(\varepsilon, t, y) = y + \cdots$ *which is T-periodic in* t and

transforms the equation to the autonomous Hamiltonian system $\dot{x} = \varepsilon F^{\#}(x) = \varepsilon F^1(x) + \varepsilon^2 F^2(x) + \cdots$. Show that $F^1(x) = (1/T)\int_0^T F_1(\tau, x)\, d\tau$, i.e., F^1 is the average of F_1 over a period. This is called the method of averaging; see Bogoliubov and Mitropolskii (1961). [Hint: Use Theorem E.2 and remember $F_0^0 = 0$.]

5. Use the notation of the previous problem. Show that if $F^1(\xi) = 0$ and $\partial F^1(\xi)/\partial x$ is nonsingular, then the equation $\dot{x} = \varepsilon F_{\#}(\varepsilon, t, x)$ has a T-periodic solution $\phi(t) = \xi + O(\varepsilon)$.

6. Analyze the forced Duffing's equation, $\ddot{x} + x = \varepsilon\{\delta x + \gamma x^3 + A\cos t\} = 0$, three different ways, and show that the seemingly different methods give the same intrinsic results. The parameter δ is called the detuning and is a measure of the difference between the natural frequency and the external forcing frequency. Remember that a one-degree-of-freedom autonomous system has a phase portait given by the level lines of the Hamiltonian.
   a. Write the system in action-angle coordinates, and compute the first term in the normal form, F_0^1, as was done for Duffing's equation. Analyze the truncated equation by drawing the level lines of the Hamiltonian. (See Section VI.II.B.)
   b. Write the system in complex coordinates and compute the first term in the normal form, F_0^1, as was done for Duffing's equation in Section B. Analyze the equation as in **a**.
   c. Make the "van der Pol" change of coordinates

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

   and then compute the first average of the equations via Problems 4 and 5. Analyze the equations. See McGehee and Meyer (1974).

7. Consider a Hamiltonian of two degrees of freedom of the form C.1, $x \in \mathbb{R}^4$. Let $H_0(x)$ be the Hamiltonian of two harmonic oscillators. Change to action-angle variables $(I_1, I_2, \phi_1, \phi_2)$ and let $H_0 = \omega_1 I_1 + \omega_2 I_2$. Use theorem C.1 to show that the terms in the normal form are of the form $aI_1^{p/2} I_2^{q/2} \cos(r\phi_1 + s\phi_2)$ or $bI_1^{p/2} I_2^{q/2} \sin(r\phi_1 + s\phi_2)$, a and b constants, if and only if $r\omega_1 + s\omega_2 = 0$, and the terms have the d'Alembert character. See Henrard (1970b).

8. Consider a Hamiltonian $H(x)$ with general solution $\phi(t, \xi)$. Observe that the ith component of ϕ is the Lie transform of x_i, i.e., $\phi_i(t, \xi) = \mathscr{L}_H(x_i)(\xi)$, where ε is replaced by t.
   a. Show that $\phi_i(t, \xi) = [x_i + \{x_1, H\}t + \{\{x_1, H\}, H\}t^2/2 + \cdots]_{x=\xi}$.
   b. Using Macsyma, Reduce, etc., write a simple function to compute the time 1 maps given in Equations (E.36) and (E.38). (Make sure that you compute the time series far enough to pick up all the quadratic and cubic terms in the initial conditions.)

9. Prove Theorem B.4, the uniqueness theorem. [Hint: Show that if the normal form is not unique then there are two different Hamiltonian H and K which are both in normal form and a generating function W carrying one into the other. Show that the terms in the series expansion for W must lie in the kernel $\mathscr{Q}_i$. Then show that this implies that $W \equiv 0$.]

CHAPTER VIII

Bifurcations of Periodic Orbits

This chapter and Chapter IX use the theory of normal forms developed in Chapter VI. They contain an introduction to generic bifurcation theory and its applications. Bifurcation theory has grown into a vast subject with a large literature; so, this chapter can only present the basics of the theory. The primary focus of this chapter is the study of periodic solutions—their existence and evolution. Periodic solutions abound in Hamiltonian systems. In fact, a famous Poincaré conjecture is that periodic solutions are dense in a generic Hamiltonian system, a point that was established in the C^1 case by Pugh and Robinson (1977).

A. Bifurcations of Periodic Solutions and Points

Recall that in Section V.E the study of periodic solutions of a Hamiltonian system was reduced to the study of one-parameter family of symplectic maps—the Poincaré map in an integral surface. The integral surface is in a level set of the Hamiltonian, and the parameter is the value of the Hamiltonian on that level set. If the Hamiltonian has n degrees of freedom, then the phase space is $2n$ dimensional, and the section in the integral surface has dimension $2n - 2$. This effects a reduction of dimension by 2.

A fixed point of the Poincaré map corresponds to a periodic solution of the Hamiltonian system. The questions answered in this section are: (1) When can a fixed point be continued? (2) What typically happens when you cannot continue a fixed point? (3) Are there other periodic points near a fixed point? However, to keep the notation simple, the discussion will be limited to symplectic maps of two dimensions which depend on one parameter. This corre-

sponds to a two-degree-of-freedom autonomous system or a one-degree-of-freedom periodic system. In two dimensions, a map is symplectic if and only if it is area preserving; so, henceforth that term will be used. A warning should be given: the proper generalization of the theory presented below would be to symplectic maps, not just volume-preserving maps.

Even the restriction to area-preserving maps is not enough for a complete classification because the number of types of bifurcations is manifold. Therefore, only the "generic case" will be considered in this section. The word "generic" can be given a precise mathematical meaning in the context of bifurcation theory, but here only the intuitive meaning will be given in order to avoid a long mathematical digression. Consider the set of all smooth area-preserving mappings depending on some parameter; then a subset of that set is *generic* if it has two properties: (1) it is open and (2) it is dense. A subset is *open* if a small smooth perturbation of a mapping in the subset is also in the subset. So the defining properties of elements of the subset are not sensitive to small perturbations, or the elements are "stable" under perturbations. A subset is *dense* if any element in the set can be approximated by an element of the subset. The set of area-preserving mappings satisfying the properties listed in the propositions can be shown to be generic [Meyer (1970)].

1. Elementary Fixed Points

Let $P: \mathbb{I} \times \mathbb{O} \to \mathbb{O}: (\mu, x) \to P(\mu, x)$ be a smooth function where $\mathbb{I} = (-\mu_0, \mu_0)$, $\mu_0 > 0$, is an interval in $\mathbb{R}$, and $\mathbb{O}$ is an open neighborhood of the origin in $\mathbb{R}^2$. For fixed $\mu \in \mathbb{I}$, let $P_\mu = P(\mu, \cdot): \mathbb{O} \to \mathbb{O}$ be area preserving; so, $\det(\partial P(\mu, x)/\partial x) \equiv 1$. Let the origin be a fixed point of P when $\mu = 0$, i.e., $P(0, 0) = 0$. The eigenvalues of $A = \partial P(0, 0)/\partial x$ are the *multipliers* of the fixed point. In two dimensions, the eigenvalues of the symplectic matrix, A, are (1) real reciprocals, or (2) on the unit circle, or (3) both equal to -1, or (4) both equal to $+1$. If the multipliers are different from $+1$, the fixed point is *elementary*.

Proposition 1. *An elementary fixed point can be continued. That is, if $x = 0$ is an elementary fixed point for P when $\mu = 0$, then there exists a $\mu_1 > 0$ and a smooth map $\xi: (-\mu_1, \mu_1) \to \mathbb{O}$ with $P(\xi(\mu), \mu) \equiv \xi(\mu)$. Moreover, the multipliers of the fixed point $\xi(\mu)$ vary continuously with μ; so, if $x = 0$ is elliptic (respectively hyperbolic) when $\mu = 0$, then so is $\xi(\mu)$ for small μ.*

PROOF. The implicit function theorem applies to $G(\mu, x) = P(\mu, x) - x = 0$ because $G(0, 0) = 0$, and $\partial G(0, 0)/\partial x = A - I$ is nonsingular; so, there is a $\xi(\mu)$ such that $G(\mu, \xi(\mu)) = P(\mu, \xi(\mu)) - \xi(\mu) = 0$. The multipliers of $\xi(\mu)$ are the eigenvalues of $\partial P(\mu, \xi(\mu))/\partial x$, and the eigenvalues of a matrix vary continuously (not always smoothly) with a parameter. ∎

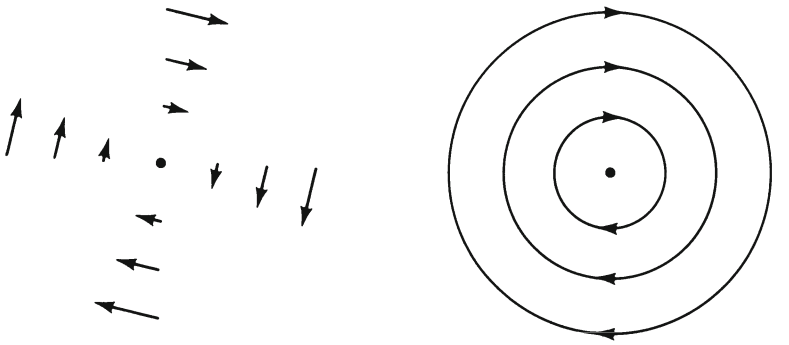

Figure A.1. Rendering of elliptic points.

In particular, an elliptic (respectively hyperbolic) fixed point can be continued to an elliptic (respectively hyperbolic) fixed point.

There will be several figures in this chapter. These figures will show the approximate placement of the fixed points and their type, elliptic or hyperbolic, as parameters are varied. That is all they are meant to convey. They will be drawn as if the diffeomorphism where the time one map of a differential equaton. Thus, for example, the drawing of an elliptic point will show concentric circles about the fixed point. Do *not* assume that the circles are invariant curves for the map. These curves suggest that the mapping approximately rotates the points. Figure A.1 shows two depictions of an elliptic fixed point. The one on the left shows that the points near the elliptic point move a discrete distance and is a more accurate depiction, whereas the figure on the right indicates invariant curves. The figure on the right is slightly misleading, but is less cluttered and therefore will be used in this chapter.

2. Extremal Fixed Points

Consider the case when the multipliers are equal to $+1$. In this case, the simple implicit function theorem argument fails and for a good reason. Many different things can happen depending on the nonlinear function; so, the simple conclusions of Proposition 1 may not hold in this case. As an extreme, consider the case when $A = I$ and $P(\mu, x) = x + \mu p(x)$, where $p(x)$ is an arbitrary function. The fixed points of $P(\mu, x)$ for $\mu \neq 0$ are the zeros of $p(x)$; since $p(x)$ is arbitrary, the fixed point set can be quite complicated—in fact, it can be any closed set in $\mathbb{R}^2$. In light of this potential complexity, only the typical or generic situation for a one-parameter family will be considered.

Definition. The origin is an *extremal fixed point* for P when $\mu = 0$, if there are symplectic coordinates (u, v) so that P: $(\mu, u, v) \to (u', v')$, where

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ \delta \end{pmatrix} + \begin{pmatrix} \cdots \\ \beta u^2 + \cdots \end{pmatrix} + \cdots, \tag{1}$$

and $\alpha = \pm 1$, $\beta \neq 0$ and $\delta \neq 0$. First, note that it is assumed that when $\mu = 0$ the linear mapping is already in Jordan normal and the matrix is not simple ($\alpha = \pm 1$). Second, since $\delta \neq 0$, the perturbation does not leave the origin as a fixed point. Last, note that one nonlinear term is nonzero. It is not necessary to put the full map into normal. However, if (1) is in the normal form as discussed in Section VII.E.5, then the assumption that $\beta \neq 0$ is the assumption that the first nonlinear term in the normal form appears with a nonzero coefficient.

Proposition 2. *Let $0 \in \mathbb{O} \subseteq \mathbb{R}^2$ be an extremal fixed point for P when $\mu = 0$. Then there is a smooth curve $\sigma: (-\tau_2, \tau_2) \to \mathbb{I} \times \mathbb{O}: \tau \to (\nu(\tau), \xi(\tau))$ of fixed points of P, $P((\nu(\tau), \xi(\tau)) = \xi(\tau)$, with $\tau = 0$ giving the extremal fixed point, $\tau(0) = (0, 0)$. The extremal point divides the curve of fixed points into two arcs. On one arc, the fixed points are all elliptic, and on the other, the fixed points are all hyperbolic. Moreover, the parameter μ achieves a nondegenerate maximum or minimum when at the extremal fixed point; so, there are two fixed points when μ has one sign and no fixed points when μ has the other. The proof contains precise information on the relationship between the signs and the nature of the fixed points.*

PROOF. The equations to be solved are

$$\begin{aligned} 0 &= u' - u = \alpha v + \mu\gamma + \cdots, \\ 0 &= v' - v = \mu\delta + \beta u^2 + \cdots, \end{aligned} \tag{2}$$

Since $\alpha \neq 0$ and $\delta \neq 0$, these equations can be solved for v and μ as a function of u. (Note the difference between this proof and the proof of Proposition 1—one of the variables solved for in this proof is the parameter.) The solution is of the form $\bar{v}(u) = O(u^2)$ and $\bar{\mu}(u) = (-\beta/\delta)u^2 + O(u^3)$; so, the map is $\sigma: \tau \to (\bar{\mu}(\tau), \tau, \bar{v}(\tau))$. The extreme point is obtained when $\tau = 0$. Note that if $\beta\delta > 0$, the $\bar{\mu}$ obtains a nondegenerate maximum when $\tau = 0$, and if $\beta\delta < 0$, then $\bar{\mu}$ obtains a nondegenerate minimum when $\tau = 0$.

The Jacobian of the map along this solution is $\left(\begin{smallmatrix} 1 & \alpha \\ 2\beta\tau & 1 \end{smallmatrix}\right) + \cdots$, and so the multipliers are $1 \pm (2\alpha\beta\tau)^{1/2} + \cdots$. Hence, when $\alpha\beta > 0$, the fixed point is elliptic for $\tau < 0$ and hyperbolic for $\tau > 0$ and vice versa when $\alpha\beta < 0$. ∎

Figure A.2 shows the curve σ in $\mathbb{I} \times \mathbb{O}$. $\mathbb{I}$ is the horizontal axis and $\mathbb{O}$ is depicted as a one-dimensional space, the vertical axis. In the case shown the μ achieves a nondegenerate maximum on the curve at the origin. Consider the case depicted in Figure A.2. For μ negative there are two fixed points in $\mathbb{O}$, one elliptic and one hyperblic; see Figure A.3(a). As μ approaches zero through negative values, these fixed points come together until they collide and become a degenerate fixed point when $\mu = 0$; see Figure A.3(b). For positive μ there are no fixed points in $\mathbb{O}$.

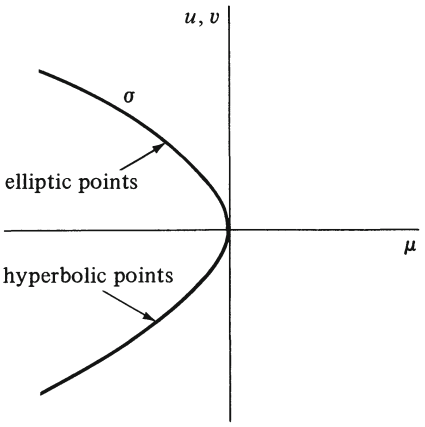

Figure A.2. The curve σ.

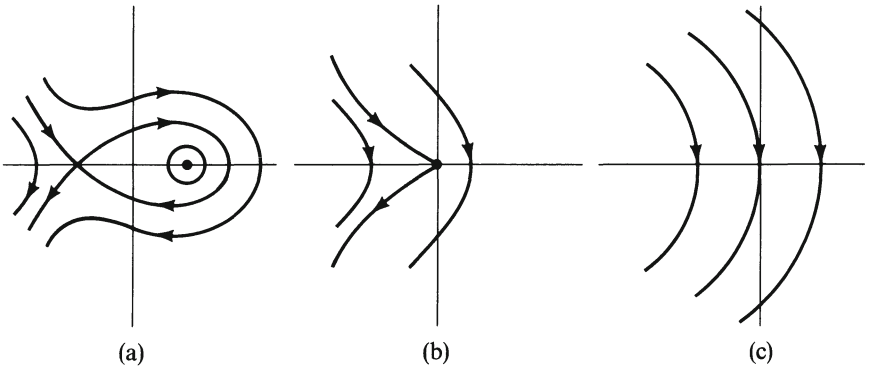

Figure A.3. Extremal fixed point. (a) $\mu < 0$; (b) $\mu = 0$; (c) $\mu > 0$.

3. Period Doubling

The solutions given by the implicit function theorem are locally unique; so, there is a neighborhood of an elementary or an extremal fixed point which contains no other fixed points. But, there may be periodic points of higher period near one of these fixed points. There are no periodic points near a hyperbolic or extremal fixed point (see the Problems Sections), but there may be one near an elliptic fixed point. Let $x = 0$ be an elementary fixed point for P when $\mu = 0$; so, by Proposition 1, there is a smooth $\xi(\mu)$ of fixed points. This fixed point can be shifted to the origin by considering $P'(u, \mu) = P(n + \xi(\mu), \mu) - \xi(\mu)$. Assume that this shift has been done, and revert to the original notation, i.e., assume that $P(0, \mu) \equiv 0$.

Let $P_\mu(x) = P(x, \mu) = Ax + \cdots$, then $P_\mu^k(x) = P_\mu \circ P_\mu \circ \cdots \circ P_\mu(x) = A^k x + \cdots$. A k-periodic point satisfies the equation $P_\mu^k(x) = A^k x + \cdots = x$ which has a unique solution, $x = 0$, unless $A^k - I$ is singular, or one of the

eigenvalues of A is a kth root of unity. Thus, k-periodic points may exist near a fixed point with a multiplier which is a kth root of unity. In fact, generically they do bifurcate from fixed points whose multipliers are kth roots of unity. In this subsection, the case when $k = 2$ is considered, i.e., when the multipliers are -1.

The map $\mathscr{C}$: $\mu \to \partial P(0, \mu)/\partial x$ is a curve in $Sp(2, \mathbb{R})$, the set of all 2×2 real matrices with determinant equal to $+1$. $Sp(2, \mathbb{R})$ is a three-dimensional space because there is one algebraic identity among the four entries of the matrix. Let $\mathscr{M}$ be a subspace of $Sp(2, \mathbb{R})$. If $\mathscr{M}$ is a discrete set of points or a curve, then a small perturbation of $\mathscr{C}$ would miss $\mathscr{M}$, and if $\mathscr{C}$ already misses $\mathscr{M}$, then a small perturbation of $\mathscr{C}$ would still miss $\mathscr{M}$. Thus, one open and dense condition (a generic condition) is for $\mathscr{C}$ to miss a discrete set or a curve in $Sp(2, \mathbb{R})$. If $\mathscr{M}$ is a surface in the three-dimensional space $Sp(2, \mathbb{R})$, then the curve $\mathscr{C}$ would in general hit $\mathscr{M}$ in a discrete set of points and cross the surface with nonzero velocity. This is generic when $\mathscr{M}$ is a surface; see Figure A.4.

The set of matrices, $\mathscr{M}_2 = \{A \in Sp(2, \mathbb{R})$: trace $A = -2\}$, is the set of matrices in $Sp(2, \mathbb{R})$ with eigenvalue equal to -1. It is a surface because the matrices satisfy the additional algebraic identity, trace $A = -2$. The set $\{-I\} \in \mathscr{M}_2$ is a discrete point; thus, generically, the curve $\mathscr{C}$ intersects $\mathscr{M}_2 \setminus \{-I\}$ in a discrete set of points, and at these points, $d(\text{trace } \mathscr{C}(\mu))/d\mu \neq 0$. Thus, along a curve of elementary fixed points, there are isolated points where the multipliers are -1, and the Jacobian is not simple. At these points, the map can be put into the normal form described in Section VII.E.5. It is also generic for the first term in the normal form to appear with a nonzero coefficient. This informal discussion leads to the following definition.

Definition. The origin is a *transitional periodic point* for P at $\mu = 0$ if there are symplectic coordinates (u, v) so that P: $(u, v) \to (u', v')$, where

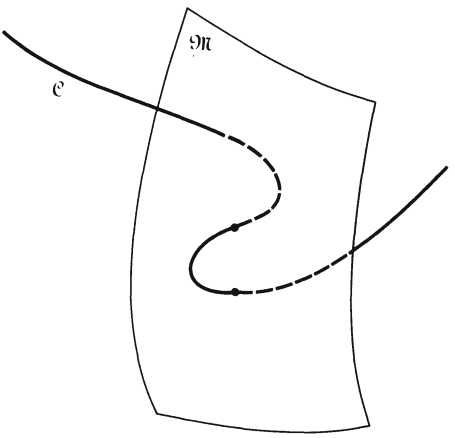

Figure A.4. Intersections in $Sp(2, \mathbb{R})$.

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} -1 & \alpha \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \mu \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} \cdots \\ \beta u^3 + \cdots \end{pmatrix} + \cdots, \tag{3}$$

and $\alpha = \pm 1, c \neq 0, \beta \neq 0$.

There are three conditions in this definition. First, when $\mu = 0$, the multipliers are -1, and the Jacobian matrix is not diagonalizable, $\alpha \neq 0$. Second, since

$$\det \left\{ \begin{pmatrix} -1 & \alpha \\ 0 & -1 \end{pmatrix} + \mu \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \cdots \right\} = 1 - \mu(a + d + \alpha c) + \cdots = +1,$$

$c = -(a + b)/\alpha$, and so the condition $c \neq 0$ implies that the derivative of the trace of the Jacobian is nonzero. Third, $\beta \neq 0$ is the condition that the first term in the normal form when $\mu = 0$ is nonzero. It is not necessary that the map be put into normal form completely; simply eliminate all the quadratic terms, and then assume that $\beta \neq 0$.

Proposition 3. *Let the origin be a transitional fixed point for P when $\mu = 0$ and μ small. If $\alpha c < 0$, then the origin is a hyperbolic fixed point when $\mu > 0$ and the origin is an elliptic fixed point when $\mu < 0$ (vise versa when $\alpha c < 0$). If $\beta c > 0$ (respectively $\beta c < 0$), then there exists a periodic orbit of period 2 for P_μ when $\mu < 0$ (respectively $\mu > 0$), and there does not exist a periodic orbit for $\mu \geq 0$ (respectively $\mu \leq 0$). As μ tends to zero from the appropriate side, the period 2 orbit tends to the transition fixed point. For fixed μ, the stability type of the fixed point and the period 2 orbit are opposite. That is, if fixed μ the origin is elliptic, then the periodic point is hyperbolic and vice versa. See Figure* A.5.

Remark. The fixed point is called a transition point because the stability type of the fixed point changes from hyperbolic to elliptic, or vice versa. At the transition point, a new period 2 point appears on one side of $\mu = 0$; this is called *period doubling* in the literature. One says that the period 2 point *bifurcates from the transition point.*

PROOF. The first part of the proposition follows by the remark preceding the statement. This remark shows that the trace of the Jacobian at the origin is $-2 - \mu\alpha c + \cdots$. Compute that the second iterate of the map is $(u, v) \to (u'', v'')$, where

$$\begin{aligned} u'' &= u - 2\alpha v + \cdots, \\ v'' &= v - 2\mu c u - 2\mu d v - 2\beta u^3 + \cdots. \end{aligned} \tag{4}$$

Since $\alpha \neq 0$, the equation $u'' - u = -2\alpha v + \cdots = 0$ can be solved for v as a function of μ and u. Call this solution $\bar{v}(u, \mu)$. The lowest-order terms in $\bar{v}$ are of the form $k\mu u^2$ and $k' u^3$, where k and k' are constants. Substitute this solution into the equation $v'' - v$ to get

$$v'' - v = -2\mu c u - 2\beta u^3 + \cdots. \tag{5}$$

Since the origin is always a fixed point, u is a common factor, and so, the equation to solve is

$$(v'' - v)/u = -2\mu c - \beta u^2 + \cdots . \tag{6}$$

Since $c \neq 0$, Equation (6) can be solved for μ as a function of u; call this solution $\bar{\mu}(u) = -(\beta/2c)u^2 + \cdots$. If $\beta c > 0$, then there are two real solutions, $u_\pm(\mu) = \pm\sqrt{-2c\mu/\beta + \cdots}$ for $\mu < 0$, and none for $\mu \geq 0$, and vice versa when $\beta c < 0$. Thus, $(u_\pm(\mu), \bar{v}(u_\pm(\mu), \mu))$ are two points of P_μ^2, but since they are not the origin, they are not fixed points of P_μ. Therefore, they are periodic points of period 2. The Jacobian is

$$\frac{\partial(u'', v'')}{\partial(u, v)} = \begin{pmatrix} 1 & -2\alpha \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \cdots & \cdots \\ -2c\mu - 6\beta u^2 & \cdots \end{pmatrix} + \cdots . \tag{7}$$

Since $u^2 = -2c\mu/\beta + \cdots$ along these solutions, the multipliers are $1 \pm \sqrt{-20\alpha c\mu + \cdots}$ and so are hyperbolic if $\alpha c\mu < 0$ and elliptic when $\alpha c\mu > 0$. ■

There are two basic cases. Case A: the periodic point is elliptic; Case B: the periodic point is hyperbolic. These are depicted in Figure A.5. In the figure it is assumed that $\alpha c > 0$ and Case A is when $\beta c > 0$ and Case B is when $\beta c < 0$.

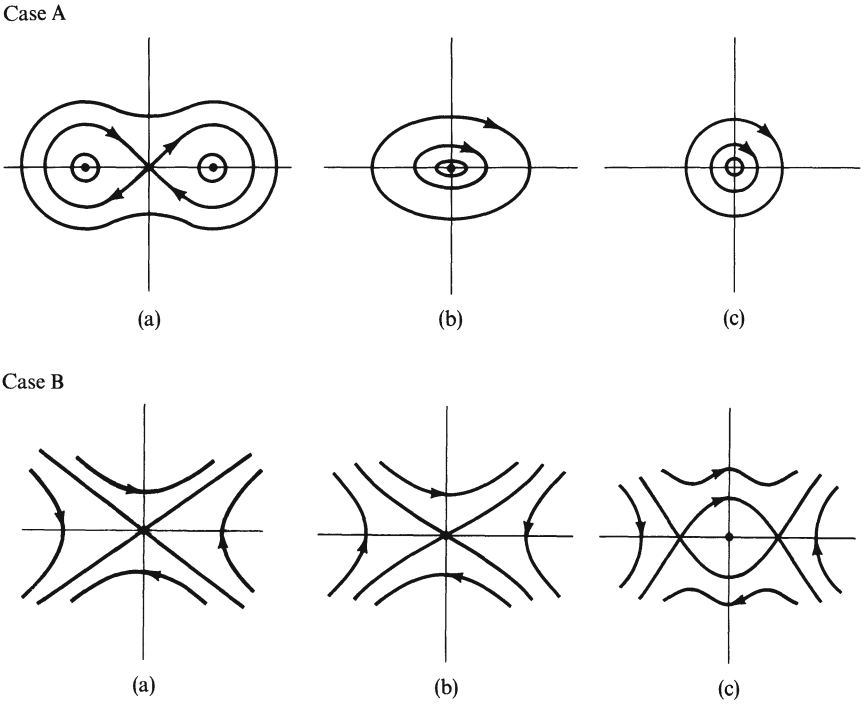

Figure A.5. Transitional point. Case A: (a) $\mu < 0$, (b) $\mu = 0$, (c) $\mu > 0$. Case B: (a) $\mu < 0$, (b)$\mu = 0$, (c) $\mu > 0$.

4. k-Bifurcation Points

From the discussion of the last section, periodic points are likely near a fixed point that has multipliers which are kth roots of unity. In the last subsection the generic case of a fixed point with multiplier -1, a square root of unity, was discussed, and in this section, the remaining cases will be discussed. Recall the normal forms given in Section VII.E.4.

Definition. The origin is a *k-bifurcation point*, $k \geq 3$, for P when $\mu = 0$, if there are symplectic action-angle coordinates (I, ϕ) so that $P(I, \phi) \to (I', \phi')$, where

$$\begin{aligned} I' &= I - 2\gamma I^{k/2} \sin(k\phi) + \cdots, \\ \phi' &= \phi + (2\pi h/k) + \alpha\mu + \beta I + \cdots + \gamma I^{(k-2)/2} \cos(k\phi) + \cdots, \end{aligned} \tag{8}$$

and $\alpha \neq 0$, $\gamma \neq 0$ when $k = 3$; $\alpha \neq 0$, $\gamma \neq 0$, $\beta \pm \gamma \neq 0$ when $k = 4$; $\alpha \neq 0$, $\beta \neq 0, \gamma \neq 0$ when $k \geq 5$.

The linearized map is $I' = I$, $\phi' = \phi + (2\pi h/k) + \alpha\mu$. So when $\mu = 0$, the multipliers are $\exp(\pm 2\pi hi/k)$, a kth root of unity. The assumption $\alpha \neq 0$ is the assumption that the multipliers pass through the kth root of unity with nonzero velocity. When $k \geq 5$, the terms with ϕ dependence are higher order, and the map when $\mu = 0$ is of the form $I' = I + \cdots$, $\phi' = \phi + (2\pi h/k) + \beta I + \cdots$. The assumption that $\beta \neq 0$ is the *twist* assumption, and a map satisfying this assumption is called a *twist map*. Twist maps will be discussed in the next section, and in Chapters IX and X. The assumption that $\gamma \neq 0$ is the assumption that the first angle-dependent term in the normal form appears with a nonzero coefficient. This term is referred to as the *resonance term*, and it is very important to the bifurcation analysis given below. The resonance term is of lower order than the twist term when $k = 3$ and vice versa when $k \geq 5$. When $k = 4$, they are both of the same order. Therefore, the case $k = 3$ and $k = 4$ is special and must be treated separately.

Proposition 4. *Let the origin be a 3-bifurcation point for P when $\mu = 0$ and μ is small. Then there is a hyperbolic periodic orbit of period 3 which exists for both positive and negative values of μ and the periodic point tends to the 3-bifurcation as $\mu \to 0$ from either side.* (See Figure A.6.)

PROOF. Compute the third iterate of the map P_μ as P_μ^3: $(I, \phi) \to (I^3, \phi^3)$, where

$$\begin{aligned} I^3 &= I - 2\gamma I^{3/2} \sin(3\phi) + \cdots, \\ \phi^3 &= \phi + 2\pi h + 3\alpha\mu + 3\gamma I^{1/2} \cos(3\phi) + \cdots. \end{aligned} \tag{9}$$

The origin is always a fixed point; so, I is a common factor in the formula for I^3. Since $\gamma \neq 0$, the equation $(I^3 - I)/(2\gamma I^{3/2}) = \sin(3\phi) + \cdots$ can be solved for six functions $\phi_j(I, \mu) = j\pi/3 + \cdots$, $j = 0, 1, \ldots, 5$. For even j, $\cos 3\phi_j = +1 + \cdots$, and for odd j, $\cos 3\phi_j = -1 + \cdots$. Substituting these solutions into the ϕ equation gives $(\phi^3 - \phi_j - 2h\pi)/3 = \alpha\mu \pm \gamma I^{1/2} + \cdots$. The equations

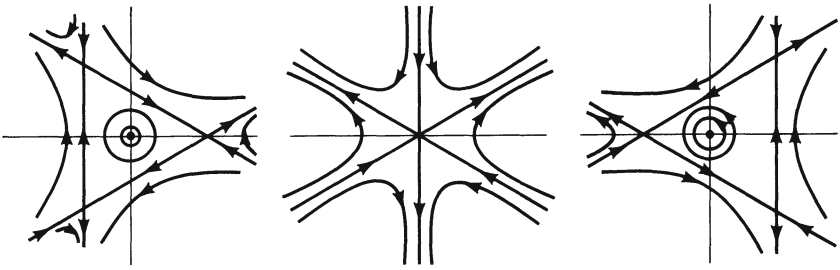

Figure A.6. 3-Bifurcation point. (a) $\mu < 0$, (b) $\mu = 0$, (c) $\mu > 0$.

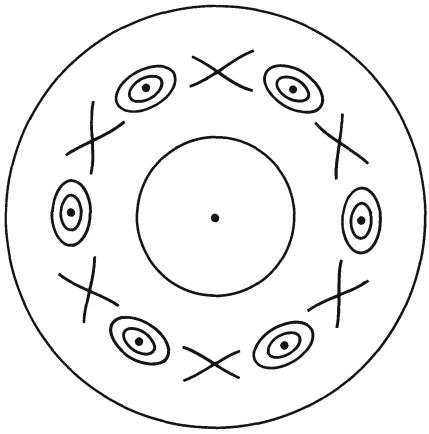

Figure A.7. A 6-bifurcation point.

with a plus sign have a positive solution for I when $\alpha\gamma\mu$ is negative, and the equations with the minus sign have a positive solution for I when $\alpha\gamma\mu$ is positive. The solutions are of the form $I_j^{1/2} = \mp\alpha\mu/\gamma + \cdots$. Compute the Jacobian along these solutions to be

$$\frac{\partial(I^3, \phi^3)}{\partial(I, \phi)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \mp 6\gamma I_j^{3/2} \\ (\pm 3\gamma/2) I_j^{1/2} & 0 \end{pmatrix} + \cdots, \tag{10}$$

and so the multipliers are $1 \pm 3I_j|v|$, and the periodic points are all hyperbolic. ∎

Proposition 5. *Let the origin be a k-bifurcation point, $k \geq 5$, for P when $\mu = 0$. Then when $\alpha\beta < 0$ (respectively $\alpha\beta > 0$) there exist an elliptic and also a hyperbolic periodic orbit of period k for $\mu > 0$ (respectively $\mu < 0$) and no periodic orbit of period k when $\mu < 0$ (respectively $\mu > 0$). As $\mu \to 0$ from the appropriate side, both the elliptic and hyperbolic orbits tend to the k-bifurcation point. (See Figure A.7.)*

Remark. These periodic orbits are said to *bifurcate from the fixed point when* $\mu = 0$. Each orbit consists of k points, and there are exactly two periodic orbits.

