LECTURES ON NONLINEAR ORBIT DYNAMICS

Alex J. Dragt
University of Maryland, College Park, Maryland 20742

TABLE OF CONTENTS

)94-243X/82/870147-167$3.00

INTRODUCTION

The treatment of nonlinear effects in orbit dynamics is essential for two reasons. First, nonlinear effects are almost unavoidable. They arise from spatial and temporal inhomogeneities in electric and magnetic fields, from fringe fields, from beam self forces, from the beam-beam interaction in the case of colliding beams, and from higher order kinematic terms arising in expansions of the equations of motion. In this context, it is necessary to deal with nonlinear behavior in order to understand and avoid harmful effects. Second, nonlinear effects can sometimes be exploited to good ends. Current applications include chromaticity control with sextupole magnets, beam stabilization through Landau damping induced by sextupole and octupole magnets, achromatic sections composed of sextupole compensated dipoles, and quadrupoles and nonlinear beam extraction.

Present methods of nonlinear orbit calculation usually involve either Hamiltonian perturbation theory or the direct numerical integration of trajectories. Other important methods include "particle tracking" and extended matrix calculations. Particle tracking approximates nonlinear effects by impulsive momentum kicks. Extended matrix calculations are based on a Green's function method for treating in lowest order the effect of quadratic terms in the equations of motion.

These lectures describe a new approach to the analysis of nonlinear orbit dynamics. Special attention is given to the Hamiltonian nature of the equations of motion, and Lie algebraic tools are developed to exploit this symmetry. It is shown that the use of Lie algebraic concepts provides a concise and powerful method for describing and computing nonlinear effects. Applications are made to charged particle beam transport, light optics, and orbits in circular machines including the case of colliding beams.

1. THE UBIQUITY OF LAGRANGIAN AND HAMILTONIAN DYNAMICS

1.1 Hamilton's Equations with Time as an Independent Variable[1-3]

It is a remarkable discovery that all the fundamental dynamical laws of Nature are expressible in Lagrangian or Hamiltonian form. The relativistic Lagrangian for the motion of a particle of rest mass m_o and charge q in an electromagnetic field is given by the expression

$$L(\vec{r},\vec{v},t) = -m_o c^2(1-\vec{v}^2/c^2)^{1/2} - q\psi(\vec{r},t) + q\,\vec{v}\cdot\vec{A}(\vec{r},t). \quad (1.1)$$

Here ψ and $\vec{A}$ are the scalar and vector potentials defined in such a way that the electromagnetic fields $\vec{E}$ and $\vec{B}$ are given by the standard relations

$$\vec{B} = \vec{\nabla} \times \vec{A}$$
$$\vec{E} = -\vec{\nabla}\psi - \partial\vec{A}/\partial t. \tag{1.2}$$

Exercise 1.1: Lagrange's equations of motion for a system having n degrees of freedom are

$$(d/dt)(\partial L/\partial \dot{q}_i) - (\partial L/\partial q_i) = 0, \tag{1.3}$$

where $(q_1 \ldots q_n)$ is any set of generalized coordinates. In the case that the generalized coordinates are taken to be the usual Cartesian coordinates, verify that Lagrange's equations for the Lagrangian (1.1) reproduce the required relativistic equations of motion for a charged particle under the influence of the Lorentz force.

The momentum p_i canonically conjugate to the variable q_i is defined by the relation

$$p_i = L/\partial \dot{q}_i. \tag{1.4}$$

The Hamiltonian H associated with a Lagrangian L is defined by the Legendre transformation

$$H(q,p,t) = \sum_i p_i \dot{q}_i - L(q,\dot{q},t). \tag{1.5}$$

Note that the Hamiltonian is to be expressed as a function of the variables q,p,t. That is, the variables $\dot{q}$ are to be eliminated in terms of the p's.

Exercise 1.2: For the Lagrangian (1.1), show that the canonical momenta in Cartesian coordinates are given by the equation

$$\vec{p} = m_o\vec{v}/(1-\vec{v}^2/c^2)^{1/2} + q\vec{A}. \tag{1.6}$$

Note that the first term in (1.6) is just the relativistic mechanical momentum. Consequently, the relation (1.6) can also be written in the form

$$\vec{p} - q\vec{A} = \vec{p}^{\,mech}. \tag{1.7}$$

Exercise 1.3: Show that the Hamiltonian associated with the Lagrangian (1.1) is given in Cartesian coordinates by the expression

$$H = [m_o^2 c^4 + c^2(\vec{p}-q\vec{A})^2]^{1/2} + q\psi. \tag{1.8}$$

Exercise 1.4: Find the canonical momenta and Hamiltonian associated with the Lagrangian (1.1) when cylindrical coordinates ρ, ϕ, z are used as generalized coordinates.
Answer:

$$p_\rho = m_o\dot{\rho}/(1-v^2/c^2)^{1/2} + qA_\rho \tag{1.9a}$$
$$p_z = m_o\dot{z}/(1-v^2/c^2)^{1/2} + qA_z \tag{1.9b}$$
$$p_\phi = m_o\rho^2\dot{\phi}/(1-v^2/c^2)^{1/2} + q\phi A_\phi. \tag{1.9c}$$

$$H = \{m_o^2 c^4 + c^2[(p_\rho-qA_\rho)^2 + (p_z-qA_z)^2 + (p/\rho-qA_\phi)^2]\}^{1/2} + q\psi. \tag{1.10}$$

Hamilton's equations of motion for the 2n canonical variables $(q_1...q_n)$ and $(p_1...p_n)$ are given in terms of the Hamiltonian $H(q,p,t)$ by the rules

$$\dot{q}_i = \partial H/\partial p_i \quad , \quad \dot{p}_i = -\partial H/\partial q_i. \tag{1.11a,b}$$