PROOF. Compute the kth iterate of the map P_μ as P_μ^k: $(I, \phi) \to (I^k, \phi^k)$, where

$$\begin{aligned} I^k &= I - 2\gamma I^{k/2} \sin(k\phi) + \cdots, \\ \phi^k &= \phi + 2h\pi + \alpha k\mu + \beta k I + \cdots. \end{aligned} \tag{11}$$

Since the origin is a fixed point for μ, the first equation is divisible by $I^{k/2}$. By the implicit function theorem, there are $2k$ solutions of $(I^k - I)/(-2\gamma I^{k/2}) = \sin(k\phi) + \cdots = 0$; call them $\phi_j(I, \mu) = j\pi/k + \cdots$. Substitute these solutions into the equation $(\phi^k - \phi - 2h\pi)/k = \alpha\mu + \beta I + \cdots = 0$. For each of the ϕ_j, this second equation has a solution $I_j = -\alpha\mu/\beta + \cdots$ which gives a positive I provided $\alpha\beta\mu < 0$.

The Jacobian at these solutions is

$$\frac{\partial(I^k, \phi^k)}{\partial(I, \phi)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -2k\gamma I_j^{k/2} \cos(k\phi_j) \\ k\beta & 0 \end{pmatrix} + \cdots, \tag{12}$$

and so the multipliers are $1 \pm \sqrt{2k\gamma\beta I_j^{k/2}} + \cdots$, where the plus sign is taken for even j because $\cos(k\phi_j) = +1 + \cdots$ for even j, and the minus sign is taken for odd j. ∎

The case when $k = 4$ is sometimes like the case when $k = 3$ and sometimes like the case when $k \geq 5$ depending on the relative size of the twist term and the resonance term.

Proposition 6. *Let the origin be a* 4*-bifurcation point for* P *when* $\mu = 0$. *Then*:

Case A. If $\beta \pm \gamma$ *have different signs, then there is a hyperbolic periodic orbit of period* 4 *which exists for both positive and negative* μ *and tends to the* 3*-bifurcation as* $\mu \to 0$ *from either side.*

Case B. If $\beta \pm \gamma$ *have the same sign, then when* $\alpha(\beta \pm \gamma) < 0$ [*respectively* $\alpha(\beta \pm \gamma) > 0$], *there exists an elliptic and a hyperbolic periodic orbit of period* 4 *for* $\mu > 0$ (*respectively* $\mu < 0$) *and no periodic orbit of period* 4 *when* $\mu < 0$ (*respectively* $\mu > 0$). *As* $\mu \to 0$ *from the appropriate side, both the elliptic and hyperbolic orbits tend to the* 4*-bifurcation point.*

PROOF. Compute the 4th iterate of the map P_μ as P_μ^4: $(I, \phi) \to (I^4, \phi^4)$, where

$$\begin{aligned} I^4 &= I - 2\gamma I^2 \sin(4\phi) + \cdots, \\ \phi^4 &= \phi + 2h\pi + 4\alpha\mu + 4\{\beta + \gamma \cos(4\phi)\} I + \cdots. \end{aligned} \tag{13}$$

Since the origin is a fixed point for all μ, the first equation is divisible by I^2. By the implicit function theorem there are eight solutions of $(I^4 - I)/(-2\gamma I^2) = \sin(4\phi) + \cdots = 0$; call them $\phi_j(I, \mu) = j\pi/4 + \cdots$. Substitute these solutions

into the equation

$$\begin{aligned}(\phi^4 - \phi - 2h\pi)/4 &= \alpha\mu + \{\beta + \gamma\cos(4\phi_j)\}I + \cdots \\ &= \alpha\mu + \{\beta \pm \gamma\}I + \cdots = 0.\end{aligned} \tag{14}$$

For each of the ϕ_j, this equation has a solution $I_j = -\alpha\mu/\{\beta \pm \gamma\} + \cdots$ which gives a positive I provided $\alpha\{\beta \pm \gamma\}\mu < 0$. So if $\beta \pm \gamma$ have different signs, then one group of four solutions exists for positive μ and the order group for negative μ; this is Case A and is similar to Proposition 4. If $\beta \pm \gamma$ are of the same sign, all eight solutions exist for μ on one side on 0; this is Case B and is similar to Proposition 5. The calculation of the multipliers is similar to the calculations given above. ∎

B. Duffing Revisited

This section develops some new ideas in order to analyze two types of bifurcations which occur in Duffing's equation. The first is an extremal, and the second is a k-bifurcation.

1. Duffing at 1–1 Resonance

Here the classical Duffing equation is considered, even though it has been discussed in many texts. Most of the classical treatments miss the fact that there is an extremal periodic solution as defined in Section A.2, and therefore their treatment is incomplete. Consider the classical Duffing equation

$$\ddot{x} + \omega_n^2 x + \gamma x^3 = A\cos\omega_e t, \tag{1}$$

or

$$\begin{aligned}\dot{x} &= \omega_n y = \frac{\partial H}{\partial y}, \\ \dot{y} &= -\omega_n x - \left(\frac{\gamma}{\omega_n}\right)x^3 + \left(\frac{A}{\omega_n}\right)\cos\omega_e t = -\frac{\partial H}{\partial x},\end{aligned} \tag{2}$$

where

$$H = (\omega_n/2)(y^2 + x^2) + (\gamma/\omega_n)x^4/4 - (A/\omega_n)x\cos\omega_e t. \tag{3}$$

When the nonlinearity is absent, $\gamma = 0$, and there is no external forcing, $A = 0$. This is simply the harmonic oscillator with a frequency ω_n, the *natural frequency*. The general solution is $\phi_g = \alpha\cos\omega_n t + \beta\sin\omega_n t$.

If the nonlinearity is absent, $\gamma = 0$, the external force is present, $A \neq 0$, and the two frequencies are unequal, $\omega_e \neq \omega_n$, then the equation has a particular solution $\phi_p = B\cos\omega_e t$, $B = A/(\omega_e^2 - \omega_n^2)$. In this case, the particular solution

is the unique solution which is periodic with the same frequency as the *external frequency*, ω_e, and period $T = 2\pi/\omega_e$. The variational equation for this solution is obtained by setting $A = \gamma = 0$ in (2), and the period map is computed to be $(x, y) \to (x', y')$, where

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \omega_n T & \sin \omega_n T \\ -\sin \omega_n T & \cos \omega_n T \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{4}$$

Thus, the multipliers of this solution are $\exp(\pm 2\pi i \omega_n/\omega_e)$, which are not equal to $+1$ provided $\omega_n/\omega_e \neq 0, \pm 1, \pm 2, \pm 3, \ldots$. In this case, the particular solution is elliptic, hence elementary, and so it can be continued into the nonlinear problem for small $\gamma \neq 0$. In summary, if $\omega_n/\omega_e \neq 0, \pm 1, \pm 2, \pm 3, \ldots$, then for small forcing and small nonlinearity, there is a small periodic solution of (1) with the same period as the external forcing. In the classical literature this solution is sometimes referred to as the *harmonic*.

The question of interest, then, is what happens when $\omega_n/\omega_e = 0, \pm 1, \pm 2, \pm 3, \ldots$. To this end consider the case when ω_n/ω_e is near $+1$ by setting $\omega_n^2 = 1 - \varepsilon\delta$, $\omega_e = 1$, where ε is a small parameter. The interesting thing happens not just when the ratio of the frequencies is 1, but also when the ratio is near 1; so, the parameter δ is introduced. It is called the *detuning*. Assume that the nonlinearity and forcing are small by using the following replacements: $\gamma \to -\varepsilon\gamma$, $A \to \varepsilon A$. That is, consider the equation

$$\ddot{x} + x = \varepsilon\{\delta x + \gamma x^3 + A \cos t\}, \tag{5}$$

or

$$\begin{aligned} \dot{x} &= y = \frac{\partial H}{\partial y}, \\ \dot{y} &= -x + \varepsilon\{\delta x + \gamma x^3 + A \cos t\} = -\frac{\partial H}{\partial x}, \end{aligned} \tag{6}$$

where

$$H = (1/2)(x^2 + y^2) - \varepsilon\{\delta x^2/2 + \gamma x^4/4 + Ax \cos t\}. \tag{7}$$

Change to action-angle variables by setting $x = \sqrt{2I} \cos \phi$, $y = \sqrt{2I} \sin \phi$, so that the Hamiltonian becomes

$$H = I - \varepsilon\{\delta I \cos^2 \phi + \gamma I^2 \cos^4 \phi + A\sqrt{2I} \cos \phi \cos t\}. \tag{8}$$

By Theorem VII.E.2, the normal form for (8) depends on I and $(\phi + t)$. To find the first term in the normal form, substitute the identities $\cos^2 \phi = (1 + \cos 2\phi)/2$, $\cos^4 \phi = (3 + 4 \cos 2\phi + \cos 4\phi)/8$, $\cos \phi \cos t = (\cos(\phi + t) + \cos(\phi - t))/2$ into (8), and keep only the terms in I and $(\phi + t)$ obtain the normal form:

$$H = I - \varepsilon\{(\delta/2)I + (3\gamma/8)I^2 + A2^{-1/2}I^{1/2} \cos(\phi + t)\} + \cdots, \tag{9}$$

Integrate the normalized equations from 0 to 2π to get the period map to be

$(I, \phi) \to (I', \phi')$, where

$$\begin{aligned} I' &= I + \varepsilon\{\pi 2^{1/2} A I^{1/2} \sin \phi\} + \cdots, \\ \phi' &= \phi - 2\pi + \varepsilon\pi\{\delta + gI + aI^{-1/2} \cos \phi\} + \cdots, \end{aligned} \tag{10}$$

where $g = 3\gamma/2$ and $a = (2A)^{-1/2}$. Solving the equation $(I' - I)/\varepsilon = 0$ using the implicit function theorem gives two solutions, $\phi_+(\varepsilon, \phi) = 0 + O(\varepsilon)$ and $\phi_-(\varepsilon, \phi) = \pi + O(\varepsilon)$, with $\sin \phi_\pm = 0 + O(\varepsilon)$ and $\cos \phi_\pm = \pm 1 + O(\varepsilon)$. Substituting these solutions into the equation $(\phi' - \phi)/2\pi\varepsilon = 0$ gives $\delta + gI \pm aI^{-1/2} + O(\varepsilon) = 0$. Solve this equation for δ by the implicit function theorem to get

$$\delta = -gI \mp aI^{-1/2} + \cdots; \tag{11}$$

see Figure B.1, where g and a are taken to be positive. Thus, if ε is small, for each (I, δ) satisfying (11) there are two 2π periodic solutions of (9). The graph of (11) when $\varepsilon = 0$ has a maximum when $ag > 0$ and a minimum when $ag < 0$ at $|a/g|^{2/3}$. In either case, the second derivative is nonzero at extrema, and so (11) has a maximum/minimum at a point $I' = |a/g|^{2/3} + O(\varepsilon)$, $\delta' = (3/2)(2a^2 g)^{1/3} + \cdots$ even when ε is nonzero and small. Thus, the qualitative features of the graph of (11) are the same as when $\varepsilon = 0$; see Figure B.1.

Consider the case when a and g are positive as pictured in Figure B.1. The free parameter is the detuning, δ. When $\delta < \delta'$ there are three values of I that satisfy Equation (11) and hence six periodic solutions. As δ approaches δ' from below, two of these I values approach each other, and when $\delta = \delta'$ they collide. For $\delta > \delta'$, there is only one I solution. This reminds one of an extremal bifurcation where an elliptic and a hyperbolic point come together.

In order to calculate the multipliers, calculate the Jacobian by

$$\frac{\partial(I', \phi')}{\partial(I, \phi)} = \begin{pmatrix} 1 & \pm\varepsilon(2I)^{-1/2} A \\ -\varepsilon\pi d\delta/dI & 1 \end{pmatrix} + \cdots. \tag{12}$$

Note the term in the lower left, $\partial\phi'/\partial I = -\varepsilon\pi d\delta/dI$, is exact. The characteristic equation of (12) is $(\lambda - 1 + \cdots)^2 \pm (\varepsilon^2\pi(2A)^{-1/2} + \cdots)\, d\delta/dI$. The two solutions near the extrema have different signs for $d\delta/dI$; so, one is elliptic and one is hyperbolic. Thus, the extrema (11) correspond to an extremal periodic point.

2. k-Bifurcations in Duffing's Equation

In applications it is difficult to verify the hypothesis $\gamma \neq 0$ for a k-bifurcation point when k is large. This is difficult even with an algebraic processor because the map must be put into normal form to very high order. In applications the period map is typically known only approximately, and it is not in full normal form. In several examples it is easy to compute that $\alpha \neq 0$ and $\beta \neq 0$. With this information alone and an ingenious idea of Birkhoff, the

existence of a bifurcation can be detected, but the uniquencess of the periodic orbits cannot. Use the notation of Section A.4 in this section.

Definition. The origin is a *weak k-bifurcation point*, $k \geq 5$, for P when $\mu = 0$, if there are symplectic action-angle coordinates (I, ϕ) so that P is as in Equation (A.8) with $\alpha \neq 0$ and $\beta \neq 0$.

Proposition 1. *Let the origin be a weak k-bifurcation point for P, $k \geq 5$, when $\mu = 0$. Then when $\alpha\beta < 0$ (respectively $\alpha\beta > 0$) P, has ℓ, $\infty \geq \ell \geq 2$, periodic orbits for each μ, $\mu > 0$ (respectively $\mu < 0$). As $\mu \to 0$ from the appropriate side, all the periodic orbits tend to the origin (the weak k-bifurcation point).*

Remark. This is simply an existence theorem for each fixed μ. In particular, the number of periodic orbits ℓ may depend on μ, and the orbits may not vary continuously in μ except at $\mu = 0$.

PROOF. As in the proof of Proposition A.5, compute P_μ^k to be as given in Equations (A.11). Since $\alpha \neq 0$ and $\beta \neq 0$, the equation $(\phi^k - \phi - 2h\pi)/k = \alpha\mu + \beta I + \cdots = 0$ can be solved for I to give $I^*(\phi, \mu) = -\alpha\mu/\beta + \cdots$, Let Γ_μ denote the closed curve $\{(I, \phi): I = I^*(\phi, \mu)\}$, the circle of zero rotation. Since P_μ^k is area preserving, the curve Γ_μ and its image $P_\mu^k(\Gamma_\mu)$ must intersect, i.e., $\Gamma_\mu \cap P_\mu^k(\Gamma_\mu) \neq \varnothing$. For small μ, both Γ_μ and $P_\mu^k(\Gamma_\mu)$ are smooth curves that meet a ray from the origin in only one point. Let $x \in \Gamma_\mu \cap P_\mu^k(\Gamma_\mu)$. Since $x \in \Gamma_\mu$, its angular coordinate does not change under P_μ^k, and since $x \in \Gamma_\mu \cap P_\mu^k(\Gamma_\mu)$ its radial coordinate does not change under P_μ^k; so, x is a fixed point of P_μ^k. Thus, there is at least one periodic orbit. Using the same argument as found in Birkhoff (1926) one can show that if there are a finite number of points in the intersection then one must have index $+1$, and one must have index -1; so, there are at least two periodic orbits. ∎

By a similar argument Birkhoff proved the following.

Proposition 2. *Near a general elliptic point there are periodic points of arbitrary high period.*

In Birkhoff's theorem there is no parameter, so, he had to make careful estimates to show that the curve of zero rotation Γ_0, existed for large k. See Birkhoff (1927) for the complete proof.

Consider the forced Duffing's equation (1) again; only this time, assume that only the forcing term is small by substituting $\gamma \to \varepsilon\gamma$. Normalize the time by setting $\omega_n = 1$ and $\omega_e = \omega$. Thus, consider the Hamiltonian

$$H = (1/2)(y^2 + x^2) + \gamma x^4/4 - \varepsilon A x \cos \omega t. \tag{13}$$

By the argument at the beginning of the previous subsection Duffing's equation has a small, order ε, $2\pi/\omega$-periodic solution, the harmonic, provided

$1/\omega \neq 0, \pm 1, \pm 2, \pm 3, \ldots$; so, assume that ω is away from these values. When $\varepsilon = 0$ the harmonic is the constant zero function, and its multipliers are $\exp(\pm 2\pi/\omega)$. Fix $k \geq 5$. By the implicit function theorem, there is a smooth function $\omega_{h/k}(\varepsilon)$ such that the multipliers of the harmonic are $\exp(\pm 2\pi hi/k)$ and $\omega_{h/k}(\varepsilon) = k/h + \cdots$ for small ε. Define a new parameter μ by $\mu = \omega - \omega_{h/k}(\varepsilon)$, the detuning parameter. When $\mu = 0$, the harmonic has multipliers which are kth roots of unity. When $\mu \neq 0$ but μ and ε are small, the multipliers of the harmonic are $\exp(\pm i(2\pi h/k + \alpha(\varepsilon)\mu + \cdots))$ where $\alpha(0) = -2\pi(h/k)^2 \neq 0$.

The period map about the harmonic can be put into normal form through the twist term, because the low resonance cases have been excluded. When $\varepsilon = 0$ and $\mu = 0$, the computations in Section VII.B.1 show that in new action-angle variables, one has

$$H = I + (3\gamma/8)I^2 + O(\varepsilon). \tag{14}$$

By integrating the equations from $t = 0$ to $t = 2\eta k/h$, the period map is

$$\begin{aligned} I' &= I + O(\varepsilon), \\ \phi' &= \phi - 2\pi h/k - (3\pi k\gamma/2h)I + O(\varepsilon). \end{aligned} \tag{15}$$

So when $\varepsilon = 0$ and $\mu = 0$, the twist coefficient is $-(3\pi k\gamma/2h) \neq 0$, and by continuity, it is nonzero for small ε. Therefore, Proposition 1 applies. For each $k \geq 5$ and small ε, Duffing's equation with Hamiltonian (13) has a small 2π-periodic solution with multipliers $\exp(\pm 2\pi h/k)$ when $\omega = \omega_{h/k}(\varepsilon)$. At least two periodic solutions of period $2k\pi/\omega$ bifurcate from the harmonic as ω varies from $\omega_{h/k}(\varepsilon)$. These periodic solutions occur for $\omega > \omega_{h/k}(\varepsilon)$ when $\gamma > 0$, and they occur for $\omega < \omega_{h/k}(\varepsilon)$ when $\gamma < o$.

These solutions are called *subharmonics* in the classical literature.

C. Schmidt's Bridges

In Section VI.C, the circular orbits of the Kepler problem were continued into the restricted problem to give two families of periodic solutions for small values of the mass ratio parameter μ. These families are known as the direct and retrograde orbits, depending on whether they rotate in the same or opposite direction as the primaries in the fixed coordinate system. In Section VI.G, some of the elliptic periodic solutions of the 2-body problem were continued into the restricted problem as symmetric periodic orbits.

Schmidt (1972) showed that these elliptic periodic solutions lie on families of symmetric periodic solutions which connect the direct and retrograde orbits. That is, for small μ, there is a smooth family of symmetric periodic solutions of the restricted problem, $\phi(t, \mu, \alpha)$, where α is the parameter of the family such that for $-1 < \alpha < +1$, $\phi(t, 0, \alpha)$ is an elliptic periodic solution, $\phi(t, 0, -1)$ is a direct circular periodic solution, and $\phi(t, 0, +1)$ is a retrograde

circular periodic solution of the Kepler problem in rotating coordinates. Of course, this family contains a collision orbit, but there is a natural way to continue a family through a collision. Such a family is called a *bridge of periodic solutions* (*connecting the direct and retrograde orbits*).

The complete justification of Schmidt's bridges would take too much time, but one of the bifurcations will be given here. Consider the restricted problem for small μ in Poincaré coordinates (Section IV.E.2); so, the Hamiltonian is

$$H = -\frac{1}{2P_1^2} - P_1 + \frac{1}{2}(Q_2^2 + P_2^2) + O(\mu). \tag{1}$$

These coordinates are valid in a neighborhood of the direct circular orbits when $\mu = 0$. Recall the Q_2 is an angular coodinate, and when $\mu = 0$, the direct circular orbits are $Q_2 = P_2 = 0$. In Section VI.C these periodic orbits were continued into the restricted problem for small μ, and these solutions have Q_2, P_2 coordinates which are $O(\mu)$. This result will be reproved below.

The condition for an orthogonal crossing of the line of syzygy in these coordinates is

$$Q_1 = m\pi, \qquad Q_2 = 0, \tag{2}$$

where m is an integer. Let $Q_1(t, p_1, p_2, \mu)$, $Q_2(t, p_1, p_2, \mu)$, $P_1(t, p_1, p_2, \mu)$, $P_2(t, p_1, p_2, \mu)$ be the solution which satisfies $Q_1 = Q_2 = 0$, $P_1 = p_1$, $P_2 = p_2$ when $t = 0$. Then the equations to solve for a symmetric T-periodic solution are

$$\begin{aligned} Q_1(\tfrac{1}{2}T, p_1, p_2, \mu) &= \tfrac{1}{2}(1/p_1^3 - 1)T - m\pi + O(\mu) = 0, \\ Q_2(\tfrac{1}{2}T, p_1, p_2, \mu) &= p_2 \sin(\tfrac{1}{2}T) + O(\mu) = 0. \end{aligned} \tag{3}$$

The direct circular orbits correspond to $m = \pm 1$; take $+1$ for definiteness. When $\mu = 0$ these equations have a solution $p_1^{-3} = j$ (arbitrary), $p_2 = 0$, $T = 2\pi/(j - 1)$. Since

$$\frac{\partial(Q_1, Q_2)}{\partial(t, p_2)} = \begin{pmatrix} \frac{1}{2}(j-1) & 0 \\ 0 & \sin(\pi/(j-1)) \end{pmatrix}, \tag{4}$$

which is not zero when $j \neq (s + 1)/s$, $s = 1, 2, 3, \ldots$, the implicit function theorem implies that these solutions can be continued into the restricted problem for small μ. This is a second proof of the existence of the direct circular orbits.

Assume that the Q_2, P_2 coordinates have been shifted so that the circular orbits are at $Q_2 = P_2 = 0$ for all small μ. This only affects the $O(\mu)$ terms in (1). Let k and n be relatively prime integers. The first equation in (3) has a solution $T = 2\pi n$, $p_1^{-3} = k/n$, $m = k - n$ when $\mu = 0$, and since $\partial Q_1/\partial t = (1/p_1^{-3} - 1)/2 = (k - n)/2n \neq 0$ it can be solved for $T = T(p_1, p_2, \mu) = 2(k - n)\pi(1/p_1^{-3} - 1)^{-1} + O(\mu)$. Substitute this solution into the second equation in (3) to get

$$\mathscr{L}_2(p_1, p_2, \mu) = Q_2\left(\frac{T}{2}, p_1, p_2, \mu\right) = p_2 \sin\left\{\frac{(k-n)\pi}{(1/p_1^3 - 1)}\right\} + O(\mu) = 0 \tag{5}$$

as the equation to solve. Since the circular orbit has been shifted to the Q_2, P_2 origin, Equation (5) is satisfied when $p_2 = 0$; so, p_2 is a factor. Thus, to solve (5) it is enough to solve

$$\sin\left\{\frac{(k-n)\pi}{(1/p_1^3 - 1)}\right\} + O(\mu) = 0. \tag{6}$$

This equation has a solution, $p_1 = p_1(p_2, \mu) = (n/k)^{1/3} + O(\mu)$ again by the implicit function theorem. This gives rise to a periodic solution for all p_2 that are small, including $p_2 = 0$. So this family is parameterized by p_2, $0 \leq p_2 \leq \delta$ (small), for μ small. The period of the solutions in this family is approximately $2n\pi$ for $p_2 \neq 0$. Where $p_2 = 0$, this periodic solution is the direct circular orbit established before. See Figure C.1.

D. Bifurcation at $\mathscr{L}_4$

One of the most interesting bifurcations occurs in the restricted problem at the libration point $\mathscr{L}_4$ as the mass ratio parameter passes through the critical mass ratio of Routh, μ_1. Recall that the linearized equations at $\mathscr{L}_4$ have two pairs of pure imaginary eigenvalues, $\pm\omega_1 i$, $\pm\omega_2 i$ for $0 < \mu < \mu_1$, eigenvalues $\pm i\sqrt{2}/2$ of multiplicity two for $\mu = \mu_1$, and eigenvalues $\pm\alpha \pm \beta i$, $\alpha \neq 0$, $\beta \neq 0$ for $\mu_1 < \mu \leq \frac{1}{2}$; see Section II.G. For $\mu < \mu_1$ and μ near μ_1, Lyapunov's Center Theorem VI.B.1, establishes the existence of two families of periodic solutions emanating from the libration point $\mathscr{L}_4$, and for $\mu_1 < \mu \leq \frac{1}{2}$, the Stable Manifold Theorem, V.F.1, asserts that there are no periodic solutions near $\mathscr{L}_4$. What happens to these periodic solutions as μ passes through μ_1?

In a lovely paper, Buchanan (1941) proved, up to a small computation, that there are still two families of periodic solutions emanating from the libration point $\mathscr{L}_4$ even when $\mu = \mu_1$. This is particularly interesting because the linearized equations have only one family. The small computation of a coefficient of a higher-order term was completed by Deprit (1969), thus showing that Buchanan's theorem did indeed apply to the restricted problem. Palmore (1969) investigated the question numerically and was led to the conjecture that the two families detach as a unit from the libration point and recede as μ increases from μ_1. Finally, Meyer and Schmidt (1971) established a general theorem which established Palmore's conjecture using the calculation of Deprit (1969). Unfortunately, a spurious factor of $\sqrt{2}$ occurred in the application of Deprit's calculation. Subsequently, this theorem has been reproved by several authors by essentially the same method. It has become known as the Hamiltonian Hopf bifurcation. (There have been many people who have contributed to this problem, but Hopf was not one of them.)

By the discussion in Section II.C, the normal form for a quadratic Hamiltonian (linear Hamiltonian system) with eigenvalues $\pm\omega i$, $\omega \neq 0$, with multiplicity 2, which is nonsimple, is

$$Q_0 = \omega(\xi_2\eta_1 - \xi_1\eta_2) + (\delta/2)(\xi_1^2 + \xi_2^2), \tag{1}$$

where $\delta = \pm 1$ which gives rise to the linear system of equations $\dot{z} = A_0 z$, where

$$z = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \eta_1 \\ \eta_2 \end{bmatrix}, \qquad A_0 = \begin{bmatrix} 0 & \omega & 0 & 0 \\ -\omega & 0 & 0 & 0 \\ -\delta & 0 & 0 & -\omega \\ 0 & -\delta & \omega & 0 \end{bmatrix}. \tag{2}$$

Consider a smooth quadratic perturbation of Q_0, i.e., a quadratic Hamiltonian of the form $Q(v) = Q_0 + vQ_1 + \cdots$, where v is the perturbation parameter. By the discussion in Sections VI.C and VI.D, there are three qualities that are important in the theory of normal forms for this problem, namely,

$$\Gamma_1 = \xi_2\eta_1 - \xi_1\eta_2, \qquad \Gamma_2 = \tfrac{1}{2}(\xi_1^2 + \xi_2^2), \qquad \Gamma_3 = \tfrac{1}{2}(\eta_1^2 + \eta_2^2). \tag{3}$$

The higher-order terms in $Q(v)$ are in normal form if they are functions of Γ_1 and Γ_3 only. Assume that $Q(v)$ is normalized through terms in v; so that $Q_1 = a\Gamma_1 + b\Gamma_3$ or

$$Q(v) = \omega\Gamma_1 + \delta\Gamma_2 + v(a\Gamma_1 + b\Gamma_3) + \cdots. \tag{4}$$

Only a Luddite would fail to use complex coordinates at this point; so, introduce new coordinates by

$$\begin{aligned} y_1 &= \xi_1 + i\xi_2, \qquad & y_2 &= \xi_1 - i\xi_2, \\ y_3 &= \eta_1 - i\eta_2, \qquad & y_4 &= \eta_1 + i\eta_2. \end{aligned} \tag{5}$$

This change of coordinates is symplectic with multiplier 2. Note that the reality conditions (the Luddite's bugaboo) are $y_1 = \bar{y}_2$ and $y_3 = \bar{y}_4$. We will keep the form of $Q(v)$ and make the change in the Γ's; so,

$$\Gamma_1 = i(y_2y_4 - y_1y_3), \qquad \Gamma_2 = y_1y_2, \qquad \Gamma_3 = y_3y_4. \tag{6}$$

The equations of motion are $\dot{w} = (B_0 + vB_1 + \cdots)w$, where

$$w = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}, \qquad B_0 = \begin{bmatrix} \omega i & 0 & 0 & 0 \\ 0 & \omega i & 0 & 0 \\ 0 & -\delta & \omega i & 0 \\ -\delta & 0 & 0 & -\omega i \end{bmatrix},$$

$$B_1 = \begin{bmatrix} ai & 0 & 0 & b \\ 0 & ai & b & 0 \\ 0 & 0 & ai & 0 \\ 0 & 0 & 0 & -ai \end{bmatrix}. \tag{7}$$

The characteristic polynomial of $Q(v)$ is

$$\{\lambda^2 - (\omega + va)^2\}^2 + 2vb\delta\{\lambda^2 + (\omega + va)^2\} + v^2b^2\delta^2, \tag{8}$$

which has roots

$$\lambda = \pm(\omega + va)i \pm \sqrt{-b\delta v} + \cdots. \tag{9}$$

So the coefficient a controls the way the eigenvalues move in the imaginary direction, and the coefficient b controls the way the eigenvalues split off the imaginary axis. The assumption that $b \neq 0$ means that the eigenvalues move off the imaginary axis when $b\delta v < 0$.

Now consider a nonlinear Hamiltonian system depending on the parameter v which has $Q(v)$ as its quadratic part and when $v = 0$ has been normalized in accordance with the discussion in Section VII.C through the fourth-order terms, i.e., consider

$$H(v) = \omega\Gamma_1 + \delta\Gamma_2 + v(a\Gamma_1 + b\Gamma_3) + \tfrac{1}{2}(c\Gamma_1^2 + 2d\Gamma_1\Gamma_3 + e\Gamma_3^2) + \cdots, \tag{10}$$

where the ellipsis stands for terms which are at least second order in v or fifth order in the y's Scale the variables by

$$\begin{aligned} y_1 &\to \varepsilon^2 y_1, & y_2 &\to \varepsilon^2 y_2, \\ y_3 &\to \varepsilon y_3, & y_4 &\to \varepsilon y_4, \\ v &\to \varepsilon^2 v, \end{aligned} \tag{11}$$

which is symplectic with multiplier ε^3; so, the Hamiltonian becomes

$$H = \omega\Gamma_1 + \varepsilon(\delta\Gamma_2 + vb\Gamma_3 + \tfrac{1}{2}e\Gamma_3^2) + O(\varepsilon^2). \tag{12}$$

The essential assumption is that all the terms shown actually appear, i.e., $\omega \neq 0, \delta = \pm 1, b \neq 0, e \neq 0$. The equations of motion are

$$\begin{aligned} \dot{y}_1 &= -\omega i y_1 + \varepsilon(vby_4 + e(y_3y_4)y_4), \\ \dot{y}_2 &= \omega i y_2 + \varepsilon(vby_3 + e(y_3y_4)y_3), \\ \dot{y}_3 &= \omega i y_3 - \varepsilon\delta y_2, \\ \dot{y}_4 &= -\omega i y_4 - \varepsilon\delta y_1. \end{aligned} \tag{13}$$

Note that the $O(\varepsilon^2)$ terms have been dropped for the time being. Equations (13) are of the form

$$\dot{u} = Cu + \varepsilon f(u, v) \tag{14}$$

where u is a 4-vector, f is analytic in all variables, $f(0, v) = 0$, $C = \operatorname{diag}(-\omega i, \omega i, \omega i, -\omega i)$, $\exp(CT) = I$ for $T = 2\pi/\omega$, and most importantly

$$f(e^{Ct}u, v) = e^{Ct}f(u, v). \tag{15}$$

This last property, (15), is the characterization of the normal form in the case where the matrix of the linear part is simple; see Theorem VII.C.1. The scaling has achieved this property to first order in ε. For the moment, continue to

ignore the $O(\varepsilon^2)$ terms. Let τ be a parameter (period correction parameter); then $u(t) = e^{(1-\varepsilon\tau)Ct}v$, v constant, is a solution of (15) if and only if

$$D(\tau, v, v) = \tau Cv + f(v, v) = 0. \tag{16}$$

Thus, if v satisfies (16), then $e^{(1-\varepsilon\tau)Ct}v$ is a periodic solution of (14) of period $T/(1-\varepsilon\tau) = T(1+\varepsilon\tau+\cdots)$. For Equations (13) with the $O(\varepsilon^2)$ terms omitted, one calculates

$$D(\tau, v, v) = \begin{bmatrix} (-i\omega\tau v_1 + vbv_4 + er^2v_4) \\ (i\omega\tau v_2 + vbv_3 + er^2v_3) \\ (i\omega\tau v_3 - \delta v_2) \\ (-i\omega\tau v_4 - \delta v_1) \end{bmatrix} = 0, \tag{17}$$

where $r^2 = v_3v_4$. Solving for v_1 from the last equation, substituting it into the first equation, and canceling the v_4 yields

$$\omega^2\tau^2 - \delta er^2 = \delta bv. \tag{18}$$

A solution of (18) gives rise to a 3-parameter family of periodic solutions (2-parameter family of periodic orbits) of (13) in the following way. Choose v_3 arbitrary, i.e., $v_3 = \alpha_1 + i\alpha_2$, where α_1 and α_2 are parameters. Then $v_4 = \alpha_1 - i\alpha_2$, $r^2 = \alpha_1^2 + \alpha_2^2$. Take τ arbitrary, i.e., $\tau = \alpha_3$. Then solve for v_1 and v_2 by the last two equations in (17). Fixing r determines a circle of periodic solutions which corresponds to one periodic orbit; thus, the 2-parameter family of periodic orbits is parameterized by r and τ.

The analysis depends on the sign of the two quantities δe and δb, especially δe. There are two qualitatively different cases: Case A when δe is positive, and Case B when δe is negative.

Case A: $\delta e > 0$, see Figure D.1.

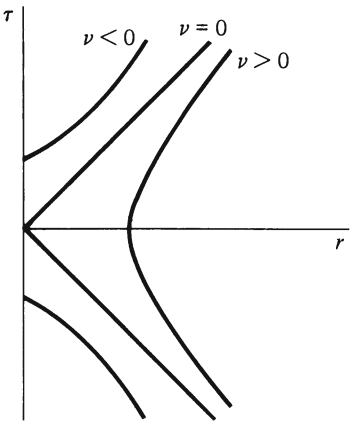

Figure D.1. Graph of (17) when δe is positive.

For definiteness, let δb be positive because the contrary case is obtained by changing the sign of v. Figure D.1 is drawn under this convention. For fixed v, the graph of (18) is a hyperbola (two lines through the origin when $v = 0$), but only the part where $r \geq 0$ is of interest. The parameter τ is the correction to the period. By the paragraph above, a fixed solution of (18), $r \neq 0$, fixes the length of v_3 and so fixes the special coordinates v_1, v_2, v_3, v_4 up to a circle. Thus, a point in the τ–r plane, $r \neq 0$, on the graph of (18) corresponds to a periodic orbit of (13) with period $T/(1 - \varepsilon\tau)$. $r = 0$, corresponds to the libration point at the origin.

By (9), when $v > 0$ and small, the eigenvalues of the linear part, $B_0 + vB_1 + \cdots$, are two distinct pairs of pure imaginary numbers; so, Lyapunov's Center Theorem implies that there are two families of periodic solutions emanating out of the origin. When $v > 0$, the graph of (18) is two curves emanating out of the line $r = 0$. This corresponds to two families of periodic solutions of (13) emanating out of the origin, and hence, it corresponds to the two Lyapunov families.

When $v = 0$, the graph of (18) is two lines emanating out of the origin, which again corresponds to two families of periodic solutions emanating out of the origin. In this case, these two families correspond to the two families of Buchanan (1941). When $v < 0$, the graph of (18) is a single curve which does not pass through the origin and thus corresponds to a single family of period orbits of (13) which do not pass through the origin.

Case A summary: The two Lyapunov families emanate from the origin when δbv is positive. These families persist when $v = 0$ as two distinct families of periodic orbits emanating from the origin. As δbv becomes negative, the two families detach from the origin as a single family and move away from the origin.

Case B: $\delta e < 0$, see Figure D.2.

For definiteness let δb be positive as before. Figure D.2 is drawn under this convention. For fixed $v > 0$, the graph of (18) is an ellipse.

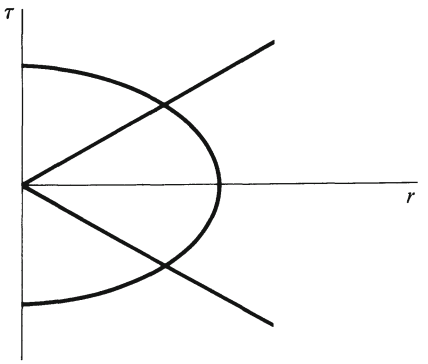

Figure D.2. Graph of (17) when δe is negative.

By (9) when $\nu > 0$ and small, the eigenvalues of the linear part, $B_0 + \nu B_1 + \cdots$, are two distinct pairs of pure imaginary numbers; so, Lyapunov's Center Theorem implies that there are two families of periodic solutions emanating out of the origin. These families correspond to the upper and lower halves of the ellipse. In this case, the two Lyapunov families are globally connected. As ν tends to zero, this family shrinks to the origin and disappears. For $\nu < 0$ there are no such periodic solutions.

Case B summary: The two Lyapunov families emanate from the origin when $\delta b \nu$ is positive and are globally connected. These families shrink to the origin as $\delta b \nu$ tends to zero through positive values. When $\delta b \nu$ is negative, there are no such periodic solutions.

Now we will show that these conclusions remain valid when the $O(\varepsilon^2)$ terms in (13) are present. If the $O(\varepsilon^2)$ terms are included, (13) is of the form $\dot{u} = Cu + f(u, \nu) + O(\varepsilon^2)$. Let $\phi(t, v, \nu, \varepsilon)$ be the general solution of this equation with $\phi(0, v, \nu, \varepsilon) = v$. Let $\psi(v, \tau, \nu, \varepsilon)$ be this solution after a time $T(1 + \varepsilon\tau)$, i.e.,

$$\begin{aligned} \psi(v, \tau, \nu, \varepsilon) &= \phi(T(1 + \varepsilon\tau), v, \nu, \varepsilon) = v + \varepsilon D(\tau, v, \nu) + O(\varepsilon^2) \\ &= v + \varepsilon\{\tau C v + f(v, \nu)\} + O(\varepsilon^2). \end{aligned} \tag{19}$$

A periodic solution corresponds to a solution of $\psi(v, \tau, \nu, \varepsilon) = v$, and so the equation to be solved is

$$\mathscr{D}(v, \tau, \nu, \varepsilon) = (\psi(v, \tau, \nu, \varepsilon) - v)/\varepsilon = D(\tau, v, \nu) + O(\varepsilon) = 0. \tag{20}$$

Equations (20) are dependent since Equations (13) admit H as an integral. Since $H(v + \varepsilon\mathscr{D}) = H(v)$, it follows by the mean value theorem that grad $H(v^*)\mathscr{D} = 0$, where v^* is a point between v and $v + \varepsilon\mathscr{D}$. Since grad $H(v^*) = (-i\omega y_3, \ldots) + \cdots$, if $\mathscr{D}_2 = \mathscr{D}_3 = \mathscr{D}_4 = 0$, then $\mathscr{D}_1 = 0$, except maybe when $y_3 = 0$. Thus, only the last three equations in (20) need to be solved because the solutions sought have $y_3 \neq 0$.

From the last two equations in (20), one solves, by the implicit function theorem for v_1 and v_2, to get $v_1 = -i\omega\tau\delta v_4 + \cdots$, $v_2 = i\omega\tau\delta v_3 + \cdots$. Substitute these solutions into the second equation to get

$$d(v_3, v_4, \tau, \nu, \varepsilon) = (\omega^2\tau^2 - e\delta r^2 - b\delta\nu)(-\delta v_3) + \varepsilon g(v_3, v_4, \tau, \varepsilon) = 0. \tag{21}$$

Since the origin is always an equilibrium point, d and g vanish when v_3 and v_4 are zero. Let $v_3 = re^{i\theta}$, $v_4 = re^{-i\theta}$, and divide (21) by $(-\delta r)$ to get

$$d'(r, \theta, \tau, \nu, \varepsilon) = (\omega^2\tau^2 - e\delta r^2 - b\delta\nu)e^{i\theta} + \varepsilon g'(r, \theta, \tau, \varepsilon) = 0. \tag{22}$$

Since $e\delta \neq 0$, this equation can be solved for r^2 to get $r^2 = R(\theta, \tau, \nu, \varepsilon) = (\omega^2\tau^2 - b\delta\nu)/e\delta + \cdots$ for all θ, all τ, $|\tau| < \tau_0$, and all ε, $|\varepsilon| < \varepsilon_0$. So $r = \sqrt{R}$ yields a real solution when $R > 0$. The analysis of the sign of R leads to the same qualitative conclusions as before.

To get back to the unscaled equations, fix $\varepsilon = \varepsilon_0/2$ for all times. The scaling is global and so it is invertible globally. Trace back from the solution $r^2 = R(\theta, \tau, \nu, \varepsilon_0/2)$ to get $\xi_1, \xi_2, \eta_1, \eta_2$ as functions of θ, τ, ν.

Theorem 1. *Consider a Hamiltonian of the form* (10) *with* $\omega \neq 0$, $\delta = \pm 1$, $b \neq 0$, $e \neq 0$.

Case A: $\delta e > 0$. *The two Lyapunov families emanate from the origin when* δbv *is small and positive. These families persist when* $v = 0$ *as two distinct families of periodic orbits emanating from the origin. As* δbv *becomes negative, the two families detach from the origin as a single family and recede from the origin.*

Case B: $\delta e < 0$. *The two Lyapunov families emanate from the origin when* δbv *is small and positive, and the families are globally connected. This global family shrinks to the origin as* δbv *tends to zero through positive values. When* δbv *is small and negative, there are no periodic solutions close to the origin.*

One can compute the multipliers approximately to show that in Case A the periodic solutions are elliptic. In Case B, the periodic solutions are initially elliptic as they emanate from the origin but go through extremal bifurcations to become hyperbolic. See the Problems Section.

Case A occurs in the restricted problem at $\mathscr{L}_4$ as the mass ratio parameter μ passes through the critical mass ratio μ_1. This theorem verifies the numeric experiments of Palmore.

E. Further Reading

Most of the major general works on Hamiltonian systems have very little on the topic of bifurcations of periodic solutions. Most of the general works on bifurcations have very little on Hamiltonian systems. Therefore, the works suggested here will be rather special in nature.

One of the most interesting books on periodic solutions of Hamiltonian systems is Poincaré (1899). It is old but not out of date. The survey paper, Deprit and Henrard (1968), gives a good introduction to the basic theory of periodic solutions of a Hamiltonian with lots of good numerical results from the restricted problem.

Hale and Chow (1982) and Guckenheimer and Holmes (1983) treat many types of bifurcation problems including bifurcations in Hamiltonian systems. These books cover a much wider area.

The results on bifurcations of periodic solutions given in this chapter are found in one form or another in Cherry (1928), Bruno (1970a,b), and Meyer (1970).

Problems

1. Let $Q = Q(q, p)$, $P = P(q, p)$ define a smooth area-preserving diffeomorphism of a neighborhood of the origin $q = p = 0$.

a. Show that $\Omega = (P - p)d(Q + q) - (Q - q)d(P + p)$ is a closed form in q, p (i.e., $d\Omega = 0$), and so by Poincaré's lemma, there is a function $G = G(q, p)$ defined in a neighborhood of the origin such that $dG = \Omega$.

b. Let the origin be a fixed point whose multipliers are not -1; so, $\xi = Q + q$, $\eta = P + p$ defines new coordinates, and $\partial G/\partial\xi = P - p$, $\partial G/\partial\eta = Q - q$. Thus, a fixed point corresponds to critical points of G. Show that if $G = \{\alpha\xi^2 + 2\beta\xi\eta + \gamma\eta^2\}/2$ with $4\Delta = \alpha\gamma - \beta^2 \neq -1$, then

$$\frac{\partial(Q, P)}{\partial(q, p)} = \frac{1}{1 + 4\Delta}\begin{pmatrix}(1 - \beta^2) - \alpha\gamma & -2\gamma \\ 2\alpha & (1 + \beta^2) - \alpha\gamma\end{pmatrix}.$$

Thus, a maximum or minimum of G corresponds to an elliptic fixed point, and a saddle point corresponds to a hyperbolic fixed point.

c. Draw the level surfaces of $G = q^2/2 + \varepsilon p + p^3/3$ as the parameter varies.

d. Show that the fixed point is an extremal fixed point if and only if $G = \partial G/\partial q = \partial G/\partial p = \partial^2 G/\partial p^2 = \partial^2 G/\partial q\partial p = 0$ and $\partial^2 G/\partial q^2 \neq 0$, $\partial^2 G/\partial p\partial\varepsilon \neq 0$, $\partial^3 G/\partial p^3 \neq 0$. See Meyer (1970).

2. Consider the forced Duffing equation at 1/3 resonance. That is, consider

$$H_* = I + \varepsilon\{\gamma I^2 \cos^4\phi + AI^{1/2}\cos\phi\cos 3t\} + (\varepsilon^2/2)\delta I\cos^2\phi,$$

which is the forced Duffing equation written in action-angle variables. Note that the Hamiltonian is $2\pi/3$-periodic in t. The normalized Hamiltonian is

$$H^* = I + \varepsilon(3\gamma/8)I^2 + (\varepsilon^2/64)\{-2A\gamma I^{3/2}\cos 3(t + \phi) + 17\gamma^2 I^3 + 16\delta I - A^2\} + \cdots.$$

a. Write the Hamiltonian, the normalized Hamiltonian, and the equations of motion in rectangular coordinates.

b. Analyze the normalized systems. Remember to bring the harmonic to the origin.

3. The bifurcation for the forced Duffing equation at 1/2 resonance is not, as predicted in the section, a generic bifurcation, and this is due to the fact that this equation has additional symmetries because the Hamiltonian is even. Consider a Hamiltonian like the forced Duffing equation which has a cubic term in the Hamiltonian (a quadratic in the equations). That is, consider

$$H_* = I + \varepsilon\{\varkappa I^{3/2}\cos^3\phi + AI^{1/2}\cos\phi\cos pt\} + (\varepsilon^2/2)\delta I\cos^2\phi.$$

Note that the Hamiltonian is π-periodic in t. The normalized Hamiltonian is

$$H^* = I + (\varepsilon^2/48)\{12A\varkappa I\cos 2(t + \phi) + 45\varkappa^2 I^2 + 24\delta I - 4A^2\} + \cdots.$$

a. Write the Hamiltonian, the normalized Hamiltonian, and the equations of motion in rectangular coordinates.

b. Analyze the normalized systems. Remember to bring the harmonic to the origin.

4. Using MACSYMA, REDUCE, MAPLE, or an algebraic processor of your choice, write a normalization routine which normalizes Duffing's equation at $q - p$ resonance, i.e., write a program which normalizes

$$H_* = qI + \varepsilon\{\delta I\cos^2\phi + \gamma I^2\cos^4\phi + AI^{1/2}\cos\phi\cos pt\}.$$

Analyze the cases $p/q = 1/3, 2/3, 3/4$, etc.

a. Show that the Jacobian of D in Equation (D.17) is

$$\frac{\partial D}{\partial v} = \begin{bmatrix} i\omega\tau & 0 & ev^2 & vb + 2er^2 \\ 0 & i\omega\tau & vb + 2er^2 & ev_3^2 \\ 0 & -\delta & i\omega\tau & 0 \\ -\delta & 0 & 0 & -i\omega\tau \end{bmatrix}.$$

b. Show that because of the dependency of the equations, the Jacobian is singular. Also show that the determinant of the minor obtained by deleting the first row and third column is ev_3^2, which is nonzero if $v_3 \neq 0$.

c. Show that the multipliers of the solutions found in Section D are of the form 1, 1, $1 + \varepsilon\mu_1 + \cdots$, $1 + \varepsilon\mu_2 + \cdots$, where μ_1 and μ_2 are the nonzero eigenvalues of $\partial D/\partial v$.

d. Show that in Case A of Section D, i.e., $\delta e > 0$, that the periodic solutions found are elliptic.

5. Use the notation of Section D.

a. Show that the value of the Hamiltonian (12) along a solution of (18) is $H = -2\omega^2\tau\delta r^2 + (2\omega^2/e)\{\delta bv\tau - \omega^3\tau^3\}$.

b. Show that in Case A the periodic solutions can be parameterized by the Hamiltonian.

6. Consider a Hamiltonian of the form

$$H = k\omega I_1 + \omega I_2 + \tfrac{1}{2}(AI_1^2 + 2BI_1I_2 + CI_2^2) + \cdots,$$

where $I_i = (x_i^2 + y_i^2)/2$, $\omega > 0$, k is a nonzero integer, and the ellipsis represents terms of degree at least 5 in the x's and y's.

a. Show that Lyapunov's Center Theorem implies the existence of a family of periodic solutions (*the short period family*) of approximate period $2\pi/k\omega$ which emanate from the origin when $|k| > 1$.

b. Use the ideas of Section D to show that there is a family of periodic solutions (*the long period family*) of approximate period $2\pi/\omega$ which emanates from the origin when $B - kC \neq 0$.

c. Using the normal form calculations for the restricted problem at $\mathscr{L}_4$, show that the long period family exists even when $\mu = \mu_i$, for $i = 4, 5, 6, \ldots$.

CHAPTER IX

Stability and KAM Theory

Questions of stability of orbits have been of interest since Newton first set down the laws that govern the motion of the celestial bodies. "Is the universe stable?" is almost a theological question. Even though the question is old and important, very little is known about the problem, and much of what is known is difficult to come by.

This chapter contains an introductions to the question of the stability of orbits of Hamiltonian systems and the celebrated KAM theory. This subject could be the subject of a complete book; so, the reader will find only selected topics presented here.

A. Elementary Stability Results

Consider the differential equation

$$\dot{x} = f(x), \tag{1}$$

where f is a smooth function from $O \subset \mathbb{R}^m$ into $\mathbb{R}^m$. Let the equation have an equilibrium point at $\xi_0 \in O$; so, $f(\xi_0) = 0$. Let $\phi(t, \xi)$ be the general solution of (1). The equilibrium point ξ_0 is said to be *positively* (*respectively negatively*) stable if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $\|\phi(t, \xi) - \xi_0\| < \varepsilon$ for all $t \geq 0$ (respectively $t \leq 0$) whenever $\|\xi - \xi_0\| < \delta$. The equilibrium point ξ_0 is said to be *stable* if it is both positively and negatively stable. In many books *stable* means positively stable, but the above convention is the common one in the theory of Hamiltonian differential equations. The equilibrium ξ_0 is *unstable* if it is not stable. The adjectives "positively" and "negatively" can be used with "unstable" too. The equilibrium ξ_0 is *asymptotically stable* if it is stable, and there is an $\eta > 0$ such that $\phi(t, \xi) \to \xi_0$ as $t \to +\infty$ for all $\|\xi - \xi_0\| < \eta$.

Recall the one result already given on stability in Theorem I.A.2, which states that a local minimum or maximum of a Hamiltonian is a stable equilibrium point. So for a general Newtonian system of the form $H = p^T Mp/2 + V(q)$, a minimum of the potential V is a stable equilibrium point because the matrix M is positive definite. It has been stated in Malkin (1952), LaSalle and Lefschetz (1961), and Courant and Hilbert (1953) that an equilibrium point of V which is not a minimum is unstable. Laloy (1976) showed that for $V(q_1, q_2) = \exp(-1/q_1^2)\cos(1/q_1) - \exp(-1/q_2^2)\{\cos(1/q_2) + q_2^2\}$, the origin is a stable equilibrium point, and yet the origin is not a local minimum for V. See Taliaferro (1980) for some positive results along these lines.

Henceforth, let the equilibrium point be at the origin. A standard approach is to linearize the equations, i.e., write (1) in the form

$$\dot{x} = Ax + g(x), \tag{2}$$

where $A = \partial f(0)/\partial x$ and $g(x) = f(x) - Ax$; so, $g(0) = \partial g(0)/\partial x = 0$. The eigenvalues of A are called the *exponents* (*of the equilibrium point*). If all the exponents have negative real parts, then a classical theorem of Lyapunov states that the origin is asymptotically stable; see LaSalle and Lefschetz (1961), Coddington and Levinson (1955), or Hale (1972). By Theorem II.C.1, the eigenvalues of a Hamiltonian matrix are symmetric with respect to the imaginary axis; so, this theorem never applies to Hamiltonian systems. In fact, since the flow defined by a Hamiltonian system is volume preserving, an equilibrium point can never be asymptotically stable.

Lyapunov also proved that if one exponent has a positive real part, then the origin is unstable. See LaSalle and Lefschetz (1961), Coddington and Levinson (1955), or Hale (1972). Thus, for the restricted 3-body problem the Euler collinear libration points, $\mathscr{L}_1$, $\mathscr{L}_2$, $\mathscr{L}_3$, are always unstable, and the Lagrange triangular libration points, $\mathscr{L}_4$ and $\mathscr{L}_5$, are unstable for $\mu_1 < \mu < 1 - \mu_1$ by the results of Section II.G.

Thus, a necessary condition for stability of the origin is that all the eigenvalues be pure imaginary. It is easy to see that this condition is not sufficient in the non-Hamiltonian case. For example, the exponents of

$$\begin{aligned} \dot{x} &= x_2 + x_1(x_1^2 + x_2^2), \\ \dot{x}_2 &= -x_1 + x_2(x_1^2 + x_2^2) \end{aligned} \tag{3}$$

are $\pm i = \pm\sqrt{-1}$, and yet the origin is unstable. (In polar coordinates, $\dot{r} = r^3 > 0$.) However, Equation (3) is not Hamiltonian.

In the second edition (1917) of Whittaker's book on dynamics, the equations of motion about the Lagrange point $\mathscr{L}_4$ are linearized, and the assertion is made that the libration point is stable for $0 < \mu < \mu_1$ on the basis of this linear analysis. In the third edition of Whittaker (1927) this assertion was dropped, and an example due to Cherry (1926) was included. A careful look at Cherry's example shows that it is a Hamiltonian system of two degrees of freedom, and the linearized equations are two harmonic oscillators with

frequencies in a ratio of 2 : 1. The Hamiltonian is in the normal form given the Theorem VI.C.1, i.e., in action-angle variables. Cherry's example is

$$H = I_1 - 2I_2 + I_1 I_2^{1/2} \cos(2\phi_1 + \phi_2). \tag{4}$$

To see that the origin is unstable, consider the Lyapunov function $W = -I_1 I_2^{1/2} \sin(2\phi_1 + \phi_2)$. Let $W(t) = W(\phi(t, p))$, and so the derivative of W among the solutions of the equations defined by (4) is $\dot{W} = 2I_1 I_2 + (\frac{1}{2})I_1^2$. Let Ω be the region where $W > 0$. In Ω, $I_1 \neq 0$; so, $\dot{W} > 0$ in Ω. Ω has points arbitrarily close to the origin, so for any $\delta > 0$ there is a point $p \in \Omega$ with $\|p\| < \delta$ and $W(p) > 0$. Assume that the solution $\phi(t, p)$ remains bounded for $t > 0$; so, $\|\phi(t, p)\| < M$. $\dot{W}(\phi(t, p)) > 0$; so, $W(t)$ is increasing, and $W(\phi(t, p)) \geq W(p)$. Since the set $K = \{x: W(x) \geq W(p) \text{ and } \|x\| \leq M\}$ is compact, there is a constant $\varkappa$, so that $\dot{W} \geq \varkappa > 0$ on K and $\phi(t, p) \in K$ for $t \geq 0$. Thus, $W(t) \geq W(p) + \varkappa t$ for all $t \geq 0$, but this is impossible because $\phi(t, p)$ was assumed bounded. Therefore, $\phi(t, p)$ leaves any neighborhood of the origin, and hence, the origin is unstable. This argument is essentially the proof of Chetaev's theorem; see LaSalle and Lefschetz (1961).

This argument can be extended to show that the origin is still unstable when higher-order terms are added to (4), and this extension can be used to prove that the Lagrange triangular point $\mathscr{L}_4$ is unstable when $\mu = \mu_2$. A similar argument shows that $\mathscr{L}_4$ is unstable when $\mu = \mu_3$. These results were established by Markeev (1966) and Alfriend (1970, 1971) by other methods.

B. The Invariant Curve Theorem

We shall return to questions about the stability of equilibrium points later, but now consider the corresponding question for maps. Let

$$F(x) = Ax + f(x) \tag{1}$$

be a diffeomorphism of a neighborhood of a fixed point $\xi \in \mathbb{R}^m$; so, $f(\xi) = 0$ and $\partial f(\xi)/\partial x = 0$. The eigenvalues of A are the *multipliers* of the fixed point.

The fixed point is said to be *positively* (*respectively negatively*) *stable* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $\|F^k(x) - \xi\| < \varepsilon$ for all $\|x - \xi\| < \delta$ *and* $k > 0$ (respectively $k < 0$). The fixed point is *stable* if it is both positively and negatively stable. The fixed point is *unstable* if it is not stable. The adjectives "positively" and "negatively" can be used with "unstable" too. The fixed point is called *asymptotically stable* if it is positively stable, and there is an $\eta > 0$ such that $F^k(x) \to 0$ as $k \to +\infty$ for all $\|x - \xi\| < \eta$.

Analogs of Lyapunov's theorems exist for diffeomorphisms and are developed in the Problems Sections. If all the multipliers have modulus less than 1 then the fixed point is asymptotically stable. By Theorem II.C.1, a symplectic matrix cannot have all of its eigenvalues less than 1 in modulus, so the above

theorem does not apply to symplectomorphisms. Also, if one eigenvalue of A has modulus greater than 1, then the fixed point is unstable. Thus, a necessary condition for stability of a fixed point is that all its multipliers be of unit modulus. But as in the case of equilibrium points, this is not sufficient; see the Problems Section.

Let us specialize by letting the fixed point be the origin in $\mathbb{R}^2$ and by letting (1) be area preserving (symplectic). Assume that the origin is an elliptic fixed point; so, A has eigenvalues λ and $\lambda^{-1} = \bar{\lambda}$, $|\lambda| = 1$. If $\lambda = 1$, $\sqrt[2]{1} = -1$, $\sqrt[3]{1} = e^{\pm 2\pi i/3}$, or $\sqrt[4]{1} = i$, then typically the origin is unstable; see the Problems Section or Meyer (1971).

Therefore, let us consider the case when λ is not an mth root of unity for $m = 1, 2, 3, 4$. In this case, the map can be put into normal form up through terms of order three by Theorem VI.E.3, i.e., there are symplectic action-angle coordinates, I, ϕ, such that in these coordinates, $F\colon (I, \phi) \to (I', \phi')$, where

$$\begin{aligned} I' &= I + c(I, \phi), \\ \phi' &= \phi + \omega + \alpha I + d(I, \phi), \end{aligned} \tag{2}$$

$\lambda = \exp(\omega i)$, and c, d and $O(I^{3/2})$.

For the moment assume c and d are zero; so, the map (2) takes circles $I = I_0$ into themselves, and if $\alpha \neq 0$, each circle is rotated by a different amount. The circle $I = I_0$ is rotated by an amount $\omega + \alpha I_0$. When $\omega + \alpha I_0 = 2\pi p/q$, where p and q are relatively prime integers, then each point on the circle $I = I_0$ is a periodic point of period q. Since they are not isolated, they are not elementary in the sense of Section V.E. By the discussion in Section A, some of these periodic points persist even when c and d are not zero, but in general they are elliptic or hyperbolic, and so they are isolated and finite in number; see Chapter X or Meyer (1970). Thus, the circles with rational rotation are destroyed by the small terms c and d.

If $\omega + \alpha I_0 = 2\pi\delta$, where δ is irrational, then the orbits of a point on the circle $I = I_0$ are dense by the discussion in Section I.B.6 ($c = d = 0$ still). One of the most celebrated theorems in Hamiltonian mechanics states that many of these circles persist as invariant curves. In fact, there are enough invariant curves encircling the fixed point that they assure the stability of the fixed point. This is the so-call "invariant curve theorem" stated below.

In Section V.B, the billiards table problem and the linear crystal model were introduced and shown to define an area-preserving mapping of the annulus. In Chapter X a discussion of area-preserving maps of the annulus is given. The invariant curve theorem applies to examples of this type also, and it will be stated for maps a little more general than those defined by a symplectomorphism at an elliptic fixed point.

The invariant curve theorem depends on some interesting number theoretic facts. In Section I.B, it was shown that for any irrational number δ and any number $\varepsilon > 0$, there are relatively prime integers p and q such that

$$\left|\delta - \frac{p}{q}\right| < \frac{\varepsilon}{q}. \tag{3}$$

Formula (3) says more than an irrational can be approximated by a rational; it gives some information on how well an irrational can be approximated by rationals. One can construct irrational numbers which are really well approximated by rationals, say, for which $|\delta - p/q| < \varepsilon/q^2$ or $< \varepsilon/q^{223}$, by taking a decimal expansion with extremely long sequences of zeros. However, there are irrationals which are badly approximated by rational in the sense of the following theorem.

Theorem 1. *Let I be any closed interval, and let $K > 0$ be a fixed constant. The set $I(K)$ of irrational numbers, δ, in a closed interval I for which*

$$\left|\delta - \frac{p}{q}\right| > \frac{K}{q^3} \tag{4}$$

for all integers p and $q \neq 0$ is dense in I and has positive measure. The measure of $I(K)$ tends to the measure of I as $K \to 0$. Moreover, all algebraic irrationals, like $\sqrt{2}$, etc., belong to $I(K)$ for some K.

PROOF. See Arnold (1963a, 1978). ∎

Such irrationals are said to be *badly approximated by rationals.* The invariant curve theorem states that the circles whose rotation number are badly approximated by rationals persist under small perturbations.

Theorem 2 (The Invariant Curve Theorem). *Consider the mapping $F\colon (I, \phi) \to (I', \phi')$ given by*

$$\begin{aligned} I' &= I + \varepsilon^{r+s} c(I, \phi, \varepsilon), \\ \phi' &= \phi + \omega + \varepsilon^s h(I) + \varepsilon^{s+r} d(I, \phi, \varepsilon), \end{aligned} \tag{5}$$

where (i) *c and d are smooth for $0 \leq a \leq I < b < \infty$, $0 \leq \varepsilon \leq \varepsilon_0$, and all ϕ,* (ii) *c and d are 2π periodic in ϕ,* (iii) *r and s are integers $s \geq 0$, $r \geq 1$,* (iv) *h is smooth for $0 \leq a \leq I < b < \infty$,* (v) *$dh(I)/dI \neq 0$ for $0 \leq a \leq I < b < \infty$, and* (vi) *if Ξ is any continuous closed curve of the form $\Xi = \{(I, \phi)\colon I = \Theta(\phi), \Theta\colon \mathbb{R} \to [a, b]$ continuous and 2π periodic$\}$, then $\Xi \cap F(\Xi) \neq \varnothing$.*

Then for sufficiently small ε, there is a continuous F-invariant curve Γ of the form $\Gamma = \{(I, \phi)\colon I = \Phi(\phi), \Phi\colon \mathbb{R} \to [a, b]$ continuous and 2π periodic$\}$. The rotation number (see Chapter X) of F on Γ is an irrational which is badly approximated by rationals in the sense of theorem 1.

Remarks. (1) The original idea for a theorem of this type was in Kolmogorov (1954, 1957), but the map was assumed analytic, and the analog of the invariant curve was shown to be analytic. In the original paper by Moser (1962),

where this theorem was proved, the degree of smoothness required of f, g, and h was very large, C^{333}, and the invariant curve was shown to be continuous. This spread led to a great deal of work to find the least degree of differentiabilty required of f, g, and h to get the most differentiabilty for the invariant curve. However, in the interesting examples, f, g, and h are analytic, and the existence of a continuous invariant curve yields the necessary stability.

(2) Assumption (v) is the *twist* assumption discussed above, and the map is a perturbation of a twist map for small ε.

(3) Assumption (vi) rules out the obvious example where F maps every point radially out or radially in. If F preserves the inner boundary $I = a$ and is area preserving, then assumption (vi) is satisfied.

(4) Since the theorem can be applied to any subinterval of $[a, b]$, the theorem implies the existence of an infinite number of invariant curves. In fact, the proof shows that the measure of the invariant curves is positive and tends to the measure of the full annulus $a \leq I \leq b$ as $\varepsilon \to 0$. If $s = 0$ and K are fixed in advance, then there is an $\varepsilon_0 > 0$ such that for all irrationals δ, $h(a) < \delta < h(b)$, satisfying (4) there is an invariant curve for F with rotation number δ.

PROOF. The proof of this theorem is quite technical. See Siegel and Moser (1971) and Herman (1983, 1986) for a complete discussion of this theorem and related results. ∎

C. A Simple Example—Duffing's Equation Again

As a simple example, consider the elliptic fixed point discussed at the beginning of the last section which is in normal form (B.2). Scale by $I \to \varepsilon^2 J$; so, Equation (B.2) becomes

$$\begin{aligned} J' &= J + \varepsilon^3 a(J, \phi), \\ \phi' &= \phi + \omega + \alpha\varepsilon^2 J + \varepsilon^3 b(J, \phi), \end{aligned} \tag{1}$$

and Theorem B.2 applies. Hence, there is an invariant curve in the punctured disk $0 < J \leq 1$ for all ε sufficiently small. Therefore, in the original unscaled variables, there is an invariant curve in $0 < I \leq \varepsilon^2$, and so there are invariant curves arbitrarily close to the origin, and so the origin is a stable fixed point.

Now consider the forced Duffing's equation with Hamiltonian

$$H = \frac{1}{2}(q^2 + p^2) + \frac{\gamma}{4}q^4 + \gamma^2 \cos \omega t, \tag{2}$$

where ω is a constants and $\gamma \neq 0$ is considered as a small parameter. This Hamiltonian is periodic with period $2\pi/\omega$. As shown in Section A, if $\omega \neq 1$, 2, 3, 4, the system has a small (order γ^2) $2\pi/\omega$ periodic solution, called the harmonic. Using the calculations in Section VI.B, the period map was shown to be

$$I' = I + O(\gamma^2),$$
$$\phi' = \phi - 2\pi/\omega - (3\pi\gamma/2\omega)I + O(\gamma^2), \tag{3}$$

where the fixed point corresponding to the harmonic has been moved to the origin. As in the above, the harmonic is stable.

D. Applications to the Restricted Problem

In Chapter VI, a small parameter was introduced into the restricted problem in three ways. First, the small parameter was the mass ratio parameter μ; second, the small parameter section was a distance to a primary; and third, the small parameter was the reciprocal of the distance to the primaries.

In all three cases an application of the invariant curve theorem can be made. Only the first and third will be given here because the computations are easy in these cases.

1. Invariant Curves for Small Mass

The Hamiltonian of the restricted problem (I.C.9) for small μ is

$$H = \frac{\|y\|^2}{2} - x^T K y - \frac{1}{\|x\|} + O(\mu). \tag{1}$$

For $\mu = 0$ this is the Hamiltonian of the Kepler problem in rotating coordinate. Be careful the $O(\mu)$ term has a singularity at the primaries. When $\mu = 0$ and Delaunay coordinates are used the Hamiltonian (1) becomes

$$H = \frac{-1}{2L^3} - G \tag{2}$$

and the equations of motion become

$$\dot{\ell} = 1/L^3, \qquad \dot{L} = 0,$$
$$\dot{g} = -1, \qquad \dot{G} = 0. \tag{3}$$

The variable g, the argument of the perihelion, is an angular variable. $\dot{g} = -1$ implies that g is steadily decreasing from 0 to -2π and so $g \equiv 0 \bmod 2\pi$ defines a cross section. The first return time is 2π. Let ℓ, L be coordinates in the intersection of the cross section $g \equiv 0$ and the level set $H =$ constant. The Poincaré map in these coordinates is

$$\ell' = \ell + 2\pi/L^3, \qquad L' = L. \tag{4}$$

Thus, when $\mu = 0$ the Poincaré in the level set is a twist map. By the invariant curve theorem some of these invariant curves persist for small μ.