For later use, it is convenient to add one more equation to the set (1.11). Consider the total time rate of change of the Hamiltonian H along a trajectory in q, p space. Using the chain rule, one finds the result

$$dH/dt = \partial H/\partial t + \sum_i [(\partial H/\partial q_i)\dot{q}_i + (\partial H/\partial p_i)\dot{p}_i]. \tag{1.12}$$

However, the quantity under the summation sign vanishes because of (1.11). It follows that the Hamiltonian has the special property

$$dH/dt = \partial H/\partial t. \tag{1.13}$$

1.2 Hamilton's Equations with a Coordinate as an Independent Variable

Note that in the usual Hamiltonian formulation (as in the usual Lagrangian formulation) the time t plays the distinguished role of an independent variable, and all the q's and p's are dependent variables. That is, the canonical variables are viewed as functions $q(t)$, $p(t)$ of the independent variable t.

In some cases, it is more convenient to take some coordinate to be the independent variable rather than the time. For example, consider the passage of a collection of particles through a rectangular magnet such as is shown in Figs. (1.1) and (1.2). In such a situation, particles with different initial conditions will require different times to pass through the magnet. If the quantities of interest are primarily the locations and momenta of the particles as they leave the exit face of the magnet, then it would clearly be more convenient to use some coordinate which measures the progress of a particle through the magnet as an independent variable. In the case of a magnet with parallel faces as shown in Figs. (1.1) and (1.2), a convenient independent variable would be the x coordinate. In the case of a wedge magnet as shown in Fig. (1.3), a convenient independent variable would be the angle ϕ of a cylindical coordinate triad ρ,ϕ,z.

Suppose some coordinate is indeed chosen to be the independent variable. Is it then still possible to have a Hamiltonian (or Lagrangian) formulation of the equations of motion? The answer in general is yes as is shown by the following theorem.

Theorem 1.1: Suppose $H(q,p,t)$ is a Hamiltonian for a system having n degrees of freedom. Suppose further that $\dot{q}_1 = \partial H/\partial p_1 \neq 0$ for some interval of time T in some region R of the phase space described by the 2n variables $(q_1 \ldots q_n)$ and $(p_1 \ldots p_n)$. Then in this region and time interval, q_1 can be introduced as an independent variable in place of the time t. Moreover, the equations of motion with q_1 as an independent variable can be obtained from a Hamiltonian which will be called K.

Proof: Consider the 2n-2 quantities $(q_2 \ldots q_n)$, $(p_2 \ldots p_n)$. They obey Hamilton's equations of motion

$$\dot{q}_i = \partial H/\partial p_i \qquad i = 2,\ldots,n \tag{1.14a}$$

$$\dot{p}_i = -\partial H/\partial q_i \qquad i = 2,\ldots,n. \tag{1.14b}$$

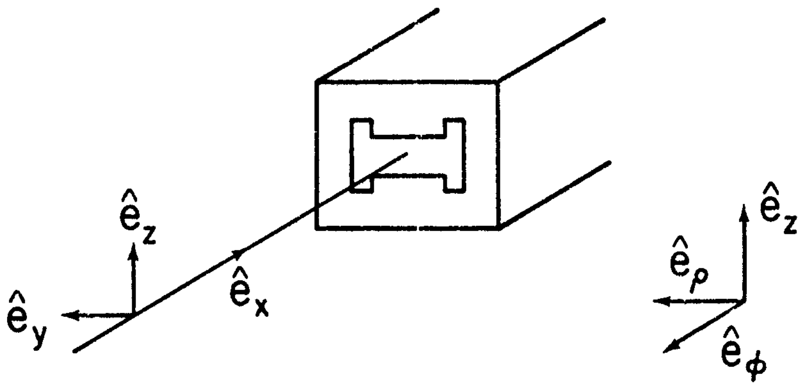


Fig. 1.1: Typical choice of a cartesion coordinate system for the description of charged particle trajectories in a magnet. Also shown, to the right, is an associated cylindrical coordinate system unit vector triad.

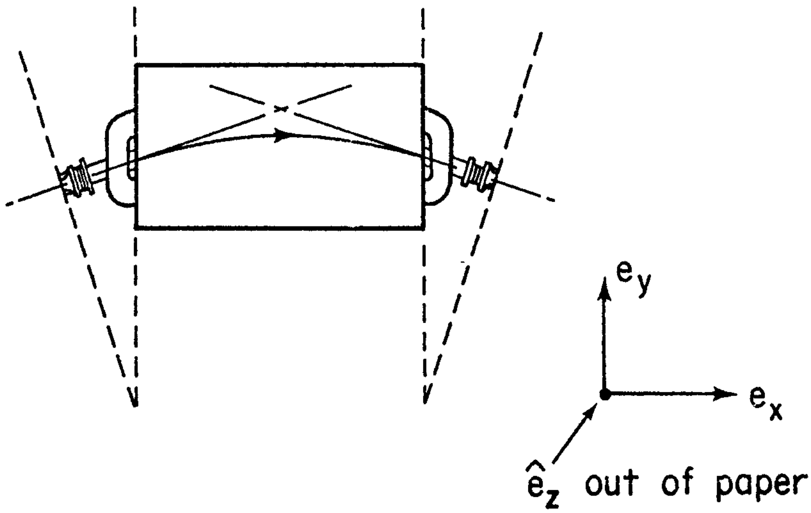


Fig. 1.2: Top view of a particle trajectory in a parallel-faced magnet.