2. The Stability of Comet Orbits

Consider the Hamiltonian of the restricted problem scaled as was done in Section VI.E in the discussion of comet orbits, i.e., the Hamiltonian VI.E.1. In Poincaré variables it is

$$H = -P_1 + \frac{1}{2}(Q_2^2 + P_2^2) - \varepsilon^3 \frac{1}{2P_1^2} + O(\varepsilon^5), \tag{5}$$

where Q_1 is an angle defined modulo 2π, P_1 is a radial variable, and Q_1, P_1 are rectangular variables. For typographical reasons drop, but don't forget, the $O(\varepsilon^5)$. The equations of motion are

$$\begin{aligned} \dot{Q}_1 &= -1 + \varepsilon^3/P_1^3, & \dot{P}_1 &= 0, \\ \dot{Q}_2 &= P_2, & \dot{P}_2 &= -Q_2. \end{aligned} \tag{6}$$

The circular solutions are $Q_2 = P_2 = 0 + O(\varepsilon^5)$ in these coordinates. Translate the coordinates so that the circular orbits are exactly $Q_2 = P_2 = 0$; this does not affect the displayed terms in the equations. The solutions of (6) are

$$\begin{aligned} Q_1(t) &= Q_{10} + t(-1 + \varepsilon^3/P_1^3), & P_1(t) &= P_{10}, \\ Q_2(t) &= Q_{20}\cos t + P_{20}\sin t, & P_2(t) &= -Q_{20}\sin t + P_{20}\cos t. \end{aligned} \tag{7}$$

Work near $P_1 = 1, Q_2 = P_2 = 0$ for ε small. The time for Q_1 to increase by 2π is

$$T = 2\pi/|-1 + \varepsilon^3/P_1^3| = 2\pi(1 + \varepsilon^3 P_1^{-3}) + O(\varepsilon^6). \tag{8}$$

Thus,

$$\begin{aligned} Q' = Q_2(T) &= Q\cos 2\pi(1 + \varepsilon^3 P_1^{-3}) + P\sin 2\pi(1 + \varepsilon^3 P_1^{-3}) \\ &= Q + vPP_1^{-3} + O(v^2), \\ P' = P_2(T) &= -Q\sin 2\pi(1 + \varepsilon^3 P_1^{-3}) + P\cos 2\pi(1 + \varepsilon^3 P_1^{-3}) \\ &= -vQP_1^{-3} + P + O(v^2), \end{aligned} \tag{9}$$

where $Q = Q_{20}, P = P_{20}$, and $v = 2\pi\varepsilon^3$. Let $H = 1$, and solve for P_1 to get

$$P_1 = -1 + \tfrac{1}{2}(Q^2 + P^2) + O(v), \tag{10}$$

and hence

$$P_1^{-3} = -1 - \tfrac{3}{2}(Q^2 + P^2) + O(v). \tag{11}$$

Substitute this into (9) to get

$$\begin{aligned} Q' &= Q + vP(-1 - \tfrac{3}{2}(Q^2 + P^2)) + O(v^2), \\ P' &= P - vQ(-1 - \tfrac{3}{2}(Q^2 + P^2)) + O(v^2). \end{aligned} \tag{12}$$

This is the section map in the energy surface $H = 1$. Change to action-angle variables, $I = (Q^2 + P^2)/2$, $\phi = \tan^{-1}(P/Q)$, to get

$$\begin{aligned} I' &= I + O(v^2), \\ \phi' &= \phi + v(-1 - 3I) + O(v^2). \end{aligned} \tag{13}$$

This is a twist map. Thus, the continuation of the circular orbits into the restricted problem is stable.

E. Arnold's Theorem

The invariant curve theorem can be used to establish a stability result for equilibrium points as well. In particular, we have Arnold's stability theorem. Arnold's theorem was originally proved independent of the invariant curve theorem; see Arnold (1963a, b), and the proof given here is taken from Meyer and Schmidt (1986a).

As discussed above, the only way an equilibrium point can be stable is if the eigenvalues of the linearized equations, i.e., the exponents, are pure imaginary. If the Hamiltonian is positive definite, then the equilibrium is stable by Theorem I.A.2. Also from Cherry's example it is not enough to consider the linear terms alone. Thus, the case to consider is the case when exponents are pure imaginary, and the Hamiltonian is not positive definite.

Consider the two-degree-of-freedom case for simplicity, and assume the Hamiltonian has been normalized a bit. Specifically, consider a Hamiltonian H in the symplectic coordinates x_1, x_2, y_1, y_2 of the form

$$H = H_2 + H_4 + \cdots + H_{2N} + H^{\dagger}, \tag{1}$$

where

(i) H is real analytic in a neighborhood of the origin in $\mathbb{R}^4$;
(ii) H_{2k}, $1 \leq k \leq N$, is a homogeneous polynomial of degree k in I_1, I_2, where $I_i = (x_i^2 + y_i^2)/2$, $i = 1, 2$;
(iii) $H^{\dagger}$ has a series expansion which starts with terms at least of degree $2N + 1$;
(iv) $H_2 = \omega_1 I_1 - \omega_2 I_2$, ω_i nonzero constants;
(v) $H_4 = \frac{1}{2}(AI_1^2 - 2BI_1I_2 + CI_2^2)$, A, B, C constants.

There are several implicit assumptions in stating that H is of the above form. Since H is at least quadratic, the origin is an equilibrium point. By (iv), H_2 is the Hamiltonian of two harmonic oscillators with frequencies ω_1 and ω_2; so, the linearized equations of motion are two harmonic oscillators. The sign convention is to conform with the sign convention at $\mathscr{L}_4$. It is not necessary to assume that ω_1 and ω_2 are positive but this is the interesting case when the Hamiltonian is not positive definite. By (ii), H_{2k}, $1 \leq k \leq N$, depends only

on I_1 and I_2; so, H is assumed to be in Birkhoff normal form (Corollary VII.C.2) through terms of degree $2N$. This usually requires the nonresonance condition $k_1\omega_1 + k_2\omega_2 \neq 0$ for all integers k_1, k_2 with $|k_1| + |k_2| \leq 2N$, but it is enough to assume that H is in this normal form.

Theorem 1 (Arnold's Stability Theorem). *The origin is stable for the system whose Hamiltonian is* (1), *provided for some* k, $0 \leq k \leq N$, $D_{2k} = H_{2k}(\omega_2, \omega_1) \neq 0$ *or, equivalently, provided* H_2 *does not divide* H_{2k}.

Moreover, arbitrarily close to the origin in $\mathbb{R}^4$, *there are invariant tori and the flow on these invariant tori is the linear flow with irrational slope as discussed in Section* I.B.5.

PROOF. Assume that $D_2 = \cdots = D_{2N-2} = 0$ but $D_{2N} \neq 0$; so, there exist homogeneous polynomials F_{2k}, $k = 2, \ldots, N-1$, of degree $2k$ such that $H_{2k} = H_2 F_{2k-2}$. The Hamiltonian (1) is then

$$H = H_2(1 + F_2 + \cdots + F_{2N-2}) + H_{2N} + H^{\dagger}. \tag{2}$$

Introduce action-angle variables $I_i = (x_i^2 + y_i^2)/2$, $\phi_i = \arctan(y_i/x_i)$, and scale the variables by $I_i = \varepsilon^2 J$, where ε is a small-scale variable. This is a symplectic change of coordinates with multiplier ε^{-2}; so, the Hamiltonian (2) becomes

$$H = H_2 F + \varepsilon^{2N-2} H_{2N} + O(\varepsilon^{2N+1}), \tag{3}$$

where $F = 1 + \varepsilon^2 F_2 + \cdots + \varepsilon^{2N-4} F_{2N-4}$.

Fix a bounded neighborhood of the origin, say $|J_i| \leq 4$, and call it O so that the remainder term is uniformly $O(\varepsilon^{2N+1})$ in O. Restrict your attention to this neighborhood henceforth. Let h be a new parameter which will lie in the bounded interval $[-1, 1]$. Since $F = 1 + \cdots$, one has

$$H - \varepsilon^{2N-1} h = KF, \tag{4}$$

where

$$K = H_2 + \varepsilon^{2N-2} H_{2N} + O(\varepsilon^{2N+1}). \tag{5}$$

Since $F = 1 + \cdots$, the function F is positive on O for sufficiently small ε so the level set when $H = \varepsilon^{2N-1} h$ is the same as the level set when $K = 0$. Let $z = (J_1, J_2, \phi_1, \phi_2)$, and let ∇ be the gradient operator with respect to these variables. The equations of motion are

$$\dot{z} = J\nabla H = (J\nabla K)F + K(J\nabla F). \tag{6}$$

On the level set when $K = 0$, the equations become

$$\dot{z} = J\nabla H = (J\nabla K)F. \tag{7}$$

For small ε, F is positive; so, reparameterize Equations (8) by $d\tau = F\, dt$, and Equations (7) become

$$z' = J\nabla K(z), \tag{8}$$

where the prime denotes $d/d\tau$.

In summary, it has been shown that in O for small ε, the flow defined by H on the level set $H = \varepsilon^{2N-1}h$ is a reparameterization of the flow defined by K on the level set $K = 0$. Thus, it suffices to consider the flow defined by K. To that end, the equations of motion defined by K are

$$\begin{aligned} J_i' &= O(\varepsilon^{2N-1}), \\ \phi_1' &= -\omega_1 - \varepsilon^{2N-2}\frac{\partial H_{2N}}{\partial J_1} + O(\varepsilon^{2N-1}), \\ \phi_2' &= +\omega_2 - \varepsilon^{2N-2}\frac{\partial H_{2N}}{\partial J_2} + O(\varepsilon^{2N-1}). \end{aligned} \tag{9}$$

From these equations, the Poincaré map of the section $\phi_2 \equiv 0 \bmod 2\pi$ in the level set $K = 0$ will be computed, and then the invariant curve theorem can be applied.

From the last equation in (9), the first return time T required for ϕ_2 to increase by 2π is given by

$$T = \frac{2\pi}{\omega_2}\left(1 + \frac{\varepsilon^{2N-2}}{\omega_2}\frac{\partial H_{2N}}{\partial J_2}\right) + O(\varepsilon^{2N-1}). \tag{10}$$

Integrate the ϕ_1 equation in (9) from $\tau = 0$ to $\tau = T$, and let $\phi_1(0) = \phi_0$, $\phi_1(T) = \phi^*$ to get

$$\begin{aligned} \phi^* &= \phi_0 + \left(-\omega_1 - \varepsilon^{2N-2}\frac{\partial H_{2N}}{\partial J_1}\right)T + O(\varepsilon^{2N-1}), \\ \phi^* &= \phi_0 - 2\pi\left(\frac{\omega_1}{\omega_2}\right) - \varepsilon^{2N-2}\left(\frac{2\pi}{\omega_2^2}\right)\left(\omega_2\frac{\partial H_{2N}}{\partial J_1} + \omega_1\frac{\partial H_{2N}}{\partial J_2}\right) + O(\varepsilon^{2N-1}). \end{aligned} \tag{11}$$

In the above, the partial derivatives are evaluated at (J_1, J_2). From the relation $K = 0$, solve for J_2 to get $J_2 = (\omega_1/\omega_2)J_1 + O(\varepsilon^2)$. Substitute this into (11) to eliminate J_2, and simplify the expression by using Euler's theorem on homogeneous polynomials to get

$$\phi^* = \phi_0 + \alpha + \varepsilon^{2N-2}\beta J_1^{N-1} + O(\varepsilon^{2N-1}), \tag{12}$$

where $\alpha = -2\pi(\omega_1/\omega_2)$ and $\beta = -2\pi(N/\omega_2^{N+1})H_{2N}(\omega_2, \omega_1)$. By assumption, $D_{2N} = H_{2N}(\omega_2, \omega_1) \neq 0$; so, $\beta \neq 0$. Along with (12), the equation $J_1 \to J_1 + O(\varepsilon^{2N-1})$ defines an area-preserving map of an annular region, say $\frac{1}{2} \leq J_1 \leq 3$ for small ε. By the invariant curve theorem (Theorem B.2) for sufficiently small ε, $0 \leq \varepsilon \leq \varepsilon_0$, there is an invariant curve for this Poincaré map of the form $J_2 = \rho(\phi_1)$, where ρ is continuous, 2π periodic, and $\frac{1}{2} \leq \rho(\phi_1, \varepsilon) \leq 3$ for all ϕ_1. For all ε, $0 \leq \varepsilon \leq \varepsilon_0$, the solutions of (9) which start on $K = 0$ with initial condition $J_1 < \frac{1}{2}$ must have J_1 remaining less than 3 for

all τ. Since on $K = 0$ on has that $J_2 = (\omega_1/\omega_2)J_1 + \cdots$, a bound on J_1 implies a bound on J_2. Thus, there are constants c and k such that if $J_1(\tau)$, $J_2(\tau)$ satisfy Equations (9), start on $K = 0$, and satisfy $|J_i(0)| \leq c$, then $|J_i(\tau)| \leq k$ for all τ and for all $h \in [-1, 1]$, $0 \leq \varepsilon \leq \varepsilon_0$.

Going back to the original variables, $(I_1, I_2, \phi_1, \phi_2)$, and the original Hamiltonian H, this means that for $0 \leq \varepsilon \leq \varepsilon_0$, all solutions of the equations defined by the Hamiltonian (1) which start on $H = \varepsilon^{2N-1}h$ and satisfy $|I_i(0)| \leq \varepsilon^2 c$ must satisfy $|I_i(t)| \leq \varepsilon^2 k$ for all t and all $h \in [-1, 1]$, $0 \leq \varepsilon \leq \varepsilon_0$. Thus, the origin is stable. The invariant curves in the section map sweep out an invariant torus under the flow. ∎

F. Stability of $\mathscr{L}_4$

The coefficients A, B, and C of Arnold's theorem for the Hamiltonian of the restricted 3-body problem were computed by Deprit and Deprit-Bartholomê (1967) specifically to apply Arnold's theorem. These coefficients were given in Section VII.D. For $0 < \mu < \mu_1, \mu \neq \mu_2, \mu_3$ they found

$$D_4 = -\frac{36 - 541\omega_1^2\omega_2^2 + 644\omega_1^4\omega_2^4}{8(1 - 4\omega_1^2\omega_2^2)(4 - 25\omega_1^2\omega_2^2)}. \tag{1}$$

which is nonzero except for one value $\mu_c \approx 0.010\,913\,667$ which seems to have no mathematical significance, is not a resonance value, and has no astronomical significance (it does not correspond to the earth–moon system, etc.).

In Meyer and Schmidt (1986b), the normalization was carried to sixth order using an algebraic processor, and $D_6 = P/Q$, where

$$\begin{aligned} P = &-\frac{3105}{4} + \frac{1338449}{48}\sigma - \frac{48991830}{1728}\sigma^2 + \frac{7787081027}{6912}\sigma^3 \\ &- \frac{2052731645}{1296}\sigma^4 - \frac{1629138643}{324}\sigma^5 \\ &\frac{1879982900}{81}\sigma^6 + \frac{368284375}{81}\sigma^7, \\ Q = &\ \omega_1\omega_2(\omega_1^2 - \omega_2^2)^5(4 - 25\sigma)^3(9 - 100\sigma), \\ \sigma = &\ \omega_1^2\omega_2^2. \end{aligned} \tag{2}$$

From this expression, $D_6 \approx 66.6$ when $\mu = \mu_c$. So by Arnold's theorem and these calculations, $\mathscr{L}_4$ is stable for $0 < \mu < \mu_1, \mu \neq \mu_2, \mu_3$. $\mathscr{L}_4$ is unstable by the results in Markeev (1966) and Alfriend (1970, 1971) when $\mu = \mu_2, \mu_3$. $\mathscr{L}_4$ is stable when $\mu = \mu_1$ by Sokol'skij (1978). This last result is interesting because the linearized system is not simple, and so the linearized equations are unstable.

G. Further Reading

There are partial higher-dimensional generalizations of both the invariant curve theorem and Arnold's theorem. Consider for example a Hamiltonian system of n degrees of freedom of the form

$$H = \sum_{i=1}^{n} \omega_i I_i + \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} I_i I_j + H^{\dagger}(I_1, \ldots, I_n, \phi_1, \ldots, \phi_n), \tag{1}$$

where $(I_1, \ldots, I_n, \phi_1, \ldots, \phi_n)$ are action-angle variables, and where $H^{\dagger}$ is smooth and of higher order. Assume for the moment that $H^{\dagger} = 0$, then the equations are

$$\dot{I}_i = 0, \qquad \dot{\phi}_i = \omega_i + \sum_{j=1}^{n} B_{ij} I_j, \tag{2}$$

and the solutions are

$$I_i = a_i, \qquad \dot{\phi}_i = b_i + \left(\omega_i + \sum_{j=1}^{n} B_{ij} I_j\right) t, \tag{3}$$

where the a_i's and b_i's are constants. The I_i's are integrals. The integral surfaces, the sets defined by $I_i = a_i$, are n-dimensional tori, and the flow on these tori is linear. If $\det(B_{ij}) \neq 0$, then the frequencies of the linear flows on the tori vary as you move from torus to torus.

The generalization of Arnold's theorem states that if $\det(B_{ij}) \neq 0$, then many of these invariant tori persist even when $H^{\dagger}$ is small. However, these tori are n-dimensional manifolds in the $2n - 1$ energy surface $H =$ constant. An n-torus does not separate in a $2n - 1$ space unless $n = 2$. Thus, for higher degrees of freedom, the tori do not trap the solutions near the origin, and so, they do not ensure stability.

The best general discussion of the KAM theory is in the two monographs Moser (1968, 1973). Both are excellent and well worth reading. A complete and clean proof of the invariant curve theorem can be found in Siegel and Moser (1971). For those interested in the fine details, see Herman (1983, 1986).

Problems

1. Let F be a diffeomorphism defined in a neighborhood O of the origin in $\mathbb{R}^m$, and let the origin be a fixed point for F. Let V be a smooth real-valued function defined on O, and define $\Delta V(x) = V(F(x)) - V(x)$.
   - **a.** Prove that if the origin is a minimum for V and $\Delta V(x) \leq 0$ on O, then the origin is a stable fixed point.
   - **b.** Prove that if the origin is a minimum for V and $\Delta V(x) < 0$ on $O \backslash \{0\}$, then the origin is an asymptotically stable fixed point.
   - **c.** Prove the analog of Chetaev's theorem.

2. Let $F(x) = Ax$ and $V(x) = x^T Sx$, where A and S are $n \times n$ matrices, and S is symmetric.

a. Show that $\Delta V(x) = x^T Rx$, where $R = A^T SA - S$.

b. Let $\mathscr{S}$ be the linear space on all $m \times m$ symmetric matrices and $\mathscr{L} = \mathscr{L}_A: \mathscr{S} \to \mathscr{S}$ be the linear map $\mathscr{L}(S) = A^T SA - S$. Show that $\mathscr{L}$ is invertible if and only if $\lambda_i \lambda_j \neq 1$ for all $i, j = 1, \ldots, m$, where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A. [Hint: First prove the result when $A = \text{diag}(\lambda_1, \ldots, \lambda_m)$. Then prove the result when $A = D + \varepsilon N$, where D is simple (diagonalizable), and N is nilpotent, $N^m = 0$, $SN = NS$, and ε is small. Use the Jordan canonical form theorem to show that A can be assumed to be $A = D + \varepsilon N$.]

c. Let A have all eigenvalues with absolute value less than 1. Show that $S = \sum_{i=1}^{\infty} (A^T)^i R A^T$ converges for any fixed R. Show S is symmetric if R is symmetric. Show S is positive definite if R is positive definite. Show that $\mathscr{L}(S) = R$; so, $\mathscr{L}^1$ has a specific formula when all the eigenvalues of A have absolute value less than 1.

3. Let $F(x) = Ax + f(x)$, where $f(0) = \partial f(0)/\partial x = 0$.

a. Show that if all the eigenvalues of A have absolute value less than 1, then the origin is asymptotically stable. [Hint: Use Problems 1b and 2c.]

b. Show that if A has one eigenvalue with absolute value greater than 1, then the origin is a positively unstable fixed point.

4. Let $r = 1$, $s = 0$, and $h(I) = \beta I$, $\beta \neq 0$ in formulas (B.5) of the invariant curve theorem.

a. Compute F^q, the qth iterate of F, to be of the form $(I, \phi) \to (I'', \phi'')$, where $I'' = I + O(\varepsilon), \phi'' = \phi + q\omega + q\beta I + O(\varepsilon)$.

b. Let $2\pi p/q$ be any number between $\omega + \beta a$ and $\omega + \beta b$, so $2\pi p/q = \omega + \beta I_0$ *where* $a < I_0 < b$. Show that there is a smooth curve $\Gamma_\varepsilon = \{(I, \phi): I = \Phi(\phi, \varepsilon) = I_0 + \cdots\}$ such that F^q moves points on Γ only in the radial direction, i.e., $\Phi(\phi)$ satisfies $\phi'' - \phi - 2\pi p = 0$. [Hint: Use the implicit function theorem.]

c. Show that since F^q is area preserving, $\Gamma \cap F^q(\Gamma)$ is nonempty, and the points of this intersection are fixed points of F^q or q-periodic points of F.

5. Using Poincaré elements show that the continuation of the circular orbits established in Section IV.C (Poincaré orbits) is of twist type and hence stable. Consider the various types of fixed points discussed in Section VIII.A.

a. Show that extremal points are unstable.

b. Show that 3-bifurcation points are unstable.

c. Show that k-bifurcation points are stable if $k \geq 5$.

d. Transitional and 4-bifurcation points can be stable or unstable depending on the case. Figure out which case is unstable. (The stability conditions are a little harder.)

CHAPTER X

Twist Maps and Invariant Curves

A. Introduction

In this chapter we study the dynamics of area-preserving (i.e., symplectic) monotone twist maps of the annulus. While seemingly quite special, we have already seen examples of such maps as time one maps of time periodic Hamiltonian systems of one degree of freedom, as Poincaré section maps for periodic orbits of Hamiltonian systems of two degrees of freedom (see Chapter V, Sections B and E). These maps also appear as dynamical systems in their own right (e.g., billiards on a convex table and one-dimensional crystals; see V.B).

The orbits of even algebraically simple monotone twist maps can be extremely complicated and a complete description of all the orbits for all but the simplest maps is too difficult a problem. Hence, we can choose one of two other, less difficult problems: either

(i) put qualitative restrictions on the behavior of all (or many) orbits (e.g., no orbit which starts in region A ever reaches region B), or
(ii) find examples of orbits which have various types of behavior (e.g., periodic orbits of various periods).

In this chapter we focus on problem (ii), particularly on finding periodic orbits, and then we show that the periodic orbits play a crucial role in the qualitative behavior of all orbits.

The study of periodic points for monotone twist maps was begun by Poincaré in his study of the restricted 3-body problem. A variety of techniques have been used since then to show the existence of periodic points. In this chapter, we will use techniques which take full advantage of the topology available (particularly the fact that we are working with maps of a *two*-dimensional annulus). The condition that the maps are symplectic (area

preserving) will be used only when absolutely necessary and will frequently be replaced by topological conditions. Other successful attacks on periodic orbits have used variational and analytic techniques (see notes in Section G for references). With a few notable exceptions, results for maps of two dimensions have not been extended to exact symplectic maps of higher-dimensional spaces. This is an active area of present research.

The exposition which follows owes a great deal to the work of Jungries, Gole, and particularly Boyland. It was the subject of classes taught at the University of Minnesota and the University of Cincinnati and the author (GRH) would like to thank those who attended these courses for their hospitality, suggestions, patience, and remarkable endurance.

In the next section we give the notation and definitions necessary for the chapter as well as some of the elementary propositions. Sections C and D are concerned with various fixed point theorems and the properties of periodic orbits. In Section E we discuss the relationship between periodic orbits, invariant sets and curves, and the KAM theorem.

B. Notation and Definitions

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the circle with unit circumference, i.e., $\mathbb{T}$ is the interval $[0, 1]$ with 1 and 0 identified. Let $\mathscr{A} = \mathbb{T} \times [0, 1]$ be the annulus and $\mathscr{C} = \mathbb{T} \times \mathbb{R}$ be the cylinder. We will be studying diffeomorphisms of $\mathscr{A}$ to itself and of $\mathscr{C}$ to itself; however, it will be easier to state the results if we have a global coordinate system (i.e., polar coordinates). So Let $A = \mathbb{R} \times [0, 1]$ be the strip. Then A is the universal cover of $\mathscr{A}$ with natural projection

$$\pi\colon A \to \mathscr{A} \tag{1}$$

which sends points (x, y) and $(x + r, y) \in A$ to the same point of $\mathscr{A}$ whenever $r \in \mathbb{Z}$. Similarly, $\mathbb{R}^2$ is the universal cover of $\mathscr{C}$ with natural projection

$$\pi\colon \mathbb{R}^2 \to \mathscr{C}. \tag{2}$$

We let X and Y denote the projections onto x and y coordinates respectively, i.e.,

$$\left.\begin{matrix} X \\ Y \end{matrix}\right\}\colon (x, y) \to \begin{cases} x \\ y \end{cases} \tag{3}$$

(the domain is either A or $\mathbb{R}^2$).

Then for any continuous map $\tilde{f}\colon \mathscr{A} \to \mathscr{A}$ (or $\tilde{f}\colon \mathscr{C} \to \mathscr{C}$) there exists a unique continuous map $f\colon A \to A$ (or $f\colon \mathbb{R}^2 \to \mathbb{R}^2$) such that

(i) $X(f(0, 0)) \in [0, 1)$,
(ii) $\pi \circ f = \tilde{f} \circ \pi$,

i.e., f is a particular lift of $\tilde{f}$ to the cover, or f in the polar coordinate re-

presentation of $\tilde{f}$. Conversely, if $f: A \to A$ (or $f: \mathbb{R}^2 \to \mathbb{R}^2$) satisfies $\forall (x, y)$, $f(x + 1, y) = f(x, y) + (1, 0)$, then there exists $\tilde{f}: \mathscr{A} \to \mathscr{A}$ (or $\tilde{f}: \mathscr{C} \to \mathscr{C}$) such that $\pi \circ f = \tilde{f} \circ \pi$.

Henceforth, we will state all results for maps from $A \to A$ or $\mathbb{R}^2 \to \mathbb{R}^2$, and we will assume the following restrictions: *All maps* $f: A \to A$ *or* $\mathbb{R}^2 \to \mathbb{R}^2$ *will be assumed to satisfy*:

(i) f is a C^1 diffeomorphism,
(ii) $X(f(0, 0)) \in [0, 1)$,
(iii) $\forall (x, y), f(x + 1, y) = f(x, y) + (1, 0)$,
(iv) f is orientation preserving,
(v) f is boundary component preserving.

Remarks. Condition (iii) is that f is the lift of a map on $\mathscr{A}$ or $\mathscr{C}$, and condition (v) is that $Y(f(x, i)) = i$ for $i = 0$ or 1 when $f: A \to A$ or $Y(f(x, y)) \to +\infty$ or $-\infty$ as $y \to +\infty$ or $-\infty$, respectively, when $f: \mathbb{R}^2 \to \mathbb{R}^2$.

Examples. (1) Let $g_0(x, y) = (x + y, y)$. This map makes sense on both A and $\mathbb{R}^2$ and we will use this notation in both situations.

(2) Let, for $k \in \mathbb{R}$, $g_k: \mathbb{R}^2 \to \mathbb{R}^2$ where

$$g_k(x, y) = \left(x + y + \frac{k}{2\pi} \sin(2\pi x),\ y + \frac{k}{2\pi} \sin(2\pi x) \right). \tag{4}$$

This is called *the standard family* of maps of the cylinder. It has been extensively studied both analytically and especially numerically (see Section G). We can replace $(k/2\pi) \sin(2\pi x)$ with any smooth function $\phi(x)$ satisfying $\forall x$, $\phi(x + 1) = \phi(x)$. The corresponding one-parameter family of maps is given by

$$(x, y) \to (x + y + k\phi(x),\ y + k\phi(x)) \tag{5}$$

and is sometimes called *a standard family*.

We next define conditions which "keep orbits in the annulus."

Definition. We say $f: A \to A$ (or $\mathbb{R}^2 \to \mathbb{R}^2$) is an *exact symplectic* map if f is symplectic with respect to the usual symplectic structure (i.e., symplectic form $\omega = dx \wedge dy$) and for any embedding $\gamma: \mathbb{R} \to A$ (or $\mathbb{R} \to \mathbb{R}^2$) satisfying $\forall x$, $\gamma(x + 1) = \gamma(x) + (1, 0)$ we have

$$\int_0^1 Y(\gamma(s)) \frac{d}{ds} X(\gamma(s))\, ds = \int_0^1 Y(f \circ \gamma(s)) \frac{d}{ds} X(f \circ \gamma(s))\, ds. \tag{6}$$

Remarks. (1) Since we are in two dimensions, the condition that f be symplectic is the same as requiring that f be area preserving, i.e., $|Df| \equiv 1$.

(2) The condition that f be exact symplectic adds to the area preservation a condition saying that the net area between a nontrivial loop on $\mathscr{C}$ and its

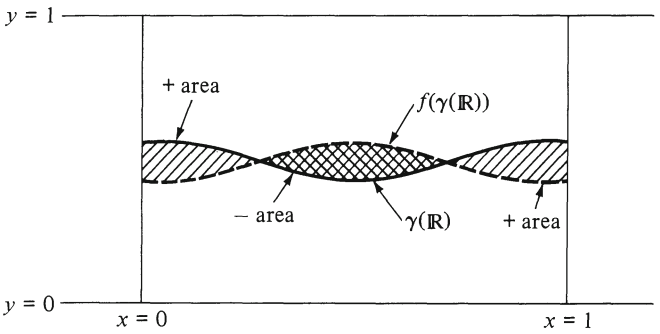

Figure B.1. Areas between $\gamma(\mathbb{R})$ and $f(\gamma(\mathbb{R}))$.

image under f is zero (see Figure B.1). For an area-preserving map $f: A \to A$, this condition will be satisfied automatically (see Problems).

In particular, the condition that f be exact symplectic is *not* satisfied by $f: \mathscr{C} \to \mathscr{C}$ given by $f(x, y) = (x, y + 1)$ even though f is area preserving (symplectic).

Sometimes it will be possible to weaken the condition that f be exact symplectic, replacing it with a topological condition like the following:

Definition. A map $f: A \to A$ is said to *satisfy condition B* if for every $\varepsilon > 0$ there exist $z_1, z_2 \in A$ and $n > 0$ such that

$$Y(z_1) < \varepsilon, \qquad Y(z_2) > 1 - \varepsilon \tag{7}$$

and

$$Y(f^n(z_1)) > 1 - \varepsilon, \qquad Y(f^n(z_2)) < \varepsilon. \tag{8}$$

A map $f: \mathbb{R}^2 \to \mathbb{R}^2$ is said to *satisfy condition B* if for every $M > 0$ there exist $z_1, z_2 \in A$ and $n > 0$ such that

$$Y(z_1) < -M, \qquad Y(z_2) > M \tag{9}$$

and

$$Y(f^n(z_1)) > M, \qquad Y(f^n(z_2)) < -M. \tag{10}$$

There is a wealth of possible dynamics for maps of two-dimensional spaces. Luckily, there is another condition satisfied by the maps that we wish to study that will give us a powerful tool for exploring the dynamics of these maps.

Definition. A map $f: A \to A$ (or $\mathbb{R}^2 \to \mathbb{R}^2$) is called a *monotone twist map* if there exists $\varepsilon > 0$ such that for all $(x, y) \in A$ [or $(x, y) \in \mathbb{R}^2$]

$$\frac{\partial X(f(x, y))}{\partial y} > \varepsilon. \tag{11}$$

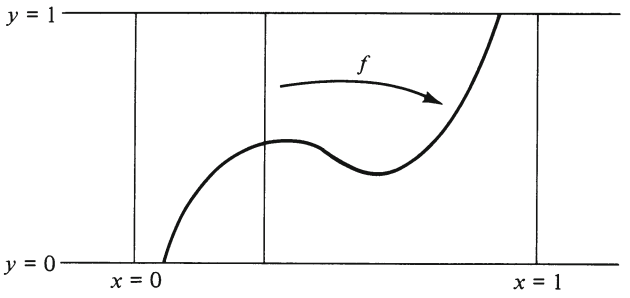

Figure B.2. Monotone twist condition.

Remarks. Geometrically, this condition states that the image of a segment $x =$ constant under f forms a graph over the x axis (see Figure B.2).

This condition can be expressed in a different way for exact symplectic maps, i.e., given $f: A \to A$, let $B = \{(x, x^1) \in \mathbb{R}^2: \{f(x, y): y \in : [0, 1]\} \cap \{(x^1, y): y \in [0, 1]\} \neq \phi\}\}$ then,

Theorem 1. *Given $f: A \to A$ is an exact symplectic map, f is a monotone twist map if and only if f has a generating function, $S: B \to \mathbb{R}$ such that*

$$f(x, y) = (x^1, y^1) \quad \text{if } y = \frac{-\partial S}{\partial x}(x, x^1), \quad y^1 = \frac{\partial S}{\partial x^1}(x, x^1). \tag{12}$$

Remarks. (1) f has a "locally defined" generating function is automatic (see Section IV.B.4), but that this function is defined on all of A is a strong restriction. There is a geometrical description of the generating function which we discuss in the Problems.

(2) The family of monotone twist maps is open in the C^1 topology, that is, any map sufficiently C^1 close to a monotone twist map is also a monotone twist map.

(3) Monotone twist maps are not closed under composition, i.e., if f and g are monotone twist maps, then $f \circ g$ might not be monotone twist. To get a family of maps closed under composition we need to consider "positive tilt" maps; see Boyland (1988).

Examples. (1) The standard family $g_k: \mathbb{R}^2 \to \mathbb{R}^2$ given above, and in fact any "standard family" of maps, are exact symplectic monotone twist maps as long as $\int_0^1 \varphi(x)\, dx = 0$.