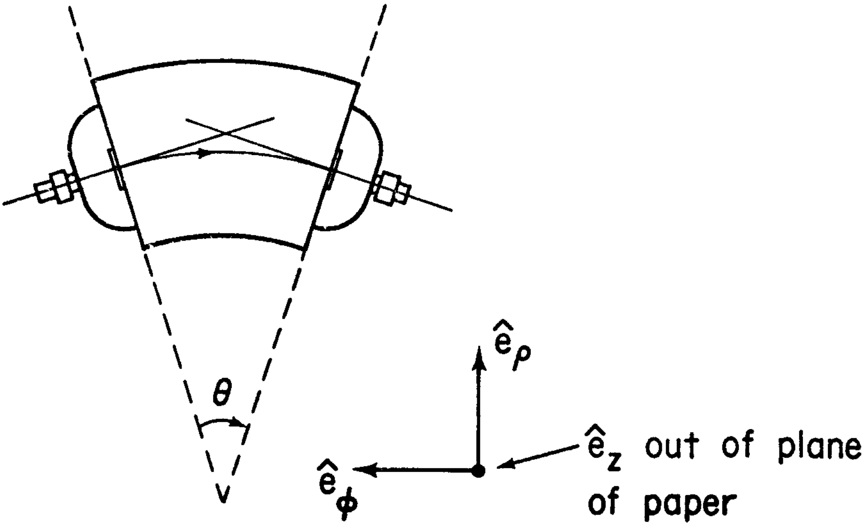


Fig. 1.3: Top view of a particle trajectory in a wedge magnet. The trajectory is conveniently described using cylindrical coordinates ρ, ϕ, z.

Denote total derivatives with respect to q_1 by a prime. Then, applying the chain rule to Eqs. (1.14), one finds the relations

$$\begin{aligned} q_i' &= dq_i/dq_1 = (dq_i/dt)(dt/dq_1) \\ &= (\partial H/\partial p_i)(\partial H/\partial p_1)^{-1} \end{aligned} \tag{1.15a}$$

$$\begin{aligned} p_i' &= dp_i/dq_1 = (dp_i/dt)(dt/dq_1) \\ &= -(\partial H/\partial q_i)(\partial H/\partial p_1)^{-1}. \end{aligned} \tag{1.15b}$$

To these 2n-2 relations it is convenient to add two more. First, suppose the time t is added to the list of coordinates as a dependent variable. Then one immediately has the relation

$$t' = dt/dq_1 = (dq_1/dt)^{-1} = (\partial H/\partial p_1)^{-1}. \tag{1.15c}$$

Second, suppose the quantity p_t defined by writing $p_t = -H$ is formally added to the list of momenta. Then, using (1.13), one finds the equation

$$\begin{aligned} p_t' &= dp_t/dq_1 = (dp_t/dt)(dt/dq_1) \\ &= -(\partial H/\partial t)(\partial H/\partial p_1)^{-1}. \end{aligned} \tag{1.15d}$$

Equations (1.15) are the desired equations of motion for the 2n variables $(t, q_2, \ldots, q_n)$, $(p_t, p_2, \ldots, p_n)$ with q_1 as an independent variable. What remains to be shown is that the quantities on the right-hand sides of Eqs. (1.15) can be obtained by applying the standard rules to some Hamiltonian K.

Look once again at the defining relation for p_t,

$$p_t = -H(q,p,t). \tag{1.16}$$

Suppose that this relation is solved for p_1 to give a relation of the form

$$p_1 = -K(t,q_2\ldots q_n;p_t,p_2\ldots p_n;q_1). \tag{1.17}$$

Such an inversion is possible according to the implicit function theorem because $\partial H/\partial p_1 \neq 0$ by assumption. Then, as the notation is intended to suggest, K is the desired new Hamiltonian.

To see that this is so, take the total differential of (1.16) to find the result

$$dp_t = -(\partial H/\partial t)dt - \sum_i (\partial H/\partial q_i)dq_i - \sum_i (\partial H/\partial p_i)dp_i. \quad (1.18)$$

Now solve (1.18) for dp_1 to get the relation

$$dp_1 = \left(\frac{\partial H}{\partial p_1}\right)^{-1} [-dp_t - (\partial H/\partial t)dt - \sum_i (\partial H/\partial q_i)dq_i$$

$$- \sum_{i\neq 1} (\partial H/\partial p_i)dp_i]. \quad (1.19)$$

Also, take the total differential of (1.17) to find the result

$$dp_1 = -(\partial K/\partial p_t)dp_t - (\partial K/\partial t)dt$$

$$- \sum_i(\partial K/\partial q_i)dq_i - \sum_{i\neq 1} (\partial K/\partial p_i)dp_i. \quad (1.20)$$

Upon comparing (1.19) and (1.20), and looking at Eqs. (1.15), one obtains the advertised result

$$\begin{aligned}
\partial K/\partial p_t &= (\partial H/\partial p_1)^{-1} = t' \\
\partial K/\partial p_i &= (\partial H/\partial p_i)(\partial H/\partial p_1)^{-1} = q_i' \quad i = 2,\ldots.n \\
\partial K/\partial t &= (\partial H/\partial t)(\partial H/\partial p_1)^{-1} = -p_t' \qquad (1.21) \\
\partial K/\partial q_i &= (\partial H/\partial q_i)(\partial H/\partial p_1)^{-1} = -p_i' \quad i = 2,\ldots.n.
\end{aligned}$$

That is, the indicated partial derivatives of K do indeed produce the required right-hand sides of Eqs. (1.15). Note that according to Eqs. (1.21), the quantity p_t may be viewed as the momentum canonically conjugate to the time t.