(2) Let $H_0: A \to \mathbb{R}$ be given by $H_0(x, y) = \frac{1}{2}y^2$. Then the Hamiltonian system associated with H_0 is

$$\dot{x} = +\frac{\partial H_0}{\partial y} = y, \qquad \dot{y} = -\frac{\partial H_0}{\partial x} = 0 \tag{13}$$

and the time one map of this Hamiltonian system is $(x, y) \to (x + y, y)$, which is an exact symplectic monotone twist map. If we let $H_1: A \times \mathbb{R} \to \mathbb{R}$ be a smooth function which satisfies

(i) $\forall (x, y, t) \in A \times \mathbb{R}, H_1(x + 1, y, t) = H_1(x, y, t) = H_1(x, y, t + 1)$,
(ii) $\forall x, t \in \mathbb{R}, \partial H_1(x, 0, t)/\partial x = 0 = \partial H(x, 1, t)/\partial x$,
(iii) H_1 has the form $H_1(x, y, t) = \frac{1}{2}y^2 + P(x, y, t)$ where P is sufficiently C^2 small,

then the time one map of the flow given by the Hamiltonian system with H_1 as Hamiltonian will also be an exact symplectic monotone twist map. That the map is exact symplectic follows because the system is Hamiltonian (see Section IV.C). The monotone twist condition comes from the fact that this time one map will be C^1 close to the time one map of the H_0 system above. Knowing that $\partial^2 H_1/\partial y^2 > 0$ gives us an "infinitesimal"-twist condition, i.e., the map which follows the flow from time t to time $t + \Delta t$ will be a monotone twist map. However, this condition $\partial^2 H_1/\partial y^2 > 0$ does not imply the monotone twist condition for the time one map of the flow for the same reason that iterates of monotone twist maps need not be monotone twist maps.

The converse of the discussion above is also true, i.e.,

Theorem 2 (Moser 1986b). *Given an exact symplectic monotone twist map $f: A \to A$ there exists a Hamiltonian $H: A \times \mathbb{R} \to \mathbb{R}$ which satisfies*

(1) $\forall (x, y, t), H(x + 1, y, t) = H(x, y, t) = H(x, y, t + 1)$

and

(2) $\forall (x, y, t), \partial^2 H(x, y, t)/\partial y^2 > 0$,

such that f is the time one map of the Hamiltonian system given by H.

Remark. This is close to Theorem V.B.1, the new element being condition (2) on H, i.e., the infinitesimal twist condition or "Lagrange condition" which is useful in variational attacks on these systems. Also we note that an analogous discussion can be given for Hamiltonians on the cylinder and maps on $\mathbb{R}^2$.

Notation. Since we will be using the adjectives exact symplectic monotone twist a great deal, we will abbreviate them to ESMT.

Next we discuss the orbits of the maps we wish to study, introducing the following notation: If $f: A \to A$, then for $n > 0$, $f^n = f \circ \cdots \circ f$ (n times) and $f^{-n} = f^{-1} \circ \cdots \circ f^{-1}$ (n times). If $f: A \to A$ and $z \in A$, then the *extended orbit* of z under f is

$$\mathrm{eo}(z, f) = \mathrm{eo}(z) = \{f^i(z) + (j, 0): i, j \in \mathbb{Z}\}. \tag{14}$$

Remark. Since we think of a map $f: A \to A$ as a lift of a map $\tilde{f}: \mathscr{A} \to \mathscr{A}$, the extended orbit of $z \in A$ is the lift of the orbit of the projection of z, i.e.,

$$\mathrm{eo}(z) = \pi^{-1}\{\tilde{f}^i(\pi(z)): i \in \mathbb{Z}\}. \tag{15}$$

Since points translated by integers in the x direction are identified by $\pi: A \to \mathscr{A}$, to obtain the extended orbit of a point $z \in A$, we take all translates of the usual orbit by vectors $(j, 0)$ such that $j \in \mathbb{Z}$.

Similarly, we must extended the usual definition of periodic point:

Definition. For $f: A \to A$, a point $z \in A$ is called a *p/q-periodic point* if

$$f^q(z) = z + (p, 0). \tag{16}$$

Remarks. (1) If $\tilde{f}: \mathscr{A} \to \mathscr{A}$ is the projection of f to $\mathscr{A}$ and $\tilde{z} = \pi(z)$, then the statement that z is a p/q-periodic point of f implies that z is a period q periodic point of $\tilde{f}$ because

$$\tilde{f}^q(\tilde{z}) = \tilde{f}^q(\pi(z)) = \pi \circ f^q(z) = \pi(z + (p, 0)) = \tilde{z}. \tag{17}$$

The p in the definition of p/q-periodic point is therefore new information. It says that the q iterates of $\tilde{z}$ under $\tilde{f}$ "go around" the annulus p times. This notion of circulation can only be made well defined by going to the lift $f: A \to A$. [See Peckham (1989).]

(2) We note that p/q periodic point of f is also a $2p/2q$ periodic point because

$$f^{2q}(z) = f^q \circ f^q(z) = f^q(z + (p, 0)) = f^q(z) + (p, 0) = z + (2p, 0). \tag{18}$$

but a $2p/2q$-periodic point need *not* be a p/q-periodic point. Hence, we make the following assumption:

Notation. When we write z is a p/q-periodic point we will assume, unless otherwise stated, that p and q are relatively prime.

The notion of average "rotation per iteration" for periodic orbits can be generalized as follows:

Definition. If $f: A \to A$ and $z \in A$, then the *rotation number* of z is

$$\rho(z, f) = \rho(z) = \lim_{n \to \infty} \frac{X(f^n(z))}{n}, \quad \text{if it exists.} \tag{19}$$

If the limit does not exist, then we say $\rho(z)$ does not exist.

Examples. (1) For $f: A \to A$ and $z \in A$, then the a p/q-periodic point of f, $\rho(z) = p/q$.

(2) For $g_0: A \to A: (x, y) \to (x + y, y)$ we have $\rho((x, y)) = y$ for all $(x, y) \in A$.

(3) For $f: A \to A$, the map f restricted to the boundary components of A will give maps of $\mathbb{R}$ which are lifts of circle homeomorphisms, i.e., if we let h_0, $h_1: \mathbb{R} \to \mathbb{R}$ be defined by

$$h_i(x) = X(f(x, i)) \qquad \text{for } i = 0 \text{ or } 1. \tag{20}$$

Then $i = 0, 1$ $\forall x \in \mathbb{R}$, $h_i(x + 1) = h_i(x) + 1$ and h_i is a homeomorphism. It is not too difficult to show [see Devaney (1986), Nitecki (1971), Coddington and Levinson (1955), or Lemma 4, where this proof is given in a different context] that $\lim_{n\to\infty} h_i^n(x)/n$ exists and is independent of x.

Notation. For $f: A \to A$ we let $\rho_0 = \rho((x, 0))$ and $\rho_1 = \rho((x, 1))$. These limits exist and are independent of x by the above.

The boundary circles of A are examples of invariant circles for maps $f: A \to A$.

Definition. Suppose $\gamma: \mathbb{R} \to A$ is a continuous, one-to-one embedding satisfying $\gamma(x + 1) = \gamma(x) + (1, 0)$ for all $x \in \mathbb{R}$. Then we say that the set $\Gamma = \gamma(\mathbb{R})$ is an *invariant circle* for $f: A \to A$ if $f(\Gamma) = \Gamma$.

Remark. Hence, an invariant circle for $f: A \to A$ is a curve which is invariant under f and which projects to a *homotopically nontrivial* loop in the annulus A. (See Figure B.3.) Hence, an invariant circle separates A into two components, one for each boundary component of A. The existence of invariant circles is one of the fundamental problems in the study of monotone twist maps and we will return to it at the end of the chapter. For now we note:

Proposition 3. *If $f: A \to A$ and Γ is an invariant circle for f, then for each $z \in \Gamma$, $\rho(z)$ exists and is independent of Γ.*

PROOF. Again $f|_\Gamma$ gives a circle homeomorphism and the techniques referenced above can be applied. ∎

We will construct many examples of periodic orbits with different rotation numbers in the following sections. It is natural to hope that we could take limits of these periodic points to obtain points of arbitrary rotation number. However, rotation number is also a limit, so there is no reason that the limit of the rotation number of a sequence of points is the rotation number of the

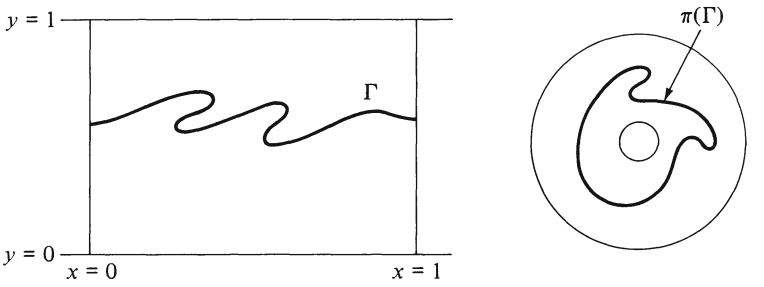

Figure B.3. Invariant curves.

limit, i.e., no reason we should be able to "switch limits." We will need to find special orbits which behave nicely with respect to limits, which motivates the following:

Definition. Suppose $f: A \to A$ is a monotone twist map and $z \in A$. Then z is called a *monotone* point or is said to have a *monotone orbit* if $\forall z_1, z_2 \in \mathrm{eo}(z)$, if $X(z_1) < X(z_2)$, then $X(f(z_1)) < X(f(z_0))$.

Remarks. (1) Hence, a monotone orbit is one on which f preserves the ordering on the extended orbit imposed by the natural ordering of the x coordinate. We will see later that another way to say this is that the orbit of a monotone point can be extended in a natural way to a homeomorphism of a circle.

(2) Clearly the definition of monotone point makes sense for arbitary annulus maps but the notion is not very useful without the monotone twist condition because the following lemmas are not true without it. These lemmas state that the set of monotone orbits are in various ways isolated from other orbits, that is, the property that makes it possible to prove their existence and that makes them useful.

Lemma 4. *Suppose $f: A \to A$ is a monotone twist map and $z_0 \in A$ is a monotone point for f. Then $\rho(z_0)$ exists.*

PROOF. We give a direct proof of this fact that follows exactly the outline of the proof of existence of rotation numbers for circle homeomorphisms. We suppose $X(z_0) \in (0, 1)$.

For any $n > 0$, fix $r \in \mathbb{Z}$ so that

$$X(z_0) + r \leq X(f^n(z_0)) < r + 1 + X(z_0). \tag{21}$$

Suppose we have that $X(f^{n(m-1)}(z_0)) < (m-1)(r+1) + X(z_0)$, then

$$X(f^{nm}(z_0)) = X(f^{n(m-1)}(f^n(z_0))). \tag{22}$$

Because z_0 is a monotone point and $f^n(z_0), z_0 + (r+1, 0) \in \mathrm{eo}(z_0)$, we have

$$X(f^{n(m-1)}(f^n(z_0)) \leq X(f^{n(m-1)}(z_0 + (r+1, 0))) \tag{23}$$

so

$$X(f^{nm}(z_0)) \leq X(z_0) + m(r+1). \tag{24}$$

By induction we conclude that $X(f^{nm}(z_0)) < m(r+1)$ for all $m > 0$. Similarly, we can show $X(f^{nm}(z_0)) \geq m\,r$ for all $m > 0$. Hence,

$$n\frac{m}{m}r \leq \frac{X(f^{nm}(z_0))}{nm} \leq \frac{m}{nm}(r+1) \quad \text{for all } m > 0. \tag{25}$$

Letting m tend to infinity we see that

$$\left| \lim_{m\to\infty} \sup \frac{X(f^{nm}(z_0))}{nm} - \lim_{m\to\infty} \inf \frac{X(f^{n(m)}(z_0))}{nm} \right| < \frac{1}{n}. \tag{26}$$

Next we note that since f is periodic in the x coordinate, for each $n > 0$ there exists a constant $C_n > 0$ such that

$$\forall i = 1, \ldots, n, \qquad \forall z \in A, \qquad |Xf^i(z) - X(z)| < C_n, \tag{27}$$

but then $|X(f^{nm+i}(z_0)) - X(f^{nm}(z_0))| < C_n$. Hence

$$\limsup_{i\to\infty} \frac{X(f^i(z_0))}{i} = \limsup_{m\to\infty} \frac{X(f^{nm}(z_0))}{nm} \tag{28}$$

and

$$\liminf_{i\to\infty} \frac{X(f^i(z_0))}{i} = \liminf_{m\to\infty} \frac{X(f^{nm}(z_0))}{nm} \tag{29}$$

i.e.,

$$\left| \limsup_{i\to\infty} \frac{Xf^i(z_0)}{i} - \liminf_{i\to\infty} \frac{Xf^i(z_0)}{i} \right| < \frac{1}{n}. \tag{30}$$

But n was arbitary. Hence $\lim_{n\to\infty} f^n(z_0)/n$ exists. ∎

Lemma 5. *Suppose $f_n: A \to A$ is a sequence of monotone twist maps $n = 0, 1, \ldots$ and $\lim_{n\to\infty} f_n = f_0$ in the C^1 topology. Suppose for each $n = 1, 2, \ldots$ there is a point $z_n \in A$ such that $X(z_n) \in [0, 1]$ and z_n has a monotone point for f_n. Suppose $z_0 = \lim_{n\to\infty} z_n$. Then z_0 is a monotone point for f_0 and $\rho(z_0, f_0) = \lim_{n\to\infty} \rho(z_n, f_n)$.*

PROOF. Suppose z_0 is not monotone for f_0. Then there exist i, j, k, and ℓ such that

$$X(f_0^i(z_0)) + k < X(f_0^j(z_0)) + \ell \quad \text{but} \quad X(f_0^{i+1}(z_0)) + k \geq X(f_0^{j+1}(z_0)) + \ell. \tag{31}$$

For n sufficiently large, we must have $X(f_n^i(z_n)) + k < X(f_n^j(z_n)) + \ell$ so $X(f_n^{i+1}(z_n)) + k < X(f_n^{j+1}(z_n)) + \ell$. Hence, by taking the limit as $n \to \infty$ we see that (31) must give

$$X(f_0^{i+1}(z_0)) + k = X(f_0^{j+1}(z_0)) + \ell. \tag{32}$$

From the monotone twist condition it follows (see Figure B.4) that

$$Y(f_0^{i+1}(z_0)) > Y(f_0^{j+1}(z_0)). \tag{33}$$

Hence, again by the monotone twist condition

$$X(f_0^{i+2}(z_0)) + k > X(f_0^{j+2}(z_0)) + \ell. \tag{34}$$

Again this implies that for n sufficiently large

$$X(f_n^{i+2}(z_n)) + k > X(f_n^{j+2}(z_n)) + \ell, \tag{35}$$

contradicting that z_n is monotone for f_n. Hence, z_0 must be a monotone point for f_0.

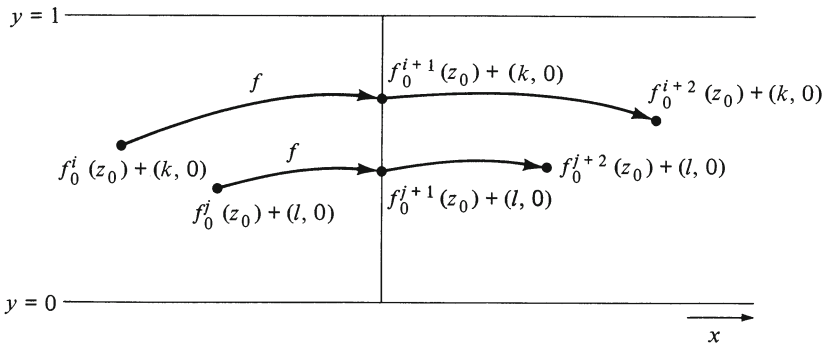

Figure B.4. Nonmonotone orbits.

From Lemma 5 we know that $\rho(z_n, f_n)$ exists for each $n = 0, 1, \ldots$. Moreover, as in the proof of Lemma 5

$$r \leq X(f_n^i(z_n)) < r + 1 \tag{36}$$

for i, $r \in \mathbb{Z}$ implies $\rho(z_n, f_n) \in [r/i, (r+1)/i]$. Hence, if we note that $r \leq X(f_0^i(z_0)) \leq r + 1$ implies that for n sufficiently large

$$r - 1 \leq X(f_n^i(z_n)) \leq r + 2, \tag{37}$$

then we see that $\lim_{n\to\infty} \rho(z_n, f_n) = \rho(z_0, f_0)$. ∎

Lemma 6. *Suppose f_n: $A \to A$, $n = 0, 1, \ldots$, is a sequence of monotone twist maps with $f_n \to f_0$ in the C^1 topology as $n \to \infty$. Fix $p, q \in \mathbb{Z}$ (p and q relatively prime) and suppose that for each $n = 1, 2, \ldots$ there is point $z_n \in A$ with z_n a p/q-periodic point for f_n. If $z_0 = \lim_{n\to\infty} z_n$, then z_0 will be a p/q-periodic point for f_0. Moreover, either*

(1) *for all n sufficiently large, z_n is monotone for f_n and hence z_0 is monotone for f_0,*

or

(2) *for all n sufficiently large, z_n is not monotone for f_n and hence z_0 is not monotone for f_0.*

PROOF. First we note that since we have for all $n \geq 1$, $f_n^q(z_n) = z_i + (p, 0)$, taking limits of both sides we have $f_0^q(z_0) = z_0 + (p, 0)$, i.e., z_0 is a p/q-periodic point for f_0. (Since p and q are relatively prime, z_0 cannot have a smaller period.)

If there exists a subsequence $z_{n_i} \to z_0$ with each z_{n_i} monotone for f_{n_i}, then z_0 will be monotone for f_0 by Lemma 5.

On the other hand, suppose $z_{n_k} \to z_0$ is a subsequence such that z_{n_k} is nonmonotone for f_{n_k}. Then for each n_k there exist i, j, and ℓ such that

$$X(f_{n_k}^i(z_{n_k})) < X(f_{n_k}^j(z_{n_k})) + \ell, \tag{38}$$

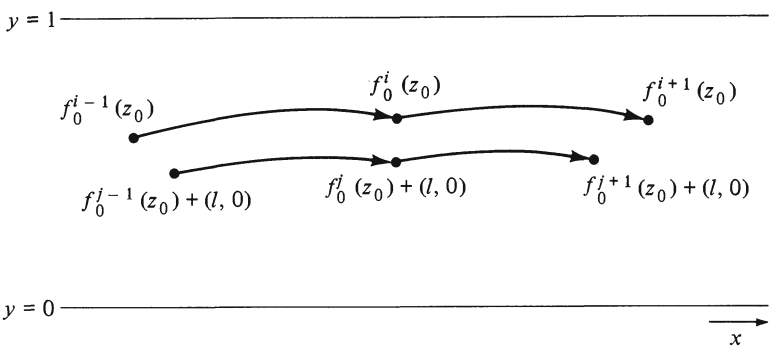

Figure B.5. Nonmonotone orbits.

but

$$X(f_{n_k}^{i+1}(z_{n_k})) \geq X(f_{n_k}^{j+1}(z_{n_k})) + \ell. \tag{39}$$

Since each z_{n_k} is a p/q-periodic point, we may assume that $0 \leq i, j \leq q$, and $0 \leq \ell \leq p$. Hence, we may choose another subsequence which we again call $z_{n_k} \to z_0$, such that i, j, and ℓ are independent of z_{n_k}. But then z_0 must satisfy

$$X(f_0^i(z_0)) \leq X(f_0^j(z_0)) + \ell, \tag{40}$$

$$X(f_0^{i+1}(z_0)) \geq X(f_0^{j+1}(z_0)) + \ell. \tag{41}$$

If strict inequality holds in (40) then z_0 is not monotone. If equality holds in (40) then we note that $f_0^i(z_0) \neq f_0^j(z_0) + (\ell, 0)$ since z_0 is a p/q-periodic point and that in order to satisfy (41) we must have $Y(f_0^i(z_0)) > Y(f^j(z_0))$. But then we have $X(f_0^{i-1}(z_0)) < X(f_0^{j-1}(z_0)) + \ell$ and again we see that z_0 is nonmonotone (see Figure B.5).

Hence we see that if z_n has a subsequence which is monotone, then z_0 is monotone, whereas if it has a subsequence that is nonmonotone, then z_0 is nonmonotone. So for all n sufficiently large, we must have z_n always monotone or z_n always nonmonotone and the proof is complete. ∎

Remarks. (1) If we think of the p/q-periodic orbits of a given monotone twist map as a set with a natural topology, then Lemma 6 says that the whole set is closed, as well as the subsets forming monotone and nonmonotone orbits. This says that the p/q monotone periodic points are isolated from the others and, hence, one can hope to use topological techniques to find them.

(2) Another, more concise, way of stating the closure properties above is to put the Hausdorff topology on the closed subsets of A. Then the set of the closures of monotone orbits for a given monotone twist map is a closed set in this Hausdorff topology [see Katok (1982)].

In the following sections we will prove the existence of many periodic points for exact symplectic monotone twist (ESMT) maps, particularly mono-

tone periodic points which behave well under limits. Then we will discuss the relation between these periodic points and the general dynamics of the map.

C. Existence of Periodic Orbits

The result known as Poincaré's Last Geometric Theorem implies that every exact symplectic monotone twist (ESMT) map $f: A \to A$ will have two distinct periodic orbits for each rational between the rotation numbers of f on the boundary components of A. This theorem was originally conjectured by Poincaré in 1912 (with a weaker twist condition than monotone twist) and was proven by Birkhoff (1913, 1925) and Brown and von Newmann (1977). Present proofs using more machinery from plane topology have greatly weakened both the area-preservation and twist hypotheses necessary and the interested reader should consult the work of John Franks (1988).

The theorem we will use is the following:

Theorem 1. *Suppose $f: A \to A$ is an exact symplectic monotone twist* (ESMT) *map with ρ_0 and ρ_1 the rotation numbers of f on $y = 0$ and $y = 1$ boundaries respectively (see Proposition B.3). If $p/q \in \mathbb{Q}$ is a rational (in lowest form) with $\rho_0 \leq p/q \leq \rho_1$, then f has at least two distinct extended orbits which are p/q-periodic.*

Remarks. (1) So the theorem implies that the projection of f to the annulus $\mathscr{A}$ has two distinct p/q-periodic orbits.

(2) A similar statement holds for ESMT maps of the cylinder with no restrictions on the size of the rational.

The remainder of this section will be devoted to a discussion of the proof of Theorem 1. The existence of p/q-periodic points is not so difficult. The proof we give will take advantage of the monotone twist condition and is due to LeCalvez (1988) and Casdagli (1987). (Again we note that weaker twist conditions suffice and refer the reader to the references of the first paragraph of this section.) That there are actually more than one p/q-periodic extended orbit for f is considerably more subtle. We will discuss the plausibility of the existence of two periodic orbits in one case (when q is odd).

Proof of Existence of p/q-Periodic Points. Fix f and $p/q \in \mathbb{Q}$ as in the theorem. We will need the following notation

$$\Sigma = \{z \in A : X(f^q(z)) = X(z) + p\}.$$

Let U_1 be the component of $A \sim \Sigma$ containing the $y = 0$ boundary of A and let V be the component of $A \sim$ closure (U_1) containing the $y = 1$ boundary of A. Finally, let $U = A \sim$ closure (V). Then U is open, $\partial U \subseteq \Sigma$, U is simply

connected $U + (1, 0) = \{z + (1, 0): z \in U\} = U$, and contains the $y = 0$ boundary of A. Let $\Gamma = \partial U$.

Claim. $f^{-1}(\Gamma) \cap \Gamma \neq \varnothing$.

PROOF OF CLAIM. Suppose not. Then $f^{-1}(\Gamma) \subseteq U$ or $f^{-1}(\Gamma) \subseteq A \sim (\text{closure } U)$, so either f^{-1} (closure U) $\subseteq U$ or closure $U \subseteq f^{-1}(U)$. But both of these cases violate the assumption that f is exact symplectic (i.e., area preserving). Hence, $f^{-1}(\Gamma) \cap \Gamma \neq \varnothing$ and the proof of the claim is complete. ∎

Claim. Every point $z \in f^{-1}(\Gamma) \cap \Gamma$ is a p/q-periodic point for f.

PROOF OF CLAIM. Suppose $z \in f^{-1}(\Gamma) \cap \Gamma$, then $z \in \Gamma$ and $f(z) \in \Gamma$ so $X(f^q(z)) = X(z) + p$ and $X(f^{q+1}(z)) = X(f(z)) + p$. Because f is a monotone twist we know that there is a unique point on the segment $\{(x, y): x = X(z) + p, 0 \leq y \leq 1\}$ such that $f(x, y) \in \{(x, y): x = X(f(z)) + p\}$, but $f^q(z) \in \{(x, y): x = X(z) + p, \quad 0 \leq y \leq 1\}$ and $f^{q+1}(z) \in \{(x, y): x = (X(f(z)) + p, 0 \leq y \leq 1\}$, so $f^q(z)$ is this unique point. However, because f is a lift of an annulus diffeomorphism, $z + (p, 0) \in \{(x, y): x = X(z) + p, 0 \leq y \leq 1\}$ has image $f(z + (p, 0)) = f(z) + (p, 0)$ in $\{(x, y): x = X(f(z)) + p, 0 \leq y \leq 1\}$ (see Figure C.1). Hence, $z + (p, 0)$ and $f^q(z)$ must be the same point, i.e., z is a p/q-periodic point and the proof of the claim is complete. ∎

Combining the claims, the proof of existence of p/q-periodic points for f is complete. ∎

PLAUSABILITY OF EXISTENCE OF TWO p/q-PERIODIC ORBITS (q ODD). The idea is to show that the points of intersection of $f^{-1}(\Gamma)$ with Γ come in different types and that there must typically be an even number of such intersections.

First, we may assume that the intersection points of $f^{-1}(\Gamma)$ with Γ are isolated because if they were not we would have infinitely many distinct p/q-periodic extended orbits.

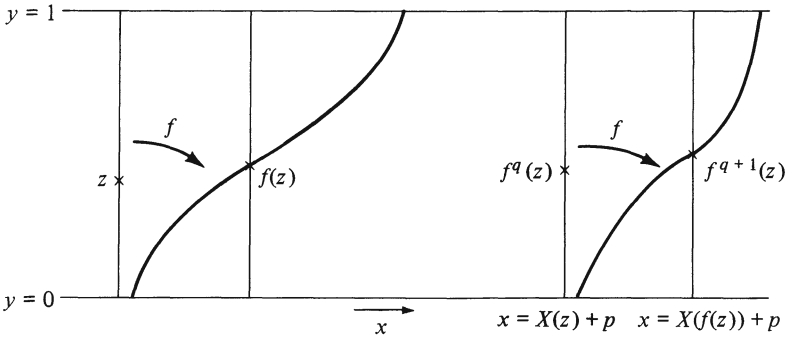

Figure C.1. Images of radial arcs through z and $f^q(z)$.

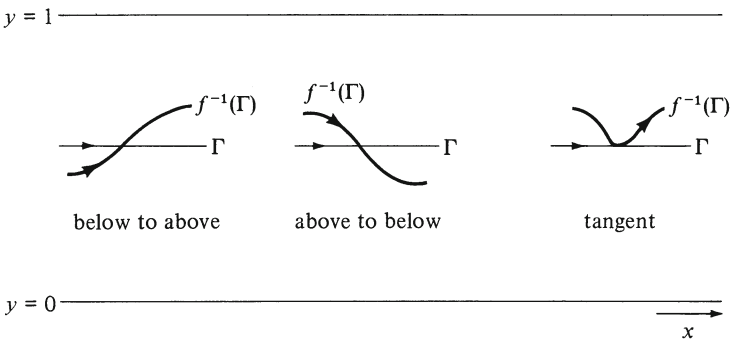

Figure C.2. Applications of monotone twist condition.

Unfortunately, there is no reason to assume that Σ will be a smooth curve in A, i.e., that 0 is a regular value of the function: $A \to \mathbb{R}$: $z \to X(f^q(z)) - X(z) - p$. However, if Σ is a smooth curve, then the proof can be easily completed as follows: If $\Gamma \subseteq \Sigma$ is a smooth curve which separates A, then $f^{-1}(\Gamma) \cap \Gamma$ has types of intersection points—those where $f^{-1}(\Gamma)$ goes from below (component of $A \sim \Gamma$ containing $y = 0$) to above Γ where $f^{-1}(\Gamma)$ is oriented from left ($x = -\infty$) to right and where $f^{-1}(\Gamma)$ goes from above to below Γ, or the intersection could be a tangency (see Figure C.2).

Now the intersections of $f^{-1}(\Gamma)$ with Γ which are tangent may be ignored because they are not structurally stable (a small perturbation of f removes them). Hence we assume that $f^{-1}(\Gamma) \cap \Gamma$ has both "below-to-above" and "above-to-below" intersections. Since f^{-1} is area preserving, both types of intersections occur. Since f^{-1} projects to a map of the annulus and Γ projects to a loop on the annulus, there must be an equal number of "below-to-above" and "above-to-below" intersections. Hence f has an even number of p/q-periodic points. If q is odd this means f has at least two p/q-periodic orbits.

To complete this proof we would have to make precise the notions of different types of intersections for level sets Σ which were not smooth arcs. This is a nontrivial exercise left to the enthusiastic reader. Complete proof of the full existence of at least two p/q-periodic extended orbits in both cases, q even and q odd, via the original ideas of Birkhoff can be found in Brown and von Newmann (1977). ■

D. Monotone Orbits

If $f: A \to A$ is a monotone twist map, then recall that a point $z \in A$ is monotone if f preserves the ordering induced on eo(z) by the x coordinate and, as we saw in Section B, the monotone orbits of a monotone twist map behave

very well under limits. The goal of this section is to prove the existence of monotone periodic orbits and to give some consequences. In particular, we prove the following theorem.

Theorem 1 (Aubry–Mather). *For $f: A \to A$, an exact symplectic monotone twist* (ESMT) *map with ρ_0 and ρ_1 the rotation numbers of f restricted to the boundaries, for every $\omega \in [\rho_0, \rho_1]$, f has a point z_ω with $\rho(z_\omega) = \omega$ and z_ω monotone. Moreover, if $\omega = p/q$, then we may choose z_ω to be a monotone p/q-periodic point.*

Remarks. (1) This theorem states that for every "possible" rotation number there is a monotone orbit with that rotation number. The monotone orbits with irrational rotation number are called "quasi-periodic" orbits.

(2) Precursors of this theorem were shown by Hedlund in the context of geodesics on a torus and by Birkhoff for orbits in the billiard problem (see Section V.B and X.E). The technique used by Aubry and Mather (independently) were variational, i.e., they created an "energy" function on a space of candidates for orbits and showed that the minimum energy candidate is actually a monotone orbit (see section G). The techniques we use below are more topological in nature, depending less on the symplectic properties and more on the two dimensionality of the annulus.

We will first show that a monotone twist map with a p/q-periodic point must have a p/q-monotone periodic point. Then, using the fixed point theorem of the Section C, we see that exact symplectic monotone twist maps must have monotone periodic points of all possible rotation numbers. Limits of these periodic orbits will give the quasi-periodic orbits with irrational rotation numbers. Finally, we will note that the same techniques can be used, under appropriate hypotheses, to produce many nonmonotone orbits.

1. Existence of Monotone Periodic Orbits

We will find periodic orbits by carefully following the iterates of a subset of A which behaves nicely under a monotone twist map.

Notation. For $z_1, z_2 \in A$ with $X(z_1) < X(z_2)$ we let

$$B(z_1, z_2) = \{z \in A: X(z_1) < X(z) < X(z_2)\}. \tag{1}$$

Also, we let

$$I^+(z_1) = \{z \in A: X(z) = X(z_1) \text{ and } Y(z) \geq Y(z_1)\}, \tag{2}$$

$$I^-(z_1) = \{z \in A: X(z) = X(z_1) \text{ and } Y(z) \leq Y(z_1)\}. \tag{3}$$

Definitions. For each $z_1, z_2 \in A$, $X(z_1) < X(z_2)$ a set $C \subseteq$ closure $(B(z_1, z_2))$ is called a *positive diagonal* if it satisfies the following conditions:

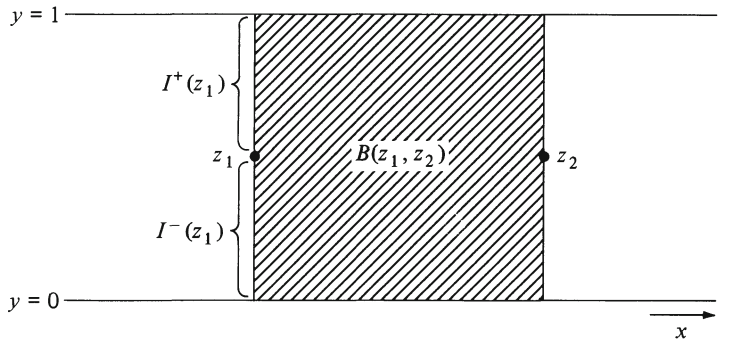

Figure D.1. The set $B(z_1, z_2)$.

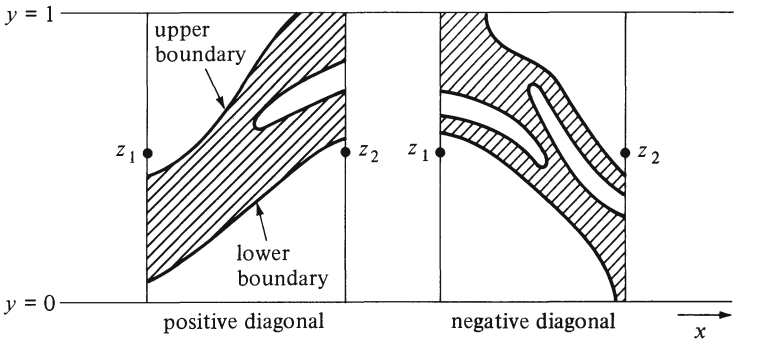

Figure D.2. Diagonals.

(i) C is the closure of its interior and the boundary of $C = \partial C$ is piecewise smooth,
(ii) C is simply connected,
(iii) $\partial C \cap (I^+(z_1) \cup I^-(z_2) \sim \{z_1, z_2\}) = \varnothing$,
(iv) ∂C contains a smooth arc connecting $I^-(z_1)$ and $I^+(z_2) \cup \{(x, 1): x \in \mathbb{R}\}$ and ∂C contains a smooth arc connecting $I^+(z_2)$ and $I^-(z_1) \cup \{(x, 0): x \in \mathbb{R}\}$.

We call C a *negative diagonal* if it satisfies (i) and (ii) above and

(iii′) $\partial C \cap (I^-(z_1) \cup I^+(z_2) \sim \{z_1, z_2\}) = \varnothing$,
(iv′) ∂C contains a smooth arc connecting $I^+(z_1)$ and $I^-(z_2) \cup \{(x, 0): x \in \mathbb{R}\}$.

Notation. If C is a positive or negative diagonal in $B(z_1, z_2)$, then there is an ordering to the components of $\partial C \cap B(z_1, z_2)$, i.e., one is "above" the other. If C is a positive diagonal, we call the component of $\partial C \cap B(z_1, z_2)$ that intersects $I^+(z_2)$ with the smallest y coordinate the *lower boundary* of C and the component of $\partial C \cap B(z_1, z_2)$ that intersects $I^-(z_1)$ with the largest y coordinate the *upper boundary* of C. For negative diagonals, replace $I^+(z_2)$ with $I^+(z_1)$ and $I^-(z_1)$ with $I^-(z_2)$. (See Figure D.2.)

The property which makes these sets useful is that they are preserved by monotone twist maps, i.e.,

Lemma 2. *Suppose $f: A \to A$ is a monotone twist map and $z_1, z_2 \in A$ satisfy $X(z_1) < X(z_2)$ and $X(f(z_1)) < X(f(z_2))$. If C is a positive diagonal of $B(z_1, z_2)$, then $f(C) \cap B(f(z_1), f(z_2))$ contains a component C_1 which is a positive diagonal of $B(f(z_1), f(z_2))$. Moreover, if we collect the components of $\partial C \cap B(z_1, z_2)$ into two disjoint sets α and β with α containing the upper boundary of C and β containing the lower boundary of C, then we may choose C_1, so that its upper boundary is in $f(\alpha)$ and its lower boundary is in $f(\beta)$.*

PROOF. The image of the upper boundary of C must connect $f(I^+(z_2)) \cup \{(x, 1): x \geq X(f(z_1))\}$ and $f(I^-(z_1))$ without intersecting $f(I^+(z_1)) \cup f(I^-(z_2))$. Similarly the image of the lower boundary of C must connect $f(I^-(z_1)) \cup \{(x, 0): x \leq X(f(z_2))\}$ and $f(I^+(z_2))$ without intersecting $f(I^+(z_1)) \cup f(I^+(z_2))$. Since f preserves orientation, this implies the lemma (see Figure D.3). ∎

Next we recall a technique from index theory which will be the main tool for producing periodic points. Suppose $g: \mathbb{R}^2 \to \mathbb{R}^2$ is a continuous map and D is a disk in $\mathbb{R}^2$ with boundary ∂D. Then for S^1, the unit circle in $\mathbb{R}^2$, we can define $\tilde{g}: \partial D \to S^1$ by $\tilde{g}(z) = (g(z) - z)/\|g(z) - z\|$ where $\|\cdot\|$ is the usual $\mathbb{R}^2$ norm. Since ∂D is homeomorphic to S^1, we can think of $\tilde{g}$ as a map between circles and define the *index of g as the number of times $\tilde{g}(z)$ goes around S^1 as z goes around ∂D.* The fundamental lemma we will use is:

Lemma 3. *If $g: \mathbb{R}^2 \to \mathbb{R}^2$ as above has nonzero index on a disk $D \leq \mathbb{R}^2$, then g has a fixed point in D. Moreover, if $g_1: \mathbb{R}^2 \to \mathbb{R}^2$ is sufficiently close to g in the sup norm topology, then g_1 will also have nonzero index on D.*

PROOF. See Milnor (1965). ∎

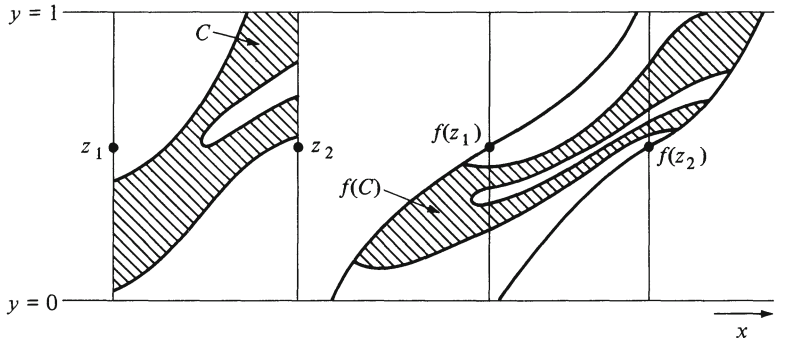

Figure D.3. Image of a positive diagonal.

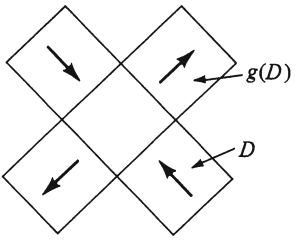

Figure D.4. A map with index -1.

The applications we will make of this lemma will be close to the case pictured in Figure D.4, where D is a rectangle tilted to the left and $g(D)$ is an intersecting rectangle tilted to the right with the boundaries mapping as indicated. The map g has index -1 on D so g contains a fixed point. Moreover, every map sufficiently close to g will also have a fixed point in D.

We will produce p/q-periodic points of a monotone twist map $f: A \to A$ by finding a subset D such that $f^q(D) - (p, 0)$ resembles Figure D.4. The set D will be a negative diagonal and $f^q(D) - (p, 0)$ will be a positive diagonal. The following lemma describes how these sets arise.

Lemma 4. *Suppose $f: A \to A$ is a monotone twist map and $z_1, z_2, w_1, w_2 \in A$ satisfy:*

(i) *z_1, z_2 are p/q-periodic points for f;*
(ii) *for $i = 0, \ldots, q$,*

$$X(f^i(z_1)) < X(f^i(z_2)),$$
$$X(f^i(w_1)) < X(f^i(w_2)),$$
$$X(f^i(w_1)) < X(f^i(z_2)),$$
$$X(f^i(z_1)) < X(f^i(w_2));$$

(iii) *for $j = 1, 2$, $\{X(w_j) - X(z_j)\}$ and $\{X(f^q(w_j)) - X(f^q(z_j))\}$ are the same sign;*
(iv) *for some i_1, i_2 between 0 and q, for $j = 1, 2$*

$$X(f^{i_j}(w_j)) - X(f^{i_j}(z_j)) \text{ and } X(w_j) - X(z_j) \text{ are opposite signs.}$$

Then there exists a negative diagonal D such that

(i) $\forall \zeta \in D$ *for* $i = 0, \ldots, q$, $X(f^i(z_1)) < X(f^i(\zeta)) < X(f^i(z_2))$,

$$X(f^i(w_1)) < X(f^i(\zeta)) < X(f^i(w_2));$$

(ii) *the map $f^q - (p, 0)$ has index -1 on D.*

Hence, $f^q - (p, 0)$ and every map sufficiently close to it has a fixed point in D.

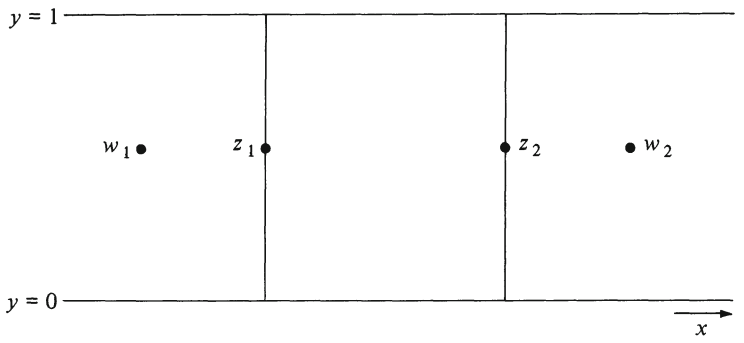

Figure D.5. Arrangement of z_1, z_2, w_1, w_2.

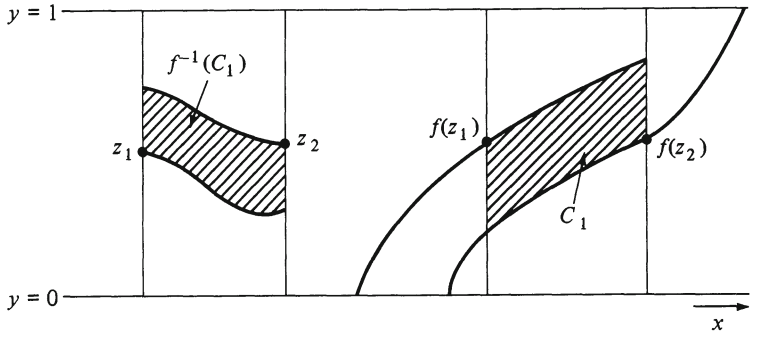

Figure D.6. The diagonal C_1.

Remark. We have not specified the set of which D is a negative diagonal because this will depend on the ordering of the iterates of z_1, z_2 and w_1, w_2 in A.

PROOF. We will consider several cases depending on the order of the points z_1, z_2, w_1, and w_2 in A.

Case 1. Suppose $X(w_1) < X(z_1) < X(z_2) < X(w_2)$. (See Figure D.5.) We will follow the image of $B(z_1, z_2)$ in a sequence of steps.

Step 1. Note that $f(B(z_1, z_2)) \cap B(f(z_1), f(z_2))$ is a positive diagonal in $B(f(z_1), f(z_2))$—call it C_1. Also, $f^{-1}(C_1)$ is a negative diagonal of $B(z_1, z_2)$. (See Figure D.6.)

Step 2. Hence using Lemma 2, we may choose a sequence C_i of positive diagonals of $B(f^i(z_1), f^i(z_2))$ such that $f^{-i}(C_i)$ is a nested sequence of diagonals of $B(z_1, z_2)$.

Step 3. We refine the choice of the C_i's by following the orbits of the w's. In particular, fix i_1 and i_2 such that $0 < i_1 < i_2 < q$; $X(f^i(w_1)) < Xf^i(z_1))$ for $i < i_1$; $X(f^i(w_1)) > X(f^i(z_1))$ for $i_1 \leq i < i_2$; and $X(f^{i_2}(w_1)) < X(f^{i_2}(z_1))$. Then if we follow the iterates of $I^+(z_1)$ under f, we must have that $f^{i_2}(I^+(z_1))$

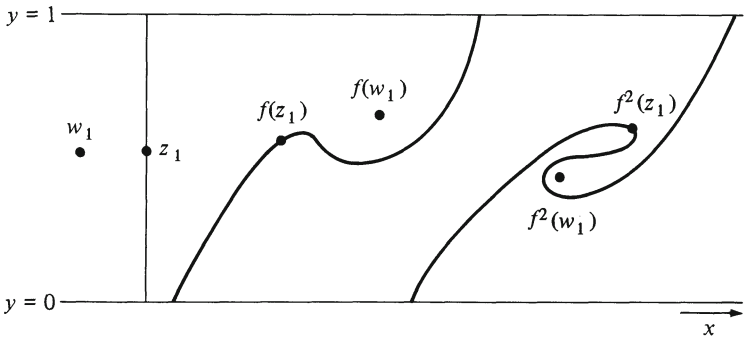

Figure D.7. Image of $B(z_1, z_2)$.

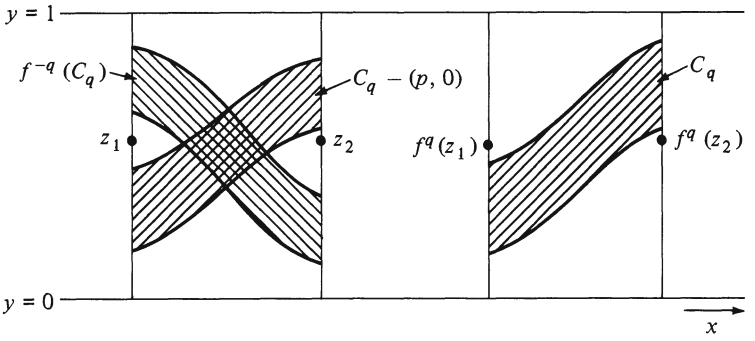

Figure D.8. Diagonal C_q and its preimage.

contains an interval which connects $I^-(f^{i_2}(z_1))$ to $I^+(f^2(z_2)) \cup \{(x, 1): x > (X(f^{i_2}(z_1))\}$ which does not contain z_1. Hence, we may choose C_{i_2} so that it does not contain $f^{i_2}(z_1)$. Similarly, using the orbit of w_2 we see that for some i, $0 < i \leq q$ we may choose C_i so that it does not contain $f^i(z_2)$. (See Figure D.7.)

Step 4. Since z_1 and z_2 are periodic, the set $\{\zeta - (p, 0): \zeta \in C_q\} = C_q - (p, 0)$ is a positive diagonal of $B(z_1, z_2)$. By construction, the set $D = f^{-q}(C_q)$ is a negative diagonal of $B(z_1, z_2)$. Also $z_1, z_2 \notin D$. Also the upper and lower boundaries of C_q are contained in $f^q(I^+(z_1))$ and $f^q(I^-(z_2))$, respectively, with $\partial C_q \cap B(f^q(z_1), f^q(z_2)) \subseteq f^q(I^+(z_1)) \cup f^q(I^-(z_2))$. Hence, $f^q - (p, 0)$ on D satisfies the conditions of Lemma 3 and has index -1 on D, so f, and every map sufficiently close to f, has a p/q-periodic point $\zeta \in D$ satisfying for $i = 0, \ldots, q$,

$$X(f^i(w_1)) < X(f^i(\zeta)) < X(f^i(w_2)),$$
$$X(f^i(z_1)) < X(f^i(\zeta)) < X(f^i(z_2)).$$

which completes Case 1. (See Figure D.8.)

For the other cases we need merely choose the initial box differently and proceed as above, i.e.,

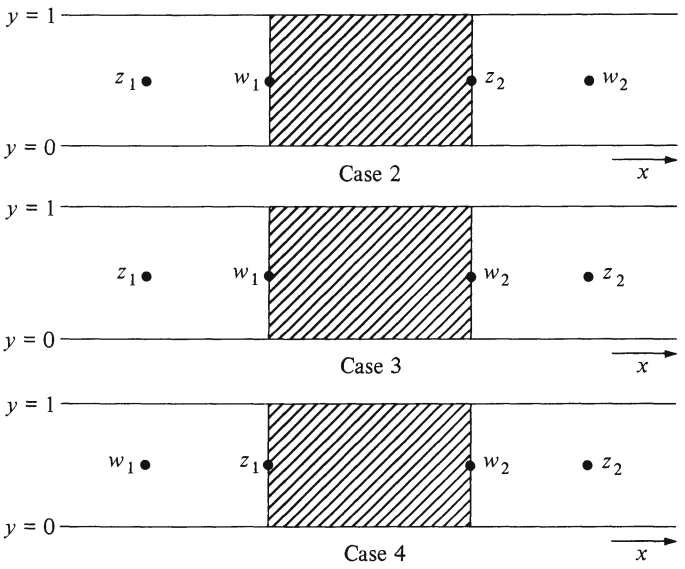

Figure D.9. The other cases.

Case 2. $X(z_1) < X(w_1) < X(z_2) < X(w_2)$.
Case 3. $X(z_1) < X(w_1) < X(w_2) < X(z_2)$.
Case 4. $X(w_1) < X(z_1) < X(w_2) < X(z_2)$.
In each case we follow the shaded region of Figure D.9 and produce a p/q-periodic point which stays to the right of z_1 and w_1 and left of z_2 and w_2. ■

So to produce p/q-monotone periodic points we must find pairs of points which do not change order as described above. In certain situations this turns out to be fairly straightforward.

Suppose $f: A \to A$ has a p/q-monotone periodic point z_0 and a p/q-nonmonotone periodic point w_0. First we fix $w_1 \in \text{eo}(w_0)$. Next we let $z_1 \in \text{eo}(z_0)$ be the point of $\{z \in \text{eo}(z_0)$: for some i, $X(f^i(z)) - X(f^i(w_1))$ and $X(z) - X(w_1)$ have opposite signs$\}$ with the largest x coordinate. Such points z_1 must exist because w_1 is nonmonotone. We let $w_2 \in \text{eo}(w_0)$ be the point of $\{w \in \text{eo}(w_0)$: $X(f^i(w)) > X(f^i(z_1))$ and $X(f^i(w)) > X(f^i(w_1))$ for all $i\}$, with the smallest x coordinate. Finally we let $z_2 \in \text{eo}(z_0)$ be the point in $\{z \in \text{eo}(z_0)$: for some i, $X(f^i(z)) - X(f^i(w_2))$ and $X(z) - X(w_2)$ have opposite signs$\}$ with smallest x coordinate. Since $X(f^i(z_2)) > X(f^i(w_2))$ for some i, we must have $X(f^i(z_2)) > X(f^i(z_1))$ for that i. But then $X(f^i(z_2)) > X(f^i(z_1))$ for all i because z_0 is a monotone point.

By definition, z_1, z_2, w_1, w_2 now satisfy the conditions of Lemma 4 and hence there is a set D such that $f^q - (p, 0)$ has index -1 on D and for all $\zeta \in D$. and for $i = 0, \ldots, q$.

$$X(f^i(z_1)) < X(f^i(\zeta)) < X(f^i(z_1)),$$
$$X(f^i(w_1)) < X(f^i(\zeta)) < X(f^i(w_2)).$$

Lemma 5. *Suppose $f: A \to A$ is a monotone twist map and f has a p/q-monotone periodic point z_1 and a p/q-nonmonotone periodic point w_0. Then there is a set $D \subseteq A$ such that $f^q - (p, 0)$ has index -1 on D and every p/q-periodic point in D is monotone. Hence, every map sufficiently close to f will have a p/q-monotone periodic point in D.*

PROOF. We will choose points $z_1, z_2 \in \text{eo}(z_0)$ and $w_1, w_2 \in \text{eo}(w_0)$ which satisfy the hypotheses of Lemma 4. First note that the set $\{z \in \text{eo}(z_0)$: for some i, $X(f^i(w_0)) - X(f^i(z))$ and $X(w_0) - X(z)$ have opposite signs$\}$ is nonempty. (If it were empty, then the orbit of w_0 would be trapped between two neighboring points of $\text{eo}(z_0)$ and hence would be monotone.) Hence, we may choose z_1 to be the point of this set with the largest x coordinate and we let $w_1 = w_0$. Now we consider two cases:

Case 1. z_1 and z_2 are neighboring points of $\text{eo}(z_0)$: In this case, since there are no points of $\text{eo}(z_0)$ between z_1 and z_2, any p/q-periodic point ζ of f in D are monotone because for all i, $X(f^i(z_1)) < X(f^i(\zeta)) < X(f^i(z_2))$, and this completes the proof of the lemma. Note in this case the p/q-periodic points in D are new periodic points (i.e., not in $\text{eo}(z_0)$ or $\text{eo}(w_0)$).

Case 2. z_1 and z_2 are not neighboring points of $\text{eo}(z_0)$: Some point $z_3 \in \text{eo}(z_0)$ with $X(z_1) < X(z_3) < X(z_1)$ is contained in a continuum of fixed points of $f^q - (p, 0)$. In this situation, we may fix $\zeta \in D$ a fixed point of $f^q - (p, 0)$ in a continuum of fixed points with index -1. Now either for every $z \in \text{eo}(z_0)$ the sign of $X(f^i(z)) - X(f^i(\zeta))$ is independent of i, or there is a $z_3 \in \text{eo}(z_0)$ and i such that $X(f^i(z_3)) - X(f^i(\zeta))$ and $X(z_3) - X(\zeta)$ have different signs. In the first case, ζ is monotone and the proof is complete. In the latter case we can repeat the above argument with z_3 replacing z_2 and ζ replacing w_2 to produce another set D_1 of index -1 for $f^q - (p, 0)$ between z_1 and z_3. Since there are only finitely many points of $\text{eo}(z_0)$ between z_1 and z_2, the process will end with the required set D. ∎

Remark. The lemma shows that either the original monotone orbit has index -1 or there is another monotone orbit with index -1. Finally, we prove that monotone periodic orbits are present whenever periodic orbits are present.

Lemma 6. *Suppose $f: A \to A$ is a monotone twist map and f has a p/q-periodic point, then f has a p/q-monotone periodic point.*

PROOF. Let $w_0 \in A$ be a p/q-periodic point for f. Then if w_0 is monotone, there is nothing to prove, so assume w_0 is nonmonotone. The idea of the proof is to find a one-parameter family of maps f_t, $t \in [0, 1]$ such that f and f_t agree on

the orbit of w_0, $f_1 = f$, and f_0 has a p/q-monotone periodic point. Now the set of t's such that f_t has a p/q-monotone periodic point is closed by the limit arguments of Section B. But it is also open by Lemma 5! Hence, all the f_t's, including $f_1 = f$, must have p/q-monotone periodic orbits. The first task is to construct the family f_t.

Fix $x_0 \in \mathbb{R}$ so that $\{x_0 + ip/q: i \in \mathbb{Z}\} \cap \{X(w): w \in \mathrm{eo}(w_0)\} = \varnothing$ and let $\varepsilon > 0$ be such that $\forall i \in \mathbb{Z}$, $\forall w \in \mathrm{eo}(w_0)$, $|(x_0 + ip/q) - X(w)| > 2\varepsilon$. We may assume (by enlarging A if necessary) that $\forall x \in \mathbb{R}$, $\{f(x, y): y \in [0, 1]\} \cap \{(x + p/q, y): y \in [0, 1]\} \neq \varnothing$. Next we let $g: [-\varepsilon, \varepsilon] \times [0, 1] \times ((0, 1) \times [0, 1]) \to [-\varepsilon, \varepsilon] \times [0, 1]$ be a two-parameter family of maps from $[-\varepsilon, \varepsilon] \times [0, 1]$ onto itself satisfying

(i) for all $y_0 \in (0, 1)$, $t \in [0, 1]$, $g(\cdot, \cdot, y_0, t)$ is a C^∞ diffeomorphism equal to the identity on a neighborhood of the boundary of $[-\varepsilon, \varepsilon] \times [0, 1]$;
(ii) for all $y_0 \in (0, 1)$, $t \in [0, 1]$, $(x, y) \in [-\varepsilon, \varepsilon] \times [0, 1]$ $X(g(x, y, y_0, t)) = x$;
(iii) for all y_0, $g(\cdot, \cdot, y_0, 1)$ is the identity on $[-\varepsilon, \varepsilon] \times [0, 1]$;
(iv) $g(0, y_0, y_0, 0) = (0, \frac{1}{2})$;

i.e., each map $g(\cdot, \cdot, y_0, t)$ slides points up and down on lines of constant x coordinate. By composing the original map f with various copies of the map g we will be able to obtain a map such that $(x_0, \frac{1}{2})$ is a monotone p/q-periodic point while the orbit of the nonmonotone point w_0 is undisturbed. Let y_i be such that $X(f(X_0 + ip/q, y_i)) = x_0 + (i + 1)p/q$, and let $\tilde{y}_i$ be given by $\tilde{y}_i = (f(x_0 + (i + 1)p/q, y_{i-1}))$. Define $g_{i,t}: A \to A$ and $\tilde{g}_{i,t}: A \to A$ by defining

$$g_{i,t}(x, y) = g(x - (x_0 + ip/q), y, y_i, t),$$

$$\tilde{g}_{i,t}(x, y) = g(x - (x_0 + ip/q), y, \tilde{y}_i, t),$$

where the right-hand side is defined and then extending to all of A so that $g_{i,t}$ and $\tilde{g}_{i,t}$ are the identity at (x, y) where $|x - (x_0 + ip/q)|$ has fractional part larger than ε and $g_{i,t}(x + 1, y) = g_{i,t}(x, y) + (1, 0)$ and $\tilde{g}_{i,t}(x + 1, y) = \tilde{g}_{i,t}(x, y) + (1, 0)$. Then, for $t \in [0, 1]$ let

$$f_t(x, y) = \tilde{g}_{0,t} \circ \cdots \circ \tilde{g}_{q-1,t} \circ f \circ g_{q-i,t}^{-1} \circ \cdots \circ g_{1,t}^{-1} \circ g_{0,t}^{-1}.$$

Then f_t has the following properties.

(i) f_t is a smooth one-parameter family of monotone twist maps;
(ii) $f_1 = f$ and f_0 has a p/q-monotone periodic point at $(x_0, \frac{1}{2})$;
(iii) for all t, f_t has a p/q-nonmonotone periodic point w_0 and the orbit of w_0 is independent of t. (See Figure D.10.)

Now we claim that $f = f_1$ must have a p/q-monotone periodic point. Let $\Theta = \{s \in [0, 1]: f_s$ has a p/q-monotone periodic point$\}$. Then $\Theta \neq \varnothing$ since $0 \in \Theta$ and Θ is closed by Lemma B.5. We claim that Θ is also open. Suppose $s \in \Theta$. Then f_s has a p/q-monotone periodic point. But f_s has a p/q-nonmonotone point as well, the point w_0. So f_s satisfies the hypotheses of Lemma 4. Hence, every map sufficiently close to f_s will also have a p/q-

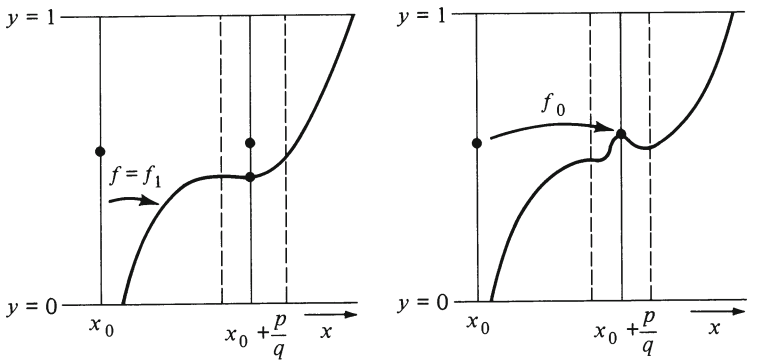

Figure D.10. Action of the deformation of f by g_{1t}.

monotone periodic point, i.e., $s \in$ interior Θ. Hence, Θ is both open and closed so $\Theta = [0, 1]$ or $f_1 = f$ has a p/q-monotone periodic point. This completes the proof of the lemma. ∎

We are now ready to give a proof of the Aubry–Mather theorem.

PROOF OF THEOREM 1. Fix $f: A \to A$ an exact symplectic monotone twist map with ρ_0, ρ_1 the rotation numbers of f restricted to the boundaries of A. Then for every rational $p/q \in [\rho_0, \rho_1]$, f has a p/q-monotone periodic point by Lemma 6. For any irrational $\omega \in [\rho_0, \rho_1]$ if we choose a sequence p_n/q_n of rationals with $\lim_{n\to\infty} p_n/q_n = \omega$ and a sequence z_n of p_n/q_n-monotone periodic points, $X(z_n) \in [0, 1]$, then some subsequence of the z_n's will converge to $z_\omega \in A$ and by Lemma B.5, z_ω will have a monotone orbit with $\rho(z_\omega) = \omega$. Hence, every rotation number possible for f is represented by a monotone orbit and the proof is complete. ∎

The techniques used above can also be applied to show the existence of many other orbits, provided there are orbits which change y coordinate dramatically. Recall that we say a map $f: A \to A$ satisfies condition B if for every $\varepsilon > 0$ there exist $z_1, z_2 \in \mathbb{Z}$ and $n > 0$ such that $Y(z_1) < \varepsilon$, $Y(z_2) > 1 - \varepsilon$ and $Y(f^n(z_1)) > 1 - \varepsilon$, $Y(f^n(z_2)) < \varepsilon$. Given a monotone twist map which satisfies condition B there will exist orbits of almost every possible behavior. In fact, the condition B can be used in place of area preservation (exact symplectic) in the Aubry-Mather theorem yielding the following theorem of Boyland.

Theorem 7. *If $f: A \to A$ is a monotone twist map satisfying condition* B *and ρ_1, ρ_2 are the rotation numbers of f on the boundary of A, then for every $\omega \in [\rho_1, \rho_2]$ there is a $z_\omega \in A$ such that z_ω is monotone and z_ω has rotation number ω. Moreover, if ω is rational, then we may assume z_ω is periodic.*

PROOF. See Boyland (1988). ∎

In the next section we will give conditions under which condition B will be satisfied by an ESMT map.

Such maps will have many monotone orbits for two reasons since Theorems 1 and 7 both apply. More importantly, maps satisfying condition B will have many periodic orbits which are nonmonotone. The idea is that the orbits given by condition B will tend to stretch positive diagonals for many different extended orbits. Informally, we can say that for a monotone twist map satisfying condition B, given any list (finite or infinite) of different periodic orbits with different rotation numbers, there is an orbit which "shadows" each of these periodic orbits in the order listed. In particular, there will be orbits for which the rotation number is not defined and many different nonmonotone periodic orbits. For a precise statement see Hall (1989). To illustrate we state a special case of these theorems.

Theorem 8. *Suppose $f: A \to A$ is a monotone twist map satisfying condition* B *with rotation numbers $\rho_0 < \rho_1$ on the boundaries. Then for any pair of numbers w_1, w_2 with $\rho_0 < w_1 < w_2 < \rho_1$ there exists points z_1 and z_2 in A such that*

$$\lim_{n\to\infty} \frac{X(f^n(z_1)) - X(z_1)}{n} = \lim_{n\to-\infty} \frac{X(f^n(z_2)) - X(z_2)}{n} = w_1$$

and

$$\lim_{n\to-\infty} \frac{X(f^n(z_1)) - X(z_1)}{n} = \lim_{n\to\infty} \frac{X(f^n(z_2)) - X(z_2)}{n} = w_2.$$

For ESMT maps we will see that either all orbits are monotone or condition B is satisfied at least in a subannulus so there are many nonmonotone orbits.

E. Invariant Circles

In the previous sections we have shown that many different periodic and quasi-periodic orbits are present in every twist map. However, all of the orbits we have considered so far can (and sometimes do) form a measure zero set in the annulus. Happily, it turns out that the existence or nonexistence of certain types of periodic orbits can have implications for the qualitative behavior of large sets of orbits. In this section we will consider the relationship between the existence of large (open) invariant sets, the existence of invariant curves and the existence of nonmonotone periodic orbits. The reader is encouraged to compare this section with Chapter IX on the KAM theorem.

The most important type of stability for ESMT maps can be illustrated by looking at the simplest map $g_0: A \to A: (x, y) \to (x + y, y)$. Here the x coordinate, or angular coordinate, changes via rigid translation. Hence in the

annulus, we expect the x coordinate to take on many different values. In fact, when y is irrational, the orbits are dense in the angular coordinate in the annulus. Since the irrationals are dense, there is no reasonable way to look for regions in A where the x variable is constrained to stay in a small set over an entire orbit.

Conversely, g_0 preserves the y coordinate. Hence, if a point has a small y coordinate, its entire orbit has a small y coordinate. Hence, it is reasonable for a given twist map to search for invariant sets which contain one boundary of A but not the other boundary. The existence of this type of invariant set corresponds to a type of "stability" for ESMT maps which arise in applications (see Chapter IX). The boundary of such an invariant set will also be an invariant set which separates A into components containing the two boundary components of A. For ESMT maps, this boundary is a particularly nice set, which motivates the following definition. (See Figure E.1).

Definition. Given $f: A \to A$, an ESMT map, an *invariant circle* for f is a set $\Gamma \subseteq A$ such that Γ is the graph of a continuous function $\varphi: \mathbb{R} \to [0, 1]$ satisfying

(1) φ is periodic, $\forall x, \varphi(x + 1) = \varphi(x)$,
(2) $\Gamma = \{x, \varphi(x)): x \in \mathbb{R}\}$ is invariant, $f(\Gamma) = \Gamma$.

Clearly, if $f: A \to A$ has an invariant circle $\Gamma = \{(x, \varphi(x)): x \in \mathbb{R}\}$ the graph of $\varphi: \mathbb{R} \to [0, 1]$, then the set $U = \{(x, y): y < \varphi(x)\}$ is an open invariant set containing the lower boundary of A. Remarkably, for ESMT maps the converse holds.

Theorem 1. *Suppose $f: A \to A$ is an ESMT map with an invariant set $U \subseteq A$ such that*

(1) *U is simply connected,*
(2) $U + (1, 0) = \{(x, y) + (1, 0) : (x, y) \in U\} = U$,
(3) *U is open and contains $\{(x, 0): x \in \mathbb{R}\}$ in its interior,*
(4) $U \cap \{(x, 1): x \in \mathbb{R}\} = \varnothing$.

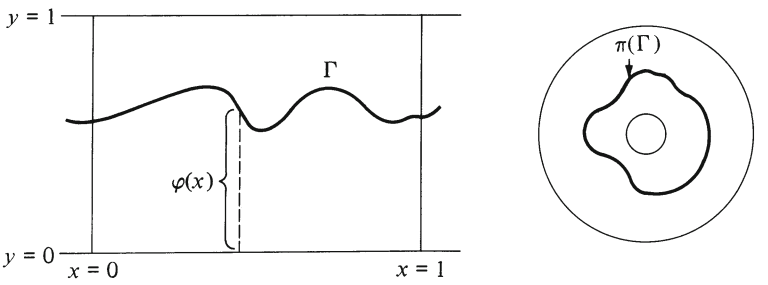

Figure E.1. Invariant circles.

Then there exists $\varphi\colon \mathbb{R} \to \{0, 1\}$, *continuous and periodic* $(\varphi(x + 1) = \varphi(x))$, *such that the boundary of* U *is the invariant circle given by the graph of* φ, *i.e.*,

$$\partial U = \Gamma = \{(x, \varphi(x))\colon x \in \mathbb{R}\}.$$

Moreover, there exists a constant K *independent of* U *(depending only on* f*) such that* φ *is Lipschitz with constant* K (i.e., $\forall_{x,1} x_2 \in \mathbb{R}$, $|\varphi(x_1) - \varphi(x_2)|/|x_1 - x_2| < K$).

This remarkable theorem says that there is a one-to-one relationship between the invariant sets which separate the boundaries of A and Lipschitz invariant circles. It was first proven by Birkhoff and a proof in modern notation can be found in Herman (1983). However, the ideas involved in the proof are both simple and elegant so we outline them here.