Exercise 1.5: Find the Hamiltonian K corresponding to the Hamiltonian H given by (1.8) when the x coordinate is taken to be the independent variable. Assume that $\dot{x} > 0$ for the trajectories in question.

Answer:

$$K = -[(p_t + q\psi)^2/c^2 - m_o^2c^2 - (p_y - qA_y)^2 - (p_z - qA_z)^2]^{1/2} - qA_x. \qquad (1.22)$$

Note that according to (1.16), p_t is usually negative.

Exercise 1.6: Find the Hamiltonian K corresponding to the Hamiltonian H given by (1.10) when the coordinate ϕ is taken to be the independent variable. Assume that $\dot{\phi} < 0$ for trajectories of interest. [This would be the case if the triad of unit vectors $\hat{e}$,$\hat{e}_\phi$, $\hat{e}_z$ is taken to form a right-handed coordinate system, if $\hat{e}_\rho$ points along the general direction of $\hat{e}_y$ of a rectangular triad $\hat{e}_x$, $\hat{e}_y$, $\hat{e}_z$, and if the two $\hat{e}_z$ vectors of the two triads agree. See Figs. (1.1) and (1.3)].

Answer:

$$K = +\rho[(p_t + q\psi)^2/c^2 - m_o^2c^2 -(p_\rho - qA_\rho)^2 - (p_z - qA_z)^2]^{1/2} - q\rho A_\phi. \qquad (1.23)$$

Exercise 1.7: Show that a uniform electric field in the x direction can be derived from the scalar and vector potentials

$$\psi = 0$$

$$\vec{A} = -Et\hat{e}_x.$$

Exercise 1.8: Show that a uniform electric field in the x direction can be derived from the scalar and vector potentials

$$\psi = -Ex$$

$$\vec{A} = 0.$$

Exercise 1.9: Show that a uniform magnetic field in the z direction can be derived from the scalar and vector potentials

$$\psi = 0$$

$$\vec{A} = -By\hat{e}_x.$$

Exercise 1.10: Show that a magnetic quadrupole field with midplane symmetry can be derived from the scalar and vector potentials

$$\psi = 0$$

$$\vec{A} = (a_2/2)(z^2 - y^2)\hat{e}_x.$$

Exercise 1.11: Show that a magnetic sextupole field with midplane symmetry can be derived from the scalar and vector potentials

$$\psi = 0$$

$$\vec{A} = a_s(yz^2 - y^3/3)\ \hat{e}_x.$$

Exercise 1.12: Show that when cylindrical coordinates are used, a uniform magnetic field in the z direction can be derived from the scalar and vector potentials

$$\psi = 0$$

$$\vec{A} = (\rho/2)B\ \hat{e}_\phi.$$

Exercise 1.13: Suppose that the electric field $\vec{E}$ is zero and the magnetic field $\vec{B}$ is static. Show that in this case p_t has the constant value

$$p_t = -[m_o^2c^4 + c^2(\vec{p}^{mech})^2]^{1/2}.$$

Suppose further that the magnetic field can be derived from a vector potential having only an x component. Show that if one is only interested in determining trajectories and not in determining transit time, then one may use the Hamiltonian

$$K = -[(\vec{p}^{mech})^2 - p_y^2 - p_z^2]^{1/2} - qA_x$$

with x treated as the independent variable.

1.3 Hamiltonian formulation of light optics

Much of the remaining lectures will be devoted, at least indirectly, to the subject of charged particle beam optics. For this reason, it is useful to also consider the analogous topic of geometrical light optics. It too can be formulated in Hamiltonian terms.

Consider the optical system illustrated schematically in Fig. (1.4). A light ray originates at the general initial point P^i with spatial coordinates $\vec{r}^i$, and moves in an initial direction specified by the unit vector $\hat{s}^i$. After passing through an optical device, it arrives at the final point P^f with coordinates $\vec{r}^f$ in a direction

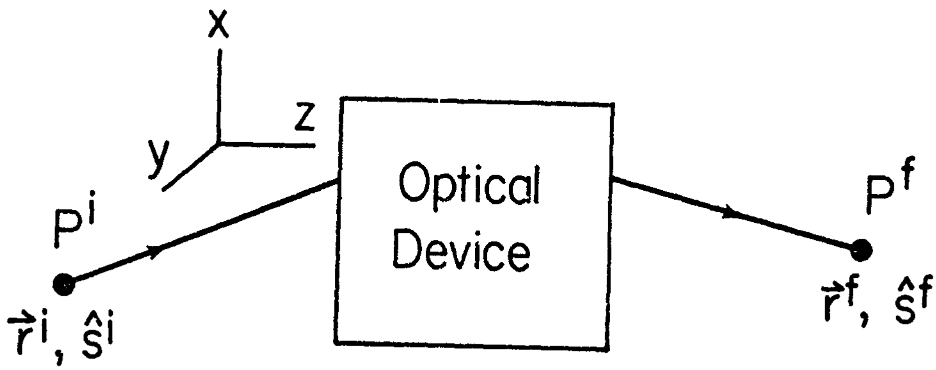


Fig. 1.4: An optical system consisting of an optical device preceded and followed by simple transit. A ray originates at P^i with location $\vec{r}^i$ and direction $\hat{s}^i$, and terminates at P^f with location $\vec{r}^f$ and direction $\hat{s}^f$.

specified by the unit vector $\hat{s}^f$. Given the initial quantities $(\vec{r}^i,\hat{s}^i)$, the fundamental problem of geometrical optics is to determine the final quantities $(\vec{r}^f,\hat{s}^f)$, and to design an optical device in such a way that the relation between the initial and final ray quantities has various desired properties.

Suppose the z coordinates of the initial and final points P^i and P^f as shown in Fig. (1.4) are held fixed. The planes $z=z^i$ and $z=z^f$ may be viewed as object and image planes respectively. Further, suppose the general light ray from P^i to P^f is parameterized using z as an independent variable. That is, the path of a general ray is described by specifying the two functions x(z) and y(z). Then the element ds of path length along a ray is given by the expression

$$\begin{aligned} ds &= [(dz)^2 + (dx)^2 + (dy)^2]^{1/2} \\ &= [1 + (x')^2 + (y')^2]^{1/2}\, dz. \end{aligned} \tag{1.24}$$

Here a prime denotes the differentiation d/dz. Consequently, the optical path length along a ray from P^i to P^f is given by the integral

$$A = \int_{z^i}^{z^f} n(z,y,z)\, [1 + (x')^2 + (y')^2]^{1/2}\, dz. \tag{1.25}$$

Here the function $n(x,y,z)=n(\vec{r})$ specifies the index of refraction at each point before and after the optical device and in the device itself.

Fermat's principle requires that A be an extremum for the path of an actual ray. Therefore, the ray path satisfies the Euler-Lagrange equations

$$\begin{aligned} &d/dz(\partial L/\partial x') - \partial L/\partial x = 0 \\ &d/dz(\partial L/\partial y') - \partial L/\partial y = 0 \end{aligned} \tag{1.26}$$

with a Lagrangian L given by the expressions

$$L = n(x,y,z)\, [1 + (x')^2 + (y')^2]^{1/2}. \tag{1.27}$$

Exercise 1.14: Calculate explicit expressions for the two "momenta" conjugate to the coordinates x and y defined by the standard relations

$$p_x = \partial L/\partial x' \quad , \; p_y = \partial L/\partial y' \tag{1.28}$$

Find the Hamiltonian H corresponding to L.
Answer:

$$p_x = n(\vec{r})\, x'/[1 + (x')^2 + (y')^2]^{1/2} \tag{1.29}$$

$$p_y = n(\vec{r})\, y'/[1 + (x')^2 + (y')^2]^{1/2},$$

$$H = -[n^2(\vec{r}) - p_x^2 - p_y^2]^{1/2}. \tag{1.30}$$

2. SYMPLECTIC MATRICES

2.1 Definitions

The purpose of this section is to define symplectic matrices and explore some of their properties in preparation for future use.

To define symplectic matrices, it is first necessary to introduce a certain fundamental $2n \times 2n$ matrix J. It is defined by the equation

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} . \tag{2.1}$$

Here each entry in J is an $n \times n$ matrix, I denotes the $n \times n$ identity matrix, and all other entries are zero.

Exercise 2.1: Show that the matrix J has the following properties:

$$J^2 = -I \text{ or } J^{-1} = -J \tag{2.2}$$

$$\det(J) = 1 \tag{2.3}$$

$$\tilde{J} = -J \tag{2.4}$$

$$\tilde{J}\, J = I. \tag{2.5}$$

Here $\tilde{J}$ denotes the transpose of J.

With this background a $2n \times 2n$ matrix M is said to be symplectic if

$$\tilde{M}\, J\, M = J. \tag{2.6}$$

Observe that symplectic matrices must be of even dimension by definition.

Exercise 2.2: Show that a symplectic matrix M has the following properties:

$$\det(M) = \pm 1 \tag{2.7}$$

$$M^{-1} = -J\,\tilde{M}\,J \text{ or } M^{-1} = J^{-1}\,\tilde{M}\,J \tag{2.8}$$

$$M\,J\,\tilde{M} = J. \tag{2.9}$$

Comment: It can be shown that det(M) actually always equals +1 for a symplectic matrix.[4] Also, as is easily checked, in the 2×2 case the necessary and sufficient condition for a matrix to be symplectic is that it have determinant +1.

Exercise 2.3: Show that the matrices I and J are symplectic.

Exercise 2.4: Supose M is a symplectic matrix. Show that M^{-1} is then also a symplectic matrix.

Exercise 2.5: Suppose M and N are symplectic matrices. Show that the produce MN is then also a symplectic matrix.

2.2 Group properties

A set of matrices forms a group G if it satisfies the following properties:

1. The identity matrix I is in G.
2. If M is in G, M^{-1} exists and is also in G.
3. If M and N are in G, so is the produce MN.

Evidently, according to exercise (2.3), Eq. (2.7), and exercises (2.4) and (2.5), the set of all $2n \times 2n$ symplectic matrices (for any particular value of n) form a group. This group is often denoted by the symbol Sp(2n).

2.3 Eigenvalue spectrum

The characteristic polynomial $P(\lambda)$ of any matrix M is defined by the equation

$$P(\lambda) = \det(M - \lambda I). \tag{2.10}$$

Evidently $P(\lambda)$ is a polynomial with real coefficients if the matrix M is real. Also, the eigenvalues of M are the roots of the equation

$$P(\lambda) = 0. \tag{2.11}$$

It follows that if M is a real matrix, then its eigenvalues must also be real or must occur in complex conjugate pairs $\lambda, \bar{\lambda}$.

Exercise 2.6: Show that a symplectic matrix cannot have $\lambda=0$ as an eigenvalue.

Suppose M is a symplectic matrix. Then it follows from (2.8) that

$$J^{-1}(\tilde{M}-\lambda I)J = M^{-1} - \lambda I = -\lambda M^{-1}(M - \lambda^{-1}I). \quad (2.12)$$

Now take the determinant of both sides of (2.12). The result is the relation

$$P(\lambda) = \lambda^{2n} P(1/\lambda). \quad (2.13)$$

It follows that if λ is an eigenvalue of a symplectic matrix, so is the reciprocal $1/\lambda$. Consequently, the eigenvalues of a symplectic matrix must form reciprocal pairs.