IDEA OF THE PROOF. The first step is to show that the boundary of the invariant set U is a graph. To do this we identify three types of points in U as follows:

(1) a point $z \in U$ is called *accessible from below* if $\{(X(z), y)\colon 0 \leq y(z)\} \subseteq U$,
(2) a point $z \in U$ is called *accessible from the left* if z is not accessible from below and there exists a continuous curve $\gamma\colon [0, 1] \to U$ such that $Y(\gamma(0)) = 0$, $\gamma(1) = z$, there is an interval $[0, a]$ such that $X(\gamma(t))$ is strictly increasing for $t \in [0, a]$ and $\gamma(t)$ is not accessible from below for any $t > a$,
(3) a point $z \in U$ is called *accessible from the right* if there exists a curve as in (2) with "$X(\gamma(t))$ strictly increasing" replaced by "$X(\gamma(t))$ strictly decreasing." (See Figure E.2.)

We note that the fact that U is simply connected implies that the sets U_B, U_L, and U_R of points in U accessible from below, left, and right, respectively, are disjoint and $U = U_B \cup U_R \cup U_L$ and $U_R + (1, 0) = U_R$, $U_L + (1, 0) = U_L$.

To complete the proof we just note that $f(U_L) \subsetneq (U_L)$, i.e., U_L maps strictly inside itself. This is clear because the curve giving ∂U_L is mapped under f to an arc inside U_L (see Figure E.3) which violates the area preservation hypothesis on f.

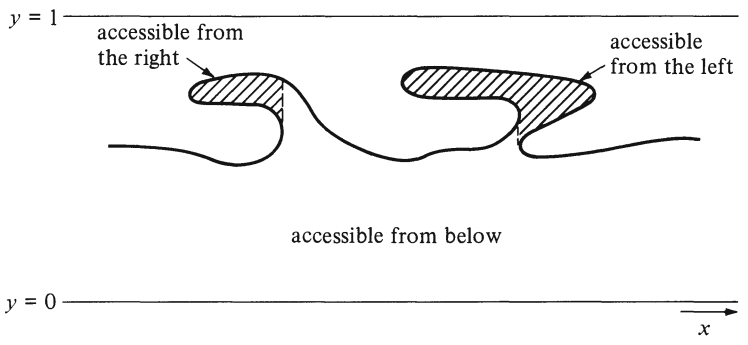

Figure E.2. Accessible regions.

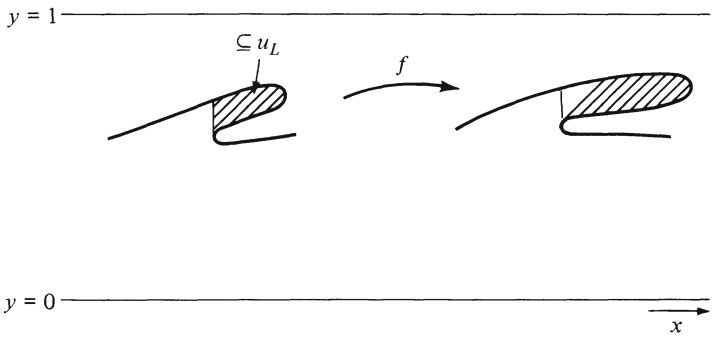

Figure E.3. A region accessible from the left and its image.

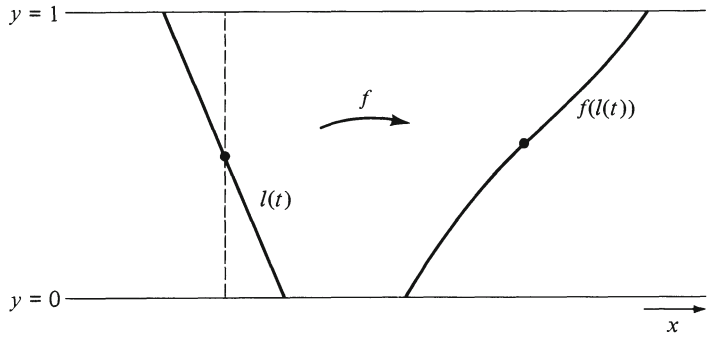

Figure E.4. Application of a twist map to an almost vertical line.

Similarly $f^{-1}(U_R) \subsetneq U_R$, again violating area preservation. Hence, $U = U_B$ and the boundary of U can be used to define φ: $\mathbb{R} \to [0, 1]$ as follows:

$$\varphi(x) = \sup\{y \in [0, 1] : (x, y) \in U\}.$$

Then φ is periodic with period one. To show that the boundary of U is actually the graph of φ, we need to show that φ is continuous. It turns out to be easier to show that φ is Lipschitz.

Since f is monotone twist, there exists a $K_1 > 0$ such that if ℓ is the line segment ℓ: $t \to (-kt + b, t)$ for $k \geq K_1$ and any $b \in \mathbb{R}$, then $d(X(f(\ell(t))))/dt > 0$. (See Figure E.4.)

Hence, if $x_1, x_2 \in \mathbb{R}$ with $x_1 < x_2$ and $\varphi(x_1) - \varphi(x_2) > K_1(x_2 - x_1)$, then we must have $X(f(x_1, \varphi(x_1))) > X(f(x_2, \varphi(x_2)))$, which contradicts the fact that U has no points accessible from the left. (See Figure E.5.) Similarly, we can apply the above argument to f^{-1} to see that there is a $K_2 > 0$ such that $\forall x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$, $\varphi(x_2) - \varphi(x_1) < K_2(x_2 - x_1)$. Taking $K = \max\{K_1, K_2\}$ we see that φ is Lipschitz with constant K and hence φ is continuous.

Noting that the constant K above depends only on f, not on U, the idea of the proof is complete. ∎

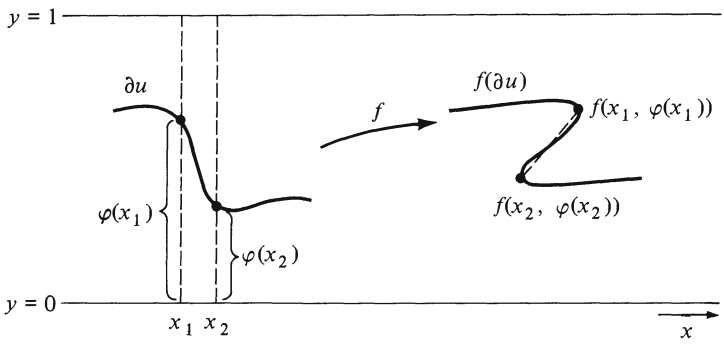

Figure E.5. Effect of twist map on an almost vertical line.

The above relationship between invariant sets and invariant circles can be restated in terms of "Condition B" of the previous sections.

Theorem 2. *Suppose $f: A \to A$ is an ESMT map which has no invariant circles in the interior of A. Then f satisfies condition* B.

Remark. This theorem is also due to Birkhoff and applies to any "annular region" between invariant circles. Since a map satisfying condition B clearly can have no interior invariant circles, we see that condition B is really equivalent to having no interior invariant circles.

PROOF. Suppose $f: A \to A$ is an ESMT map with no interior invariant circles. For each $\varepsilon > 0$ we must show that there exist points $z_1, z_2 \in A$ and $n > 0$ such that $Y(z_1) < \varepsilon$, $Y(z_2) > 1 - \varepsilon$ and $Y(f^n(z_1)) > 1 - \varepsilon$, $Y(f^n(z_2)) < \varepsilon$. This is easily seen to be equivalent to showing that there exists $n_1, n_2 > 0$ such that $Y(f^{n_1}(z_1)) > 1 - \varepsilon$ and $Y(f^{n_2}(z_2)) < \varepsilon$ (see Problems) and we will prove only the existence of z_1, the existence of z_2 being symmetric.

Fix $\varepsilon > 0$. We may assume $\varepsilon < \frac{1}{2}$ because otherwise there is nothing to prove. Now we let $U = \{z \in A: Y(f^n(z)) > 1 - \varepsilon \text{ for some } n \geq 0\}$. This set is clearly open and $f^{-1}(U) \subseteq U$ and $U + (1, 0) = U$. Since f is area preserving, we have $f^{-1}(U) = U$, so $f(U) = U$. Since U contains the boundary $y = 1$ of A, either the boundary component of U which separates A is an interior invariant circle for f or the closure of U contains points in the $y = 0$ boundary of A. Since f has no interior invariant circles, U must intersect the set $\{z \in A: Y(z) < \varepsilon\}$, any such point of intersection of U with $\{z \in A: Y(z) < \varepsilon\}$ will serve as z_1. ∎

The region in an annulus which contains no interior invariant circles is called a *zone of instability*. Since condition B is satisfied in such a zone, the techniques of Section C and D can be used to show that zones of instability of ESMT maps contain many different orbits. One of the basic problems is to

estimate the width of the zones of instability for a given ESMT map, and, in particular, determine if it has any interior invariant circles. The following theorem relates the width of the zones of instability to the existence of certain nonmonotone periodic orbits.

Theorem 3 (Boyland and Hall). *Let $f: A \to A$ be an ESMT map. Then f has an invariant circle with rotation number w if and only if for every convergent $p/q \in \mathbb{Q}$ in lowest form, every p/q-periodic orbit of f is monotone.*

Corollary 4. *If $f: A \to A$ is an ESMT map and f has a p/q-periodic orbit which is not monotone (p/q in lowest form), then f has no invariant circles with rotation number w whenever $|w - p/q| < 1/(2q^2)$.*

Remarks. (1) We recall that each $w \in [0, 1]$ may be represented as a continued fraction in the form

$$w = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}},$$

where all the a_i's are positive integers. Moreover, w is rational if and only if the continued fraction representation is finite and the representation is unique for irrational w. It is unique for rational w if the last a_n is one. The ith convergent of

$$w = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_i + \ddots}}}$$

is the rational

$$\frac{p_i}{q_i} = \cfrac{1}{a_1 + \ddots + \cfrac{1}{a_i}}.$$

The convergents of w are the closest rationals with small denominator to w, i.e., for every convergent p_i/q_i we have $|w - p_i/q_i| < 1/q_i^2$. Moreover, if $|p/q - w| < 1/(2q^2)$, then p/q is a convergent of w.

(2) Since all orbits on an invariant circle must be monotone, the fact that low period nonmonotone periodic orbits imply nonexistence of invariant circles is no surprise. What is interesting in this theorem and corollary is the estimate on the width of the interval of rotation numbers cleared by a given nonmonotone periodic orbit.

PROOF OF THE COROLLARY. If f has a nonmonotone p/q-periodic orbit, then for every w with p/q a convergent of w the theorem implies f does not have an invariant circle with rotation number w. As noted above, if $|w - p/q| < 1/(2q^2)$, then p/q is a convergent of w and the corollary follows. ∎

IDEA OF THE PROOF OF THE THEOREM. Suppose f does not have an invariant circle with rotation number w. Then (see problems) there is an interval about w such that f has no invariant circles with rotation numbers in this interval. Hence, f has a zone of instability with boundary circles having rotation numbers straddling w and hence condition B is satisfied in this zone. By the techniques of the last section many nonmonotone periodic orbits can be constructed; in particular, nonmonotone p/q periodics can be constructed for p/q arbitrarily close to w.

On the other hand, suppose f has a nonmonotone p/q periodic orbit, call it z_0. Then the distance between successive iterates of z_0 must sometimes be much larger than p/q and sometimes much smaller. Since f is a monotone twist map, this means that the orbit of z_0 must sometimes have large y coordinate and sometimes have small y coordinate and, most importantly, the points of $\text{eo}(z_0)$ are not arranged the same as rigid rotation by p/q. In particular, there will be points $z_1, z_2, z_3 \in \text{eo}(z_0)$ such that $X(z_1) < X(z_2) < X(z_3)$ but $X(f(z_1)) < X(f(z_3)) < X(f(z_2))$. Any curve passing close to $\text{eo}(z_0)$ will be mapped into a curve which is not a graph (see Figure E.6). To make the qualitative comparison between p/q and the rotation numbers of the possible invariant circles, we form a circle endomorphism by considering the x coordinate of f on $\text{eo}(z_0)$ and comparing this circle map to f on $A \sim \text{eo}(z_0)$. ∎

We end this chapter by restating the important Kolmogorov–Arnold–Moser (KAM) Theorem (see Chapter IX). We have seen above that there is a dichotomy in the behavior of orbits for ESMT maps. Invariant circles imply the existence of constraints on the rotation behavior of orbits, whereas the lack of invariant circles implies condition B and hence a multitude of different types of orbits and rotation behaviors. Indeed, the amount of regularity

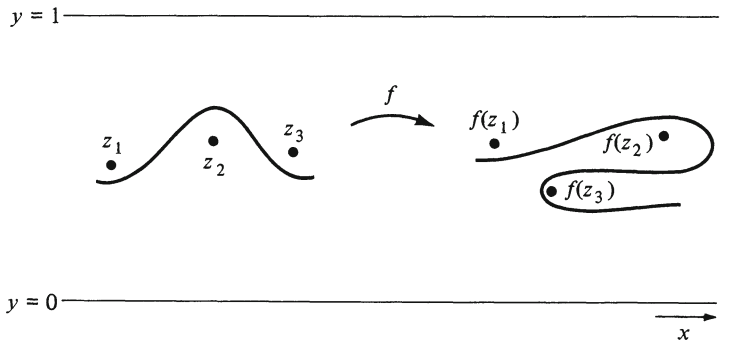

Figure E.6. An arc around $\text{eo}(z_0)$ and its image.

implied by the existence of invariant circles and the conditions that we already know imply their nonexistence makes it seem unlikely that they exist. Remarkably, invariant circles with certain rotation numbers turn out to be very robust under perturbation. We state one "KAM Theorem" due to Herman making this precise—there is a large family of such theorems and the interested reader is referred to Chapter IX and Section G.

A KAM Theorem. *Let $F_\varepsilon: A \to A$ denote a one-parameter family of ESMT maps with $\varepsilon \in [0, 1]$, which is continuous in the C^4 topology and $f_0(x, y) = (x + y, y)$). Then if $w \in [0, 1]$ is an irrational of constant type (e.g., $\exists C > 0$ such that $\forall p/q \in \mathbb{Q}$, $|w - p/q| \geq c/q^2$), there exists $\varepsilon_w > 0$ such that if $\varepsilon \in [0, \varepsilon_w]$ then f_ε has an invariant circle with rotation number w.*

Remarks. (1) We note that both the number theoretic and smoothness conditions are necessary for the theorem, i.e., for Liouville numbers w the theorem is false and for families of maps continuous only in the C^2 family the theorem is false. Since both subtle number theoretic and smoothness conditions are required, it is not surprising that the proof is both subtle and difficult.

(2) The theorem implies that when ε is small, f_ε will have many invariant circles. Hence, orbits will be trapped in narrow annular regions (see Section G for additional details and references). This is particularly useful in applications (see Chapter IX).

In light of the above theorems we can state a corollary of the KAM theorem as follows: Recall that

$$\frac{(\sqrt{5} - 1)}{2} = \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}$$

satisfies the condition of the above KAM theorem, hence

Corollary 5. *There exist an $\varepsilon > 0$ such that if $f: A \to A$ is an ESMT map whose distance from $f_0: A \to A: (x, y) \to (x + y, y)$ in the C^4 topology is less than ε, then if*

$$\frac{p_n}{q_n} = \cfrac{1}{1 + \cfrac{1}{1 + \ddots + 1}}$$

is a convergent of $(\sqrt{5} - 1)/2$, all p_n/q_n periodic points of f are monotone.

Hence, one can, very loosely, say that the KAM theorem for ESMT maps says that perturbations of the simple map $f_0: A \to A: (x, y) \to (x + y, y)$ cannot have nonmonotone periodic orbits of certain periods. The problem of creation of periodic orbits via small perturbations is a well-known one in dynamics, with the fundamental statement being the C^1-closing lemma of

Pugh and Robinson. This lemma states loosely that small C^1 perturbations can be used to create periodic orbits. The situation for smoothing "closing lemmas" is unclear at this time.

F. Applications

We end this chapter with a few comments on some applications of the above theorems on twist maps.

In the billiards problem of Section V.B, the section map constructed that corresponds with the billiard ball hitting the edge of the table turns out to be an exact symplectic (i.e., area preserving) monotone twist map. Poincaré's Last Geometric Theorem implies that there are periodic orbits of every period. The Aubry–Mather Theorem implies that there are quasi-periodic orbits with irrational rotation number. (We recall that the rotation number is for orbits of the twist map and is not so easy to compute from just the orbits of the billiard ball on the table.) Finally the KAM theorem implies that for billiard tables that are sufficiently close to circular, there are billiard ball orbits whose points of collision with the boundary are dense on the boundary. Moreover, there is an associated "stability" statement that is orbits which start close to tangent with the boundary of the table will stay close to tangent. [See Birkhoff (1927) and Moser (1973).]

For the linear crystal model of Section V.B, which gives an exact symplectic monotone twist map on the cylinder, Poincaré's Last Geometric Theorem and the Aubry–Mather Theorem imply the existence of periodic and quasi-periodic crystals for any potential. For sufficiently flat potentials, the KAM theorem yields one-parameter families of crystals between which there is no "energy barrier," i.e., the deposited layer of atoms can slide freely along the underlying surface. [See Aubry (1983) and Le Daeron, and Bangert (1988).]

Finally, Poincaré's Last Geometric Theorem and the Aubry–Mather Theorem imply the existence of periodic and quasi-periodic orbits in many different scenarios. As discussed in Chapter IX, the KAM theorem provides stability statements when the problem is nearly integrable, or near (most) linearly stable periodic orbits. The reader should consult Conley (1962), Moser (1973), Arnold and Avez (1968), and Chapter IX for more information.

G. Further Reading

The following comments are meant as a guide to the very incomplete list of references which follows. A "complete" list of references would be impossible to compile and the author hopes the readers will help fill in the obvious gaps in this list by consulting the references of these references and watching the current literature for more results.

Section A

Because of the small number of variables (two dimensions) and the simple geometric nature of the twist condition, twist maps arise in many places in dynamics. See Birkhoff (1915) for a lurid account of the appearance of twist maps in the planar circular restricted 3-body problem found by Poincaré (1890, 1899). See J. Moser (1973), C.C. Conley (1962), Chenciner and Llibre (1988) for a few of the recent applications of twist maps to celestial mechanics. Charming problems of billiards and "exterior" billiards and their relationship to twist maps is discussed in Birkhoff (1927) and Moser (1973), respectively, and the appearance of twist maps in general Hamiltonian systems is discussed in Arnold and Avez (1968). There is a great wealth of stimulating computer pictures available in the literature, for example Chirkov (1979), and a surprising application of twist maps in solid state physics is lucidly described in the excellent survey article of Bangert (1988).

The approach taken in this section is "topological." Many of the same results, and more, were first discovered by "variational" methods. These techniques were used by Birkhoff (1927) in the billards problem and were generalized to twist maps by Aubry and Le Daeron (1983), and Mather (1982) [see also Katok (1982) and Bangert (1988)]. These variational techniques have led to many additional results [e.g., Mather (1986, 1988)].

Surprisingly, the same variational techniques are useful in the study of geodesies on the 2-torus (even though there is strictly speaking no twist map present). These techniques were used by Hedlund (1932) and are clearly discussed by Bangert (1988).

The topological approach to twist maps can be found in, among others, Hall (1984, 1989), Boyland and Hall (1987), Boyland (1988), Le Calvez (1988) and Casdagli (1987). See also the last chapter of a recent textbook by Arrowsmith and Place (1990).

Several methods of generalization to higher dimensions have recently been given by Angenent (1990) and Moser (1986a). The generalizations to higher-dimensional symplectic maps and Hamiltonian systems is an area of very active present research—a small sample of the interesting work in this area includes Conley and Zehnder (1984), Bernstein and Katok (1987), Floer and Zehnder (1985), Golé (1990), and Mather and Herman (1988).

Section B

The "standard family" of maps of the cylinder [e.g., Equation (B.4)] has been extensively studied both numerically and analytically. Recent work of Veerman and Tangerman (1990) for example shows that the special form of standard families allows one to prove remarkable results concerning the exact behavior of the dynamics of these maps. Herman (1983) has used the special form of standard families to construct some truly amazing examples of twist maps containing remarkable invariant sets.

These maps, particularly "the" standard map have also been extensively studied via computer. Since the questions one asks are frequently about long term behavior of orbits or long period periodic orbits, such studies must be done with great care (and hence great amounts of computer time); see, for example, MacKay and Percival (1985). Recent work of Jungries (1988) has used the special geometric properties of twist maps to circumvent the need for tremendous numerical accuracy when using the computer to address questions about the long-term behavior of orbits. This is an excellent piece of work on twist maps and on how numerical computations can be effectively used in mathematics.

The study of diffeomorphisms of the circle makes an excellent model for the study of twist maps. Indeed, many of the difficulties encountered in studying circle maps, e.g., "small divisors," appear again in annulus maps. This topic is covered in several texts, including Devaney (1986), Nitecki (1971) and Coddington and Levinson (1955) and the reader should consult Guckenheimer and Holmes (1983), Herman (1983), and Arnold (1961) as well.

Section C

Poincaré's Last Geometric Theorem has a fascinating history. Poincaré's paper of 1912 where he conjectures the theorem [and comes close to proving it, see Golé and Hall (1990)] gives great insights into how he thought about the problem. [See also a letter concerning this paper in the last volume of Poincare's *Collected Works.*] The proof given by Birkhoff (1913) the next year used modern "index theory" techniques, and subsequent proofs have used many different methods. Work continues in extending this theorem including recent, very general versions of fixed and periodic point theorems in two dimensions by Franks (1988) and Handel and others. Extensions to higher-dimensional symplectic maps [the "Arnold conjecture"—see Arnold (1978)] have been obtained by Conley and Zehnder (1984) and have opened the gates for a great deal of work in this area (see comments for Section A).

Poincaré originally conjectured the existence of two periodic orbits. Birkhoff's 1913 paper shows the existence of two nondegenerate periodic orbits. In 1925, Birkhoff showed the existence of two periodic orbits, dropping the nondegeneracy assumption. This work has been reviewed and carefully verified by Brown and von Newmann (1977). These multiplicity statements carry into higher dimensions as seen in the work of Conley and Zehnder.

Section D

The ideas of this section are closely related to the concept of unremovable periodic orbits of Asimov and Franks (1983) and the nature of mapping classes in two dimensions [e.g., Matsuoka (1983) and Fathi, Landerback, and

Ponenaru (1979)]. The monotone twist condition gives an extremely strong tool to use in conjunction with the above theories of maps of two-dimensional surfaces. Because the techniques are topological, they apply to cases where area preservation does not hold; for example, see Casdegli (1987), Le Calvez (1988), Boyland (1988), and Arrowsmith and Place (1990).

Section E

The term "zone of instability" was used by Birkhoff to describe a region in an ESMT map which contains no invariant circles. Such a map was constructed by Birkhoff (1932) by the now familiar technique of creating a transverse homoclinic orbit by perturbing an integrable map of the annulus. The dynamics, particularly the ergodic (measure theoretic) properties of orbits in the zones of instability is still an active area of research. Of particular interest is the "transmit time" or the number of iterates it takes for an orbit to pass from a neighborhood of one boundary to a neighborhood of the other; see MacKay, Melss, and Percival (1984).

The importance of delicate number theory in the study of twist maps, particularly the nature of irrational numbers and the extent that they can be approximated by rationals with a small denominator is one of the most interesting and elegant parts of the KAM theory. [A review of the relevant number theory can be found in Hardy and Wright (1979) and Niven (1956).] These "small division" problems that arise from the number theory were known classically and were a major stumbling block to the study of twist maps. Again we note that the study of circle homeomorphisms have many of these same difficulties and the reader is referred to Arnold (1961) and Herman (1979).

Since the KAM theory involves both delicate number theoretic and differentiability hypothesis, it is not surprising that its proof is delicate and difficult. Work continues in the understanding of this theorem and the reader is referred to Moser (1973), Arnold and Avez (1968) for an overview and to Herman (1983) for recent work. That the KAM theory is related to the equally difficult C^1-Closing Lemma of Pugh and Robinson (1983) perhaps does not help to "explain" either of them, but is hopefully surprising and interesting.

Problems

1. Show that the billiard map of Section V.B. satisfies the monotone twist condition.
2. Show that the symplectic map arising in the one-dimensional crystal model of Section V.B. satisfies the monotone twist condition.
3. Show that an area-preserving map on the closed annulus A automatically satisfies the "zero flux" condition of Equation (B.6).

4. Suppose $f: \mathscr{A} \to \mathscr{A}$ is an exact symplectic monotone twist map. Let $B = \{(x, x') \in \mathbb{R}^2 : f(\{(x, y): y \in [0, 1]\}) \cap \{(x', y): y \in [0, 1]\} \neq 0\}$ and define $h: B \to \mathbb{R}$ by setting $h(x, x')$ to be the area bounded by $y = 0$, $\{(x', y): y \in [0, 1]\}$ and $f(\{(x, y): y \in [0, 1]\})$. Show that

(i) $h(x + 1, x' + 1) = h(y, x')$,
(ii) h is a generating function for f [see Katok (1982)].

5. For $f: \mathscr{A} \to \mathscr{A}$ an exact symplectic monotone twist map show that the set of invariant circles is closed (i.e., the set of points in $\mathscr{A}$ on an invariant circle for f is a closed set.)

6. Can a point in $\mathscr{A}$ be on more than one invariant circle for an exact symplectic monotone twist map $f: \mathscr{A} \to \mathscr{A}$? If so, how?

7. Suppose $f: \mathscr{A} \to \mathscr{A}$ is an exact symplectic monotone twist map and f has a 2/5-nonmonotone periodic point. What is the largest interval of rotation numbers for which f is guaranteed to have no invariant circles?

References

Abraham, R. and Marsden, J. 1978. *Foundations of Mechanics.* Benjamin Cummings, London.

Alfriend, J.

1970: The stability of the triangular Lagrangian points for commensurability of order 2, *Celest. Mech.* **1**, 351–359.

1971: Stability of and motion about L_4 at three-to-one commensurability, *Celest. Mech.* **4**, 60–77.

Angenent, S.B. 1990. Monotone recurrence relations, their Birkhoff orbits and topological entropy, *Erg. Theory Dyn. Sys.* **10**(1), 15–42.

Anosov, D.V. 1967. Geodesic flows on compact Riemannian manifolds of negative curvature, *Proceedings of the Steklov Mathematics Institute* **90**, 1–235.

Arenstorf, R.F.

1963: Periodic solutions of the restricted three-body problem representing analytic continuation of Keplerian elliptic motions, Amer. J. Math. **85**, 27–35.

1968: New periodic solutions of the plane three-body problem corresponding to elliptic motion in the lunar theory, J. Diff. Eqs. **4**, 202–256.

Arnold, V.I.

1961: Small denominators I: on the mappings of a circle into itself, Translations of the A.M.S., 2nd Series, 46, pp 213–84.

1963a: Proof of A.N. Kolmogorov's theorem on the preservation of quasiperiodic motions under small perturbations of the Hamiltonian, *Russ. Math. Surv.* **18**(5), 9–36.

1963b: Small divisor problems in classical and celestial mechanics, *Russ. Math. Surv.* **18**(6), 85–192.

1964: Instability of dynamical systems with several degrees of freedom, Sov. Math. Dokl. **5**, 581–585.

1978: *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York.

1983: *Geometric Methods in the Theory of Ordinay Differential Equations*, Springer-Verlag, New York.

Arnold, V.I. and Avez, A. 1968. *Ergodic Properties of Classical Mechanics*, Benjamin, New York.

Arrowsmith, D.K. and Place, C.M.

1982: *Ordinary Differential Equations*, Chapman and Hall, London.

1990: *Introduction to Dynamical Systems*, Clarendon Press, London.
Asimov, D. and Franks, J. 1983. Unremovable periodic orbits, in *Geometric Dynamics*, ed. J. Palis, Springer Lecture Notes in Math. No. 1007, Springer-Verlag, New York.
Aubry, S. and Le Daeron, P.Y. 1983. The discrete Frenkal–Kontorova model and the devil's staircase, *Physica* **7D**, 240–258.
Bangert, V. 1988: Mather sets for twist maps and geodesics on tori, *Dynamics Reported* **1**, 1–56.
Barrar, R.B. 1965. Existence of periodic orbits of the second kind in the restricted problem of three bodies, Astron. J. **70(1)**, 3–4.
Bernstein, D. and Katok, A. 1987. Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with convex Hamiltonian, Invent. Math. **88**, 225–241.
Besicovitch, A.S. 1932. *Almost Periodic Functions*, Cambridge University Press.
Birkhoff, G.D.
1913: Proof of Poincaré's last geometric theorem, *Trans. A.M.S.* **14**, 14–22.
1915: The restricted problem of three bodies, *Rend. Circolo. Mat. Palermo* **39**, 255–334.
1925: An extension of Poincaré's last geometric theorem, *Acta Math.* **47**, 297–311.
1927: *Dynamical Systems*, American Mathematics Society Providence, RI.
1932: Sur l'existence de regions d'instabilité en dynamique, *Annales L'Inst. H. Poincaré* **2**, 369–386.
Bogoliubov, N. and Mitropolskii, Y. 1961. *Asymptotic Methods in the Theory of Nonlinear Oscillations*, Gordon and Breach, New York.
Bohr, H. 1959. Almost Periodic Functions, Chelsea Publishers, New York.
Bowen, R.
1975: *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphism*, Lecture Notes in Mathematics, Springer-Verlag, New York.
1977: *On Axiom A Diffeomorphisms*, Conference Board of Mathematical Sciences 35, American Mathematical Society, Providence, RI.
Boyland, P.L. 1988: Rotation sets and Morse decompositions for twist maps, *Erg. Theory Dyn. Sys.* **8***, 33–62.
Boyland, P.L. and Hall, G.R. 1987. Invariant circles and the orbit structure of periodic orbits in monotone twist maps, *Topology* **26**(1), 21–35.
Brauer, F. and Nohel, J. 1969. *The Qualitative Theory of Ordinary Differential Equations*, Benjamin, New York.
Brown, M., and von Newmann, W.D. 1977. Proof of the Poincaré–Birkhoff fixed point theorem, *Michigan Math. J.* **24**, 21–31.
Bruno, A.D.
1970a: Instability in Hamiltonian systems and distribution of asteroids, *Math. USSR Sbornik* **12**(2), 271–312.
1970b: The normal form of nonlinear oscillations (in Russian), *Proc. V. Internat. Conf. on Nonlinear Osc.*, MIR, Kiev.
1972: Researches on the restricted three-body problem I; Periodic solutions of a Hamiltonian system (in Russian), Preprint 18, Institute of Applied Mathematics, Moscow.
1974: Sets of analytisity of normalizing transformations (in Russian), Institute of Applied Mathematics, Moscow.
1979: *Local Method in Nonlinear Differential Equations*, Springer-Verlag, New York.
1988: The normal form of a Hamiltonian system, *Russ. Math. Surv.* **43**(1), 25–66.
Buchanan, D. 1941. Trojan satellites—limiting case, *Trans. R. Soc. Canada* **35**, 9–25.
Burgoyne, N. and Cushman, R.
1973: Normal forms for real linear Hamiltonian systems with purely imaginary eigenvalues, *Celest. Mech.* **8**, 435–443.

1977: Conjugacy classes in linear groups, *J. Algebra* **44**, 359–362.
Casdagli, M. 1987. Periodic orbits for dissipative twist maps, Erg. Theory *Dyn. Sys.* **7**, 165–173.
Chenciner, A. and Llibre, J. 1988. A note on the existence of invariant punctured tori in the planar circular restricted three-body problem, *Erg. Theory Dyn. Sys.* **8**, 63–72.
Cherry, T.M. 1926. On periodic solutions of Hamiltonian systems of differential equations, *Phil. Trans. R. Soc. A* **227**, 137–221.
Chrikov, B.V. 1979: A universal instability of many dimensional oscillator systems, *Phys. Rep.* **52** (5).
Coddington, E. and Levinson, N. 1955. *Theory of Ordinary Differential Equations*, McGraw-Hill, New York.
Conley, C.C.
1962: On some new long periodic solutions of the plane restricted three-body problem, *Commun. Pure Appli. Math.* **XVI**, 449–467.
1969: On the ultimate behavior of orbits with respect to an unstable critical point, *J. Diff. Eqs.* **5**, 136–158.
1970: Invariant sets which carry a one form, *J. Diff. Eqs.* **8**, 587–594.
Conley, C.C. and Zehnder, E. 1984. The Birkhoff–Lewis fixed point theorem and a conjecture of V.I. Arnold, *Invent. Math.* 73, 33–49.
Contopoulos, G. 1981. The 4 : 1 resonance, *Celest. Mech.* **24**(4), 355–366.
Courant, R. and Hilbert, D. 1953: Methods of Mathematical Physics, John Wiley, New York.
Cushman, R., Deprit, A., and Mosak, R. 1983. Normal forms and representation theory, *J. Math. Phys.* **24**(8), 2102–2116.
Danby, J.M.A. 1962. *Fundamentals of Celestial Mechanics*, Macmillan, New York.
Deprit, A. 1969. Canonical transformation depending on a small parameter, *Celest. Mech.* **72**, 173–179.
Deprit, A. and Deprit-Bartholomê, A. 1967. Stability of the Lagrange points, *Astron. J.* **72**, 173–179.
Deprit, A. and Henrard, J.
1968: A manifold of periodic solutions, *Advan. Astron. Astrophys.* **6**, 6–124.
1969: Sur les orbites periodiques issues de L_4 a la resonance interne 1/4, *Astron. Astrophys.* **3**, 88–93.
1970: The Trojan manifold—survey and conjectures in periodic orbits stability and resonances, ed. G. Giacaglia, Reidel Publishing, Dordrecht.
Devaney, R. 1986. *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings, Menlo Park, CA.
Dieudonne, J.A. 1960. *Foundations of Modern Analysis*, Academic Press, New York.
Diliberto, S.P.
1961: Perturbation theorems for periodic systems, *Circ. Mat. Palermo* **9**(2), 265–299, **10**(2), 111–112.
1967: New results on periodic surfaces and the averaging principle, *Diff. Integral Eqs.*, Benjamin, New York, pp. 49–87.
Elphick, C. Tirapegui, E. Brachet, M., Coullet, P., and Iooss, G. 1987. A simple global characterization for normal forms of singular vector fields, *Physica* **29D**, 96–127.
Fathi, A., Landerbach, F., and Ponenaru, V. 1979. Travaux de Thurston sur les surfaces, *Asterique* 66–67.
Favar, J. 1933. *Lecons sur les fonctions presque periodiques*, Gauthier-Villars, Paris.
Fink, A.M. 1975. *Almost Periodic Differential Equations*, Lecture Notes in Mathematics, Springer-Verlag, New York.
Flanders, H. 1963. *Differential Forms with Applications to the Physical Sciences*, Academic Press, New York.
Floer, A. and Zehnder, E. 1985. Fixed points results for symplectic maps related to the Arnold conjecture, *Dynamical Systems and Bifurcation*, Springer-Verlag, New York.