Exercise 2.7: Verify (2.13) starting with (2.12).

The symmetry between λ and $1/\lambda$ exhibited by (2.13) can be further displayed by rewriting the equation in the form

$$\lambda^{-n} P(\lambda) = \lambda^{n} P(1/\lambda). \quad (2.14)$$

Now define another function $Q(\lambda)$ by writing

$$Q(\lambda) = \lambda^{-n} P(\lambda). \quad (2.15)$$

The functions P and Q evidently have the same zeroes. Moreover, the condition (2.14) requires that Q have the symmetry property

$$Q(\lambda) = Q(1/\lambda). \quad (2.16)$$

Equation (2.16) shows not only that the eigenvalues of a symplectic matrix must occur in reciprocal pairs; it shows that they must also occur with the same multiplicity. That is, if the root λ_o has multiplicity k, so must the root $1/\lambda_o$.

Also, if either +1 or -1 is a root, then this root must have even multiplicity. To see this, suppose for example that $\lambda=1$ is a root. Introduce the variable μ by writing $\lambda=\exp\mu$. Then (2.16) shows that Q is an even function of the variable μ and hence must have an expansion of the form

$$Q = \sum_{m=0}^{\infty} c_m \mu^{2m}. \quad (2.17)$$

Moreover, when λ is near 1, λ and μ are related by the expansion

$$\mu = \log\lambda = \log[1 + (\lambda-1)] = (\lambda-1)[1 - (\lambda-1)/2 + \cdots]. \quad (2.18)$$

Comparison of (2.17) and (2.18) shows that $\lambda=1$ is not a root unless $c_o=0$. If $c_o=0$, then $\lambda=1$ is a root of multiplicity 2. If $c_1=0$ as well, then $\lambda=1$ is a root of multiplicity 4, etc. A similar argument holds near $\lambda=-1$ upon making the substitution $\lambda=-\exp\mu$.

In summary, it has been shown that the eigenvalues of a real symplectic matrix must satisfy the following properties:

1. They must be real or occur in complex conjugate pairs.
2. They must occur in reciprocal pairs, and each member of the pair must have the same multiplicity.
3. If either ± 1 is an eigenvalue, it must have even multiplicity.

When combined, the conditions just enumerated place strong restrictions on the possible eigenvalues of a real symplectic matrix. Consider first the simplest case of a 2×2 symplectic matrix ($n=1$). Call the eigenvalues λ_1 and λ_2. Then, by the reciprocal property, it follows that

$$\lambda_1 \lambda_2 = 1. \tag{2.19}$$

Suppose, now, that λ_1 is real, positive, and greater than 1. Then λ_2 is real, positive, and less than 1. Similarly, if λ_1 is real, negative, and less than -1, than λ_2 is real, negative, and greater than -1. On the other hand, if λ_1 is complex, then $\lambda_2 = \bar{\lambda}_1$. This condition, when combined with Eq. (2.19), shows that in this case λ_1 and λ_2 must lie on the unit circle in the complex plane. Finally, there are the two special cases $\lambda_1=\lambda_2=1$ and $\lambda_1= \lambda_2=-1$.

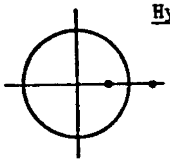
Altogether, there are five possible cases. They are listed below along with names and designations whose significance will become clear later on. See also Fig. (2.1).

1. Hyperbolic case (unstable): $\lambda_1 > 1$ and $0 < \lambda_2 < 1$.
2. Inversion hyperbolic case (unstable): $\lambda_1 < -1$ and $-1 < \lambda_2 < 0$.
3. Elliptic case (stable): $\lambda_1 = e^{i\phi}$, $\lambda_2 = e^{-i\phi}$.

   (Eigenvalues are complex conjugates and lie on unit circle).
4. Parabolic case (generally linearly unstable): $\lambda_1 = \lambda_2 = +1$.
5. Inversion parabolic case (generally linearly unstable): $\lambda_1 = \lambda_2 = -1$.

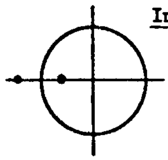
The next simplest case is that of a 4×4 symplectic matrix ($n=2$). In this case, one has to deal with four possible eigenvalues and then apply reasoning analogous to the 2×2 case. Figures (2.2)

Case 1.



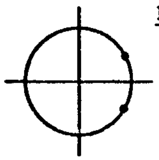
Hyperbolic (Unstable)

Case 2.



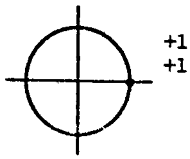
Inversion Hyperbolic (Unstable)

Case 3.



Elliptic (Stable)

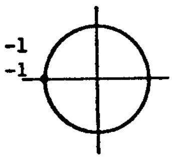
Case 4.



Parabolic

Transition between elliptic and hyperbolic cases can only occur by passage through this degenerate case. (Generally linearly unstable.)

Case 5.



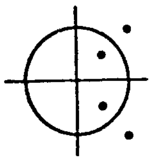
Inversion Parabolic

Transition between elliptic and inversion hyperbolic cases can only occur by passage through this degenerate case. (Generally linearly unstable.)

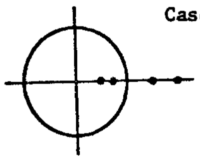
Fig. 2.1: Possible cases for the eigenvalues of a 2 × 2 real symplectic matrix.

Fig. 2.2: Possible eigenvalue configurations for a 4 × 4 symplectic matrix. The mirror image of each configuration is also a possible configuration, and therefore is not shown in order to save space.

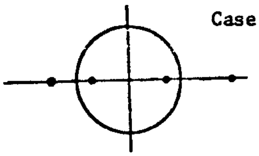
A. Generic Configurations



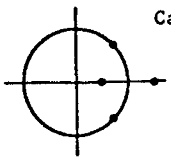
Case 1. All eigenvalues complex and off the unit circle. All eigenvalues can be obtained from a single one by the operations of complex conjugation and taking reciprocals. Unstable.



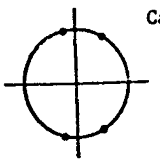
Case 2. All eigenvalues real, off the unit circle, and of same sign. Eigenvalues form reciprocal pairs. Unstable.



Case 3. All eigenvalues real, off the unit circle, and of differing signs. Eigenvalues form reciprocal pairs. Unstable.

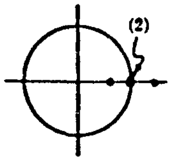


Case 4. Two eigenvalues complex and confined to unit circle. Two eigenvalues real. Eigenvalues form reciprocal pairs. Complex eigenvalues are also complex conjugate. Unstable.

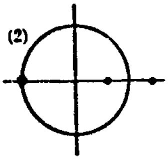


Case 5. All eigenvalues complex and confined to unit circle. Eigenvalues form reciprocal pairs which are also complex conjugate. Stable.

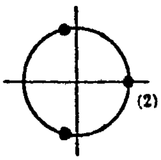
B. <u>Degenerate Configurations</u>. Transitions between generic configurations can only occur by passage through a degenerate configuration. Mirror image configurations are again possible, but not shown.



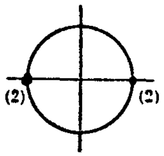
Case 1. Two eigenvalues equal, and two eigenvalues real. All of same sign. Occurs in transition between generic cases 2 and 4. <u>Unstable</u>.



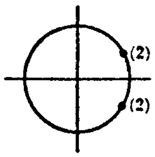
Case 2. Two eigenvalues equal, and two eigenvalues real. Signs differ. Occurs in transition between generic cases 3 and 4. <u>Unstable</u>.



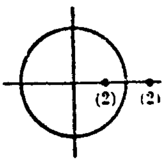
Case 3. Two eigenvalues equal, and two eigenvalues confined to unit circle. Occurs in transition between generic cases 4 and 5. <u>Generally linearly unstable</u>.



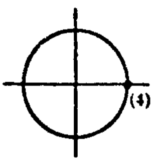
Case 4. Two eigenvalues equal +1 and two equal -1. Occurs in transition between generic cases 3 and 5, or 3 and 4, or 4 and 5. Also occurs in transition between degenerate cases 2 and 3. <u>Generally linearly unstable</u>.



Case 5. Two pairs of eigenvalues equal, and confined to unit circle. Occurs in transition between generic cases 1 and 5. Not, however, a sufficient condition to guarantee that such a transition is possible. Stability also undetermined in absence of further conditions.



Case 6. Two pairs of eigenvalues equal and real. Occurs in transition between generic cases 1 and 2. <u>Unstable</u>.



Case 7. All eigenvalues equal and have value ±1. Occurs in transitions between generic cases 1, 2, 4, and 5 and degenerate cases 1, 3, 5, and 6. <u>Generally linearly unstable</u>.

illustrate the various possibilities that can occur. Analysis of the possible spectrum of the 2n eigenvalues for the general $2n \times 2n$ symplectic matrix proceeds in a similar fashion.

Exercise 2.8: Show that the eigenvalues of a real symplectic matrix cannot all have absolute value less than 1.

Exercise 2.9: Show that the eigenvalues of a real symplectic matrix cannot all have absolute value greater than 1.

Exercise 2.10: Suppose that all the eigenvalues of a real symplectic matrix M lie on the unit circle and are distinct. Now suppose that M is changed slightly in such a way that it remains symplectic. Show that if the change in M is finite but small enough, then the eigenvalues must remain on the unit circle and must still be distinct. Show that all the generic configurations of Fig. (2.2) are unchanged by small perturbations. That is why they are called generic.

Exercise 2.11: In the case of a 4×4 symplectic matrix, show that $Q(\lambda)$ can be written in the form

$$Q(\lambda) = (\lambda+1/\lambda)^2 + 4b(\lambda+1/\lambda) + 4c, \tag{2.20}$$

where the quantities b and c are given by the relations

$$b = [P(1) - P(-1)]/[16] \tag{2.21}$$

$$c = [P(1) + P(-1)]/8 - 1. \tag{2.22}$$

Exercise 2.12: One might think that the determination of the four eigenvalues of a 4×4 symplectic matrix would in general require the solution of a quartic equation. However, using the results of exercise 2.11, show that thanks to the symplectic condition the eigenvalues can be found from the simple algebraic formulas

$$\lambda = w \pm (w^2-1)^{1/2} \tag{2.23}$$

where

$$w = -b \pm (b^2-c)^{1/2}. \tag{2.24}$$

Note that in general there are four choices of signs to be made corresponding to the four possible eigenvalues.

2.4 Lie algebraic properties

Let A be any matrix. The exponential of a matrix, written variously as e^A or exp(A), is defined by the exponential series

$$\exp(A) = \sum_{n=0}^{\infty} A^n/n! \tag{2.25}$$

Exercise 2.13: Show that the series (2.25) converges for any matrix A.