Fong, U. and Meyer, K.R. 1975. Algebras of integrals, *Revista Colombiana Math.* **IX**, 75–90.

Franks, J.

1979: Manifolds of C^r mappings and applications to differentiable dynamical systems, *Advances in Mathematics Supplementary Studies 4*, Acadermic Press, New York, pp. 271–290.

1988: Recurrence and fixed points in surface homeomorphisms, *Erg. Theory Dyn. Sys.* **8**, 99–108.

Golé, C. 1990: Monotone maps of $\pi^n \times \mathbb{R}^n$ and their periodic orbits, and Ghost tori for monotone maps, IMA preprints #677 and 678, Minneapolis.

Golé, C. and Hall, G.R. 1990: Poincaré's proof of Poincaré's last geometric theorem, IMA preprint #687, Minneapolis.

Golubitsky, M. and Schaeffer, D. 1986. *Singularities and Groups in Bifurcation Theory I*, Springer-Verlag, New York.

Groetsch, C.W. 1980. *Eelements of Applicable Functional Analysis*, Dekker, New York.

Guckenheimer, J. and Holmes, P. 1983. *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*, Springer-Verlag, New York.

Hadjidemetriou, J.D. 1975. The continuation of periodic orbits from the restricted to the general three-body problem, *Celest. Mech.* **12**, 155–174.

Hale, J.K. 1972. *Ordinary Differential Equations*, John Wiley, New York.

Hale, J.K. and Chow, S-N. 1982. *Methods of Bifurcation Theory*, Springer-Verlag, New York.

Hall, G.R.,

1984: A topological version of a theorem of Mather on twist maps, *Erg. Theory Dyn. Sys.* **4**.

1989: A topological version of a theorem of Mather on shadowing in twist maps, *Dynamical Systems and Ergodic Theory*, Banach Center Publication 23, Banach Center, Warsaw.

Halmos, P. 1958. *Finite Dimensional Vector Spaces*, Van Nostrand, Princeton, NJ.

Hardy, G.H. and Wright, E.M. 1979. *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, London.

Hartman, P. 1964. *Ordinary Differential Equations*, John Wiley, New York.

Hedlund, G.A. 1932: Geodesies on a two-dimensional torus with periodic coefficients, *Ann. Math.* **33**, 719–739.

Henrard, J.

1970a: Concerning the genealogy of long period families at L_4, *Astron. Astrophys.* **5**, 45–52.

1970b: Periodic orbits emanating from a resonant equilibrium, Celest. *Mech.* **1**, 437–466.

1970c: On a perturbation theory using Lie transforms, *Celest. Mech.* **3**, 107–120.

1973: Liapunov's center theorem for resonant equilibrium, J. *Diff. Eq.* **14**(3), 431–441.

Herman, M.R.

1979: Sur la conjugaison differentiable des diffeomorphismes du cerde, *I.H.E.S. Math.* **49**.

1983: Sur les courbes invariantes par les diffeomorphismes de l'anneau I, *Asterisque* **103**, 3–221.

1986: Sur les courbes invariantes par les diffeomorphisms de l'anneau II, *Asterisque* **144**, 3–221.

1988: Existence et nonexistence de tores invariant par des diffeomorphismes symplectiques, preprint, Ecole Polytechnique.

Hill, G.W. 1878. Researches in the lunar theory, *Am. J. Math.* 5–26, 129–147, 245–260.

Hirsch, M. 1976. *Differential Topology*, Springer-Verlag, New York.

Hirsch, M., Pugh, C., and Shub, M. 1977. *Invariant Manifolds*, Springer-Verlag, New York.

Hori, G. 1966. Theory of general perturbations with unspecified canonical variables, *Publ. Astron. Soc. Japan* **18**(4), 287–296.

Hurewicz, W. 1958: *Lectures on Ordinary Differential Equations*, MIT Press, Cambridge, MA.

Jungries, I. 1988. A method for proving that monotone twist maps have no invariant circles, preprint, Harvard University.

Kamel, A. 1970. Perturbation method in the theory of nonlinear oscillations, *Celest. Mech.* **3**, 90–99.

Katok, A. 1982. Some remarks on Birkhoff and Mather twist theorems, *Erg. Theory Dyn. Sys.* **2**.

Kelley, A. 1967. On the Liapunov subcenter theorem, *J. Math. Anal. Appl.* **18**, 472–478.

Kelley, J. 1955. *General Topology*, Van Nostrand, Princeton, NJ.

Kirchhoff, G. 1897. Vorlesunqen über Mechanik, Druck und Verlag, Leipzig.

Kobayashi, S. and Nomizu, K. 1969. *Foundations of Differential Geometry*, Interscience, New York.

Kolmogorov, A.

1954: On conservation of conditionally periodic motions under small perturbations of the Hamiltonian, Dokl. Akad. Nauk. SSSR 98, 527–530.

1957: General theory of dynamical systems and classical mechanics, Proc. 1954 Intern. Congress Math., North Holland, 315–333. (Russian)

Kummer, M.

1976: On resonant non linearly coupled oscillators with two equal frequencies, *Commun. Math. Phys.* **48**, 53–79.

1978: On resonant classical Hamiltonians with two equal frequencies, *Comm. Math. Phys.* **58**, 85–112.

Laloy, M. 1976: On equilibrium instability for conservative and partially dissipative mechanical systems, Int. J. Non-linear Mech. 2, 295–301.

LaSalle, J.P. and Lefschetz, S. 1961. *Stability by Liapunov's Direct Method with Applications*, Academic Press, New York.

Laub, A. and Meyer, K.R. 1974. Canonical forms for symplectic and Hamiltonian matrices, *Celest. Mech.* **9**, 213–238.

Le Calvez, P. 1988. Properties des attracteurs de Birkhoff, *Erg. Theory Dyn. Sys.* **8**, 241–310.

Lefschetz, S. 1963. *Differential Equations: Geometric Theory*, Wiley, New York.

Leung, A. and Meyer, K.R. 1975. Adiabatic invariants for linear Hamiltonian systems, *J. Diff. Eqs.* **17**(1), 32–43.

Lichtenberg, A.J. and Lieberman, M.A. 1983. *Regular and Stochastic Motion*, Springer-Verlag, New York.

Liu, J.C. 1985. The uniqueness of normal forms via Lie transforms and its applications to Hamiltonian systems, *Celest. Mech.* **36**(1), 89–104.

Lyapunov, A. 1892. Probleme général de la stabilité du mouvement, *Ann. Math. Studies 17*, Princeton University Press, Princeton, NJ. (Reproduction in 1947 of the French translation.)

MacKay, R.S., Meiss, J.D., and Percival I.I.C. 1984. Stochasticity and transport in Hamiltonian systems, *Phys. Rev. Lett.* **52**, 697–700.

MacKay, R.S. and Percival I.I.C. 1985. Converse KAM theory and practice, *Comm. Math. Phys.* **98**, 469–512.

Malkin, I.G. 1952: Theory of stability of motion, Office of Technical Services, Department of Commerce, Washington, DC.

Marsden, J. and Weinstein, A. 1974: Reduction of symplectic manifold with symmetrices, Rep. Math. Phys. 5, 121–130.

Markeev, A.P.
1966: On the stability of the triangular libration points in the circular bounded three-body problem, *Appl. Math. Mech.* **33**, 105–110.
1978: *Libration Points in Celestial Mechanics and Space Dynamics* (in Russian), Nauka, Moscow.
Markus, L. and Meyer, K.
1974: Generic Hamiltonian systems are neither integrable nor dense, *Mem. Am. Math. Soc. 144.*
1980: Periodic orbits and solenoids in generic Hamiltonian systems, *Am. J. Math.* **102**, 25–92.
Mather, J.N.
1982: Existence of quasi-periodic orbits for twist homeomorphisms of the annulus, *Topology* **21**, 457–467.
1986: A criterion for the nonexistence of invariant circles, *I.H.E.S. Math* **63**.
1988: Destruction of invariant circles, *Erg. Theory Dyn. Sys.* **8***, 199–214.
Mathieu, E. 1874. Mémorire sur les équations différentieles canoniques de la mácanique, *J. Math. Pures Appl.* **39**, 265–306.
Matsuoka, T. 1983. The number and linking of periodic solutions of periodic systems, *Invent. Math.* **70**, 319–340.
McGehee, R. and Meyer, K.R. 1974. Homoclinic points of area preserving diffeomorphisms. *Am. J. Math.* **96**(3), 409–421.
Meyer, K.R.
1967: On contact transformations, *Am. J. Phys.* **35**(10), 970–971.
1969: Subharmonic solutions in Hamiltonian systems, *Seminar in Differential Equations and Dynamical Systems II*, Lecture Notes in Mathematics 144, Springer-Verlag, New York, 160–168.
1970: Generic bifurcation of periodic points, *Trans. Am. Math. Soc.* **149**, 95–107.
1971: Generic stability properties of periodic points, *Trans. Am. Math. Soc.* **154**, 273–277.
1973: Symmetries and integrals in mechanics, *Dynamical Systems*, ed. M. Peixoto, Academic Press, New York, pp. 259–272.
1974: Normal forms for Hamiltonian systems, *Celest. Mech.* **9**, 517–522.
1975: Implicit function theorem and analytic differential equations, *Dynamical Systems—Warwick 1974*, Lecture Notes in Mathematics 468, Springer-Verlag, New York, pp. 62–70.
1981a: Periodic orbits near infinity in the restricted N-body problem, *Celest. Mech.* **23**, 69–81.
1981b: Periodic solutions of the N-body problem, *J. Diff. Eqs.* **39**(1), 2–38.
1981c: Hamiltonian systems with a discrete symmetry, *J. Diff. Eqs.* **41**(2), 228–238.
1981d: An application of Poincarè's recurrence theorem to academic administration. *Am. Math. Monthly* **88**(1), 32–33.
1984a: Scaling Hamiltonian systems, *SIAM J. Math. Anal.* **15**(5), 877–889.
1984b: Normal forms for the general equilibrium, *Funkcial. Ekva.* **27**(2), 261–271.
1986a: Counter-examples in dynamical systems via normal form theory, *SIAM Rev.* **28**(1), 41–51.
1986b: Bibliographic notes on generic bifurcations in Hamiltonian systems, *Multiparameter Bifurcation Theory*, Contemporary Mathematics 56, eds. *M. Golubitsky and J. Guckenheimer*, American Mathematical Society, Providence, RI.
1987: Bifurcation of a central configuration, *Celest. Mech.* **40**(3), 273–282.
1990: The geometry of harmonic oscillators, *Am. Math. Monthly*, 97, 457–465, 190–210.
1991: A Lie transform tutorial I, Computer Aided Proofs in Analysis, Springer-Verlag, New York.

Meyer, K.R. and Palmore, J.I.
1970a: A new class of periodic solutions in the restricted three-body problem, *J. Diff. Eqs.* **8**(2), 264–276.
1970b: A generic phenomenon in conservative Hamiltonian systems, *Global Analysis, Proceedings of Symposia in Pure Mathematics XIV*, American Math, Soc, Providence, pp. 185–189.
Meyer, K.R. and Schmidt, D.S.
1971: Periodic orbits near L_4 for mass ratios near the critical mass ratio of Routh, *Celest. Mech.* **4**, 99–109.
1982a: Hill's lunar equation and the three-body problem. *J. Diff. Eqs.* **44**(2), 263–272.
1982b: The determination of derivatives in Brown's lunar theory, *Celest. Mech.* **28**, 201–207.
1986a: The stability of the Lagrange triangular point and a theorem of Arnold, *J. Diff. Eqs.* **62**(2), 222–236.
1988a: Bifurcations of relative equilibria in the 4 and 5 body problems, *Erg. Dyn. Sys.*, 8, 215–225.
1988b: Bifurcations of relative equilibria in the N-body and Kirchhoff problems, *SIAM J. Math. Analy.*, 19, 1295–1313.
Meyer, K.R. and Sell, G.R.
1978: An analytic proof of the shadowing lemma, *Funkcial. Ekvac.* **30**, 127–133.
1989: Melnikov transformations, Bernoulli bundles, and almost periodic perturbations, *Trans. Am. Math. Soc.* **314**(1), 63–105.
Milnor, J. 1965. *Topology from the Differentiable Viewpoint*, The University of Virginia Press, Charlottesville, VA.
Moser, J.K.
1956: The analytic invariants of an area preserving mapping near a hyperbolic fixed point, Comm. Pure Appl. Math. 9, 673–692.
1958: New aspects in the theory of stability of Hamiltonian systems, *Comm. Pure Appl. Math.* **11**, 81–114.
1961: On the integrability of area-preserving canonical mappings near an elliptic fixed point, *Bol. Soc. Mat. Mexicana*, 16, 167–175.
1962: On invariant curves of area-preserving mappings of an annulus, *Nachr. Akad. Wiss. Gottingen Math. Phys.* Kl. 2, 6, 1–20.
1968: Lectures on Hamiltonian systems, *Mem. Am. Math. Soc.* **81**, 1–60.
1969: On a theorem of Anosov, *J. Diff. Eqs.* **5**.
1973: *Stable and Random Motion in Dynamical Systems*, Princeton University Press, Princeton, NJ.
1978: A fixed point theorem in symplectic geometry, *Acta Math.* **141**, 17–34.
1979: Periodic orbits near an equilibrium and a theorem of Alan Weinstein, *Comm. Pure Appl. Math.* **XXIX**, 727–747.
1986a: Minimal solutions of variational problems on a torus, *Ann. Inst. H. Poincaré, Non Lineaire* **3**, 229–272.
1986b: Monotone twist mappings and the calculus of variations, *Erg. Theory Dyn. Sys.* **6**, 401–413.
1986c: Recent developments in the theory of Hamiltonian systems, *SIAM Rev.* **28**(4), 459–485.
Moulton, F.R.
1914: *An Introduction to Celestial Mechanics*, Macmillan, New York.
1920: *Periodic Orbits*, Carnegie Institution of Washington, Washington, DC.
Naylor, A.W. and Sell, G.R. 1982. *Linear Operator Theory in Engineering and Science*, Springer-Verlag, New York.
Nemytskii V. and Stepanov, V. 1960. *Qualitative Theory of Differential Equations*, Princeton University Press, Princeton, NJ.

Nitecki, Z. 1971. *Differentiable Dynamics*, MIT Press, Cambridge, MA.
Niven, I. 1956. *Irrational Numbers*, Carus Monograph #11, M.A.A.
Palis, J. and de Melo, W. 1980. *Geometric Theory of Dynamical Systems*, Springer-Verlag, New York.
Palmore, J.I. 1969. *Bridges and Natural Centers in the Restricted Three-Body Problem*, University of Minnesota Report, University of Minnesota.
Peckham, B. 1989. The necessity of Hopf bifurcation for periodically forced oscillators, *Nonlinearity*, in press.
Poincaré, H.
1885: Sur les courbes definies par les equations differentielles, *J. Math. Pures Appl.* **4**, 167–244.
1890: Sur les probleme des trois corp et les equations dynamique, *Acta Math.* **13**.
1899: *Les methods nouvelles de la mechanique celeste*, Gauthier-Villar, Paris.
1912: Sur un theoreme de geometrie, *Rand. Circ. Math. Palermo*, **33**, 375–407.
Poinhoff, H.G. 1973. An analytic closing lemma, *Proceedings of the Midwest Dynamical Systems Seminar*, Northwestern University Press, Evanston, IL, pp. 128–256.
Pollard, H. 1966. *Mathematical Introduction to Celestial Mechanics*, Prentice-Hall, Englewood Cliffs, NJ.
Poston, T. and Stewart, I. 1978. *Catastrophe Theory and its Applications*, Pitman, Boston.
Pugh, C. and Robinson, C. 1977. The C^1 closing lemma, including Hamiltonians, *Erg. Theory Dyn. Sys.* **3**. 261–313.
Rimmer, R.
1978: Symmetry and bifurcation of fixed points of area preserving maps, *J. Diff. Eqs.* **29**(3), 329–344.
1983: Generic bifurcation for involuntary area preserving maps, *Mem. Am. Math. Soc.* **41**, 1–165.
Robinson, C.
1970: Generic properties of conservative systems I & II, *Am. J. Math.* **92**, 562–603, 897–906.
1971: Generic one-parameter families of symplectic matrices, *Am. J. Math.* **93**, 116–122.
Roels, J.
1971a: An extension to resonant cases of Liapunov's theorem concerning the periodic solutions near a Hamiltonian equilibrium, *J. Diff. Eqs.* **9**(2), 300–324.
1971b: Families of periodic solutions near a Hamiltonian equilibrium when the ratio of the two eigenvalues is 3, *J. Diff. Eqs.* **10**(3), 431–447.
Rolfsen, D 1976: Kuota and Links, Publish or Perish Inc, Berkeley, CA.
Rudin, W. 1973. *Functional Analysis*, McGraw-Hill, New York.
Rüssman, H. 1959. Uber die Existenz einer Normalform inhaltstreuer ellipticsher Transformationen, *Math. Ann.* **167**, 55–72.
Sánchez, D.A. 1968. *Ordinary Differential Equations and Stability Theory*, W.H. Freeman, New York.
Sanders, J. 1977. Are higher order resonances interesting? *Celest. Mech.* **166**(4), 421–440.
Sanders, J. and Verhulst, F.
1979: Approximations of higher order resonances with applications to Contopoulos' model problem, *Asymptotic Analysis*, Lectures in Mathematics 711, Springer-Verlag, New York, pp. 209–228.
1985: *Averaging methods in nonlinear dynamical systems*, Applied Mathematical Sciences, 59, Springer-Verlag, New York.
Schmidt, D.S.
1972: Families of periodic orbits in the restricted problem of three bodies connecting families of direct and retrograde orbits, *SIAM J. Appl. Math.* **22**(1), 27–37.

1974: Periodic solutions near a resonant equilibrium of a Hamiltonian system, *Celest. Mech.* **9**, 91–103.
1989: The stability of the Lagrangian point L_4, *Celest. Mech.* **45**, 201–206.
1990: Transformation to versal normal form, Computer Aided Proofs in Analysis (Ed. K.R. Meyer & D.S. Schmidt), IMA Series 28, Springer-Verlag.

Schmidt, D.S. and Sweet, D. 1973. A unified theory in determining periodic families for Hamiltonian systems at resonance, *J. Diff. Eq.* **14**(3), 597–609.

Sibuya, Y. 1960. Note on real matrices and linear dynamical systems with periodic coefficients, *J. Math. Anal. Appl.* **1**, 363–372.

Siegel, C.L. 1956. *Vorlesungen uber Himmelsmechanik*, Springer-Verlag, Berlin.

Siegel, C.L. and Moser, J.K. 1971. *Lectures on Celestial Mechanics*, Springer-Verlag, Berlin.

Smale, S.
1965: Diffeomorphisms with many periodic points, *Differential and Combinatorial Topology*, Princeton University Press, Princeton, NJ.
1967: Differentiable dynamical systems, *Bull. Am. Math. Soc.* **73**, 747–817.

Sokol'skii, A. 1978: On the stability of an autonomous Hamltonian system, J. Appl. Math. Mech. 38, 741–749.

Spivak, M. 1965: Calculus on Manifolds, W.A. Benjamin, New York.

Sweet, D. 1973. Periodic solutions for dynamical systems possessing a first integral in the resonance case, *J. Diff. Eqs.* **14**(1), 171–183.

Szebehely, V. 1967. *Theory of Orbits*, Academic Press, New York.

Szlenk, W. 1981. *An Introduction to the Theory of Smooth Dynamical Systems*, John Wiley, New York.

Taliaferro, S. 1980: Stability of two dimensional potentials, J. Diff. Eq. 35, 248–265.

Veerman, J. and Tangerman, F.M. 1990. Intersection properties of invariant manifolds of certain twist maps, preprint, SUNY Stonybrook, #1990/7.

Weinstein, A. 1971: Symplectic manifolds and their Lagrangian submanifolds, *Adv. Math.* **6**, 329–346.

Weyl, H. 1946: *Classical Groups*, Princeton University Press, Princeton, NJ.

Willianson, J.
1936: On the algebraic problem concerning the normal forms of linear dynamical systems, *Am. J. Math.* **58**, 141–163.
1937: On the normal forms of linear canonical transformations in dynamics, *Am. J. Math.* **59**, 599–617.
1939: The exponential representation of canonical matrices, *Am. J. Math.* **61**, 897–911.

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

Whittaker, E.T. 1927. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, Cambridge.

Whittaker, E.T. and Watson, G.N. 1958: A Course of Modern Analysis, 4th Ed., Cambridge Univ. Press, Cambridge.

Yakubovich, V.A. and Stazhinskii, V.M. 1975. *Linear Differential Equations with Periodic Coefficients* **1 & 2**, John Wiley, New York.

Index

Action-angle variables, 95–96
Amended potential, 27–28
Angular momentum, 19
Anomaly, true, 30, 100
Anosov system, 145
Aubry-Mather theorem, 256
Autonomous, 3, 109, 152

Bifurcation, 201–226
Billiards, 118–120

Center of mass, 18
Central configuration, 22–26
Characteristic polynomial, 45–46
Christoffel symbols, 32
Comets, 161–162
Conjugate variable, 3
Conservative, 3
Contravariant vector fields, 80
Cotangent vector, 77–79
Covariant vector fields, 80
Covector, 41
Critical point – see equilibrium point
Cross section 129–130

d'Alembert character, 96–97, 175–176
Darboux's theorem, 83–85
Degrees of freedom, 3
Delaunay elements, 104–106, 165–166
Dense orbit, 16
Dual basis, 42
Duffing's equation, 167, 175–176, 180–181, 199–200, 212–216, 225, 232–233
Dynamical system, 109–153
  discrete, 114–115
  Hamiltonian, 110

Einstein's notation, 78–86
Elliptic
  orbits, 164–166
  point, 112–113, 115, 194–199, 229–232
  sine function, 8–10, 29
Energy, 5, 10
  kinetic, 10
  potential, 10, 18
  total, 10
Equilibrium point, 5
  elementary, 129–130, 155–156

Equilibrium point (*cont.*)
N-body, 22
normal form, 182–190
relative, 23—see central configuration
restricted problem, 26–28, 59–67, 189–190, 218–224
stable—see stable, equilibrium
Equivalent, 112, 114
ESMT, 246
Euler-Moulton central configuration, 25–26
Euler's equation, 31
Exact symplectic, 243
Expansive, 152
Exponential dichotomy—see hyperbolic structure
Exterior derivative, 81–83

Fixed point, 114
Floquet-Lyapunov theorem, 53–56
Flow—see dynamical system
Flow box theorem, 120
Form
differential, 79–83
multilinear, 73–76
one (1-form), 41
symplectic, 79–83
Fundamental matrix solution, 36

General solution, 3
Generating functions, 97–99, 224
Geodesic, 30
Group, 70
general linear, 33, 70
Lie, 70–71, 153
orthogonal, 70, 127
symmetry, 153
symplectic, 70, 127

Henon map, 115–116
Hill's lunar problem, 32, 71, 167
Hill's orbits, 159–161
Homoclinic point, 140–141, 153
Hopf fibration, 15
Hyperbolic
point, 112–113, 115
set, 143
structure, 143

Ignorable coordinate, 93
Integrals, 4
classical, 18–19
local, 120–125
Invariant curve, 229–232, 248, 267
Invariant set, 4
Involution, 38, 123

Jacobi
coordinates, 92–94, 128
elliptic sine function, 8–10, 29
Jacobian, 2
J-orthogonal, 46–49

KAM Theorem, 227, 229–232, 273
Kepler problem, 19–20, 30, 99–101, 158–159, 165
Kirchhoff problem, 17, 30
Knot, 15

Lagrangian
function, 30
set of solutions, 38
subspace, 43–44
triangular solutions, 24–25, 61, 157–158, 189, 218–224, 238–239
λ-lemma, 140
Libration point, 27
Lie
group—see group
product, 29, 35
transforms, 168–199
Lie algebra, 29
Linear momentum—see momentum

Linear Hamiltonian system, 34–71
Logarithm of a matrix, 54–56, 67–69, 71
Lyapunov's center theorem, 156–157, 226

Matrix
  determinant of, 39, 53, 77
  Hamiltonian, 34–71
  matrix, 35–71
  monodromy, 54
  polar decomposition, 39
  symplectic, 35–71
Möbius transformation, 69
Moment of inertial, 23
Momentum
  angular, 93
  linear, 18, 93, 152
Monotone orbit, 249
Monotone twist map, 244
Multiplier, 54, 132–136, 152

N-body problem, 17–28, 126
Newtonian system, 10
Noether's theorem, 125–129
Nonelementary divisors, 50–53
Normal form, 168–199

One-form (1-form)—see forms
Orbit 110, 158–162
Oscillator
  harmonic, 5, 11–16
  nonlinear, 7

Pendulum equation, 29
Period map, 116–119
Periodic point, 114
  bifurcation, 209–212
  elementary, 202–203
  extremal, 203–205
  transitional, 206–208
Periodic solution, 9, 11–17, 130–136, 154–167
  bifurcation, 202–226
  elliptic, 156, 195
  hyperbolic, 157, 195
  normal form, 191–199
  Schmidt, 216–218
Phase space, 4
Poincaré
  elements, 106–107, 240
  last geometric theorem, 253
  lemma, 82, 86
  map, 131–136
  orbits, 158–159
Poisson bracket, 4, 20, 90–91
Polar coordinates, 99, 158
Pseudo-orbit, 146–152

Quasi-periodic solution, 12

Reality condition, 48
Reciprocal polynomial, 45
Reduced space, 125–129, 159–160
Remainder function, 36–37, 172–174
Restricted three body problem, 20–21, 59–66, 103–104, 162–164, 216–224, 233–235, 238
Rotation number, 247

Scaling, 102–104
Shadowing, 146, 149–152
Shift Automorphism (map), 141–143
Stable and stability, 5, 227–240
  equilibrium point, 5, 10, 227–229, 235–238
  parametric, 56–59
  manifold, 136–141
Standard family, 243
Subshift, 141–143
Symmetries, 125–127
Symplectic
  basis, 41

Symplectic (*cont.*)
coordinates, 87–107
form—see forms
isomorphic, 41
linear spaces, 40–44
matrix—see matrix
similar, 44
transformation, 87–107, 157, 159, 161, 169–199
Symplectomorphism—see symplectic transformation

Tangent vector, 77–79
Torus, 11–16
Trajectory, 110

Unstable manifold—see stable, manifold

Variational equations, 89–90
Vortex, 16

Applied Mathematical Sciences

1. *John:* Partial Differential Equations, 4th ed.
2. *Sirovich:* Techniques of Asymptotic Analysis.
3. *Hale:* Theory of Functional Differential Equations, 2nd ed.
4. *Percus:* Combinatorial Methods.
5. *von Mises/Friedrichs:* Fluid Dynamics.
6. *Freiberger/Grenander:* A Short Course in Computational Probability and Statistics.
7. *Pipkin:* Lectures on Viscoelasticity Theory.
8. *Giacoglia:* Perturbation Methods in Non-linear Systems.
9. *Friedrichs:* Spectral Theory of Operators in Hilbert Space.
10. *Stroud:* Numerical Quadrature and Solution of Ordinary Differential Equations.
11. *Wolovich:* Linear Multivariable Systems.
12. *Berkovitz:* Optimal Control Theory.
13. *Bluman/Cole:* Similarity Methods for Differential Equations.
14. *Yoshizawa:* Stability Theory and the Existence of Periodic Solution and Almost Periodic Solutions.
15. *Braun:* Differential Equations and Their Applications, 3rd ed.
16. *Lefschetz:* Applications of Algebraic Topology.
17. *Collatz/Wetterling:* Optimization Problems.
18. *Grenander:* Pattern Synthesis: Lectures in Pattern Theory, Vol. I.
19. *Marsden/McCracken:* Hopf Bifurcation and Its Applications.
20. *Driver:* Ordinary and Delay Differential Equations.
21. *Courant/Friedrichs:* Supersonic Flow and Shock Waves.
22. *Rouche/Habets/Laloy:* Stability Theory by Liapunov's Direct Method.
23. *Lamperti:* Stochastic Processes: A Survey of the Mathematical Theory.
24. *Grenander:* Pattern Analysis: Lectures in Pattern Theory, Vol. II.
25. *Davies:* Integral Transforms and Their Applications, 2nd ed.
26. *Kushner/Clark:* Stochastic Approximation Methods for Constrained and Unconstrained Systems.
27. *de Boor:* A Practical Guide to Splines.
28. *Keilson:* Markov Chain Models—Rarity and Exponentiality.
29. *de Veubeke:* A Course in Elasticity.
30. *Shiatycki:* Geometric Quantization and Quantum Mechanics.
31. *Reid:* Sturmian Theory for Ordinary Differential Equations.
32. *Meis/Markowitz:* Numerical Solution of Partial Differential Equations.
33. *Grenander:* Regular Structures: Lectures in Pattern Theory, Vol. III.
34. *Kevorkian/Cole:* Perturbation Methods in Applied Mathematics.
35. *Carr:* Applications of Centre Manifold Theory.
36. *Bengtsson/Ghil/Källén:* Dynamic Meteorology: Data Assimilation Methods.
37. *Saperstone:* Semidynamical Systems in Infinite Dimensional Spaces.
38. *Lichtenberg/Lieberman:* Regular and Stochastic Motion.
39. *Piccini/Stampacchia/Vidossich:* Ordinary Differential Equations in $\mathbf{R}^n$.
40. *Naylor/Sell:* Linear Operator Theory in Engineering and Science.
41. *Sparrow:* The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors.
42. *Guckenheimer/Holmes:* Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields.
43. *Ockendon/Taylor:* Inviscid Fluid Flows.
44. *Pazy:* Semigroups of Linear Operations and Applications to Partial Differential Equations.
45. *Glashoff/Gustafson:* Linear Operations and Approximation: An Introduction to the Theoretical Analysis and Numerical Treatment of Semi-Infinite Programs.
46. *Wilcox:* Scattering Theory for Diffraction Gratings.
47. *Hale et al:* An Introduction to Infinite Dimensional Dynamical Systems—Geometric Theory.
48. *Murray:* Asymptotic Analysis.
49. *Ladyzhenskaya:* The Boundary—Value Problems of Mathematical Physics.
50. *Wilcox:* Sound Propagation in Stratified Fluids.
51. *Golubitsky/Schaeffer:* Bifurcation and Groups in Bifurcation Theory, Vol. I.

Applied Mathematical Sciences

52. *Chipot:* Variational Inequalities and Flow in Porous Media.
53. *Majda:* Compressible Fluid Flow and System of Conservation Laws in Several Space Variables.
54. *Wasow:* Linear Turning Point Theory.
55. *Yosida:* Operational Calculus: A Theory of Hyperfunctions.
56. *Chang/Howes:* Nonlinear Singular Perturbation Phenomena: Theory and Applications.
57. *Reinhardt:* Analysis of Approximation Methods for Differential and Integral Equations.
58. *Dwoyer/Hussaini/Voigt (eds):* Theoretical Approaches to Turbulence.
59. *Sanders/Verhulst:* Averaging Methods in Nonlinear Dynamical Systems.
60. *Ghil/Childress:* Topics in Geophysical Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics.
61. *Sattinger/Weaver:* Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics.
62. *LaSalle:* The Stability and Control of Discrete Processes.
63. *Grasman:* Asymptotic Methods of Relaxation Oscillations and Applications.
64. *Hsu:* Cell-to-Cell Mapping: A Method of Global Analysis for Nonlinear Systems.
65. *Rand/Armbruster:* Perturbation Methods, Bifurcation Theory and Computer Algebra.
66. *Hlavácek/Haslinger/Necasl/Lovísek:* Solution of Variational Inequalities in Mechanics.
67. *Cercignani:* The Boltzmann Equation and Its Applications.
68. *Temam:* Infinite Dimensional Dynamical System in Mechanics and Physics.
69. *Golubitsky/Stewart/Schaeffer:* Singularities and Groups in Bifurcation Theory, Vol. II.
70. *Constantin/Foias/Nicolaenko/Temam:* Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations.
71. *Catlin:* Estimation, Control, and the Discrete Kalman Filter.
72. *Lochak/Meunier:* Multiphase Averaging for Classical Systems.
73. *Wiggins:* Global Bifurcations and Chaos.
74. *Mawhin/Willem:* Critical Point Theory and Hamiltonian Systems.
75. *Abraham/Marsden/Ratiu:* Manifolds, Tensor Analysis, and Applications, 2nd ed.
76. *Lagerstrom:* Matched Asymptotic Expansions: Ideas and Techniques.
77. *Aldous:* Probability Approximations via the Poisson Clumping Heuristic.
78. *Dacorogna:* Direct Methods in the Calculus of Variations.
79. *Hernández-Lerma:* Adaptive Markov Processes.
80. *Lawden:* Elliptic Functions and Applications.
81. *Bluman/Kumei:* Symmetries and Differential Equations.
82. *Kress:* Linear Integral Equations.
83. *Bebernes/Eberly:* Mathematical Problems from Combustion Theory.
84. *Joseph:* Fluid Dynamics of Viscoelastic Fluids.
85. *Yang:* Wave Packets and Their Bifurcations in Geophysical Fluid Dynamics.
86. *Dendrinos/Sonis:* Chaos and Socio-Spatial Dynamics.
87. *Weder:* Spectral and Scattering Theory for Wave Propagation in Perturbed Stratified Media.
88. *Bogaevski/Povzner:* Algebraic Methods in Nonlinear Perturbation Theory.
89. *O'Malley:* Singular Perturbation Methods for Ordinary Differential Equations.
90. *Meyer/Hall:* Introduction to Hamiltonian Dynamical Systems and the N-body Problem.
91. *Straughan:* The Energy Method, Stability, and Nonlinear Convection.