Similarly, the logarithm of a matrix A is defined by the series

$$\log(A) = \log[I - (I-A)] = -\sum_{m=1}^{\infty} (I-A)^m/m. \tag{2.26}$$

Exercise 2.14: Show that the series (2.26) converges for A sufficiently near the identity matrix I. Note that when A is near the identity, then log(A) is near the zero matrix.

As might be expected, the exponential and logarithmic functions are related. Specifically, if one has

$$B = \log(A), \tag{2.27a}$$

then it follows that

$$A = \exp(B). \tag{2.27b}$$

Exercise 2.15: Verify the relation given by Eqs. (2.27) in the case that A is sufficiently near the identity using the definitions (2.25) and (2.26).

With this background, suppose that M is a real symplectic matrix near the identity. Then M can be written in the form

$$M = \exp(\varepsilon B) = I + \varepsilon B + \varepsilon^2 B^2/2! + \cdots \tag{2.28}$$

where ε is a small number and B is a real matrix. Next insert the expansion (2.28) into the symplectic condition (2.6). Then one finds the relation

$$(I + \varepsilon \tilde{B})\ J\ (I + \varepsilon B) = J + O(\varepsilon^2). \tag{2.29}$$

Upon assuming that (2.29) holds term by term in powers of ε, it follows that B must satisfy the condition

$$\tilde{B}J + JB = 0. \tag{2.30}$$

Exercise 2.16: Derive (2.30) directly from (2.28) using (2.8).

To understand the implications of the condition (2.30), suppose that B is written in the form

$$B = JS. \tag{2.31}$$

Exercise 2.17: Verify that it is always possible to find a real matrix S such that (2.31) is true.

Upon inserting (2.31) into (2.30), one finds the equivalent condition

$$\tilde{S}JJ + JJS = 0 \quad \text{or} \quad \tilde{S} = S. \tag{2.32}$$

That is, S must be a symmetric matrix.

Conversly, suppose that B is any matrix of the form (2.31) with S real and symmetric. Then the matrix M defined by (2.28) is symplectic. To see this, simply compute! One finds the results

$$\begin{aligned}
M &= \exp(\varepsilon JS), \\
\tilde{M} &= \exp(-\varepsilon SJ), \\
\tilde{M}JM &= \exp(-\varepsilon SJ)J \exp(\varepsilon JS) \\
&= JJ^{-1} \exp(-\varepsilon SJ)J \exp(JS) \\
&= J \exp(-\varepsilon J^{-1}SJ^2) \exp(JS) \\
&= J \exp(-\varepsilon JS) \exp(JS) \\
&= J.
\end{aligned}$$

Exercise 2.18: Verify the details of the calculation above using the series definition of the exponential function as given by (2.23).

What has been shown is that any symplectic matrix sufficiently near the identity can be written in the form exp(JS) with S small and symmetric, and vice versa.

A set of matrices forms a Lie algebra if it satisfies the following properties:

1. If the matrix A is in the Lie algebra, then so is the matrix aA where a is any scalar.
2. If two matrices A and B are in the Lie algebra, then so is their sum.
3. If two matrices A and B are in the Lie algebra, then so is their commutator [A,B]. The commutator is defined by the relation

$$[A,B] = AB - BA. \tag{2.33}$$

The reason for introducing the concept of a Lie algebra at this stage is to point out that the set of matrices of the form JS with S symmetric is a Lie algebra.

Exercise 2.19: Verify that the set of matrices of the form JS is indeed a Lie algebra by showing that conditions 1 through 3 above are satisfied.

It is a remarkable fact that there is a close connection between the concept of a Lie algebra and that of a group. The connection arises from a deep property of the exponential function which generally bears the names Campbell-Baker-Hausdorff. Their result may be stated as follows: Let A and B be any two matrices (of the same dimension). Form the matrices exp(sA) and exp(tB) where s and t are parameters. Then, for s and t sufficiently small, it is possible to write

$$\exp(sA)\,\exp(tB) = \exp(C), \qquad (2.34)$$

where C is some other matrix. The remarkable fact is that C is a member of the Lie algebra generated by A and B. That is, C is a sum of elements formed only from A and B and their multiple commutators. Specifically, one has the relation

$$C(s,t) = sA + tB + (st/2)[A,B] + (s^2t/12)[A,[A,B]]$$
$$+ (st^2/12)[B,[B,A]] + O(s^3t, s^2t^2, st^3). \qquad (2.35)$$

No terms of the form A^2, B^2, AB, A^2B, etc. occur! In general, the series for C contains an infinite number of terms and may only converge for sufficiently small s and t.

The proof of this theorem is difficult and is given elsewhere.[5] For present purposes, it shows that given any Lie algebra L of matrices, there exists a corresponding Lie group G. Furthermore, the rules for multiplying any two group elements are contained within the Lie algebra. To see the truth of this assertion, consider all matrices of the form $g(s) = \exp(s\ell)$ with ℓ contained in L. According to the previous result, one has

$$\exp(s\ell)\,\exp(t\ell') = \exp \ell''$$

for s, t sufficiently small. Also

$$g(0) = I \text{ and } g^{-1}(s) = g(-s).$$

Thus these matrices, at least those sufficiently near the identity, form a group. Once the group has been obtained near the identity, it can be extended to a global group by successively multiplying the different g's already obtained.

It has already been shown that symplectic matrices form a group. Furthermore, it has been shown that symplectic matrices near the identity can be written as the exponentials of elements of a Lie algebra. It follows that symplectic matrices form a Lie group.

Properties 1 and 2 of a Lie algebra indicate that the elements of a Lie algebra form a linear vector space. It is therefore natural to speak of the dimension of a Lie algebra and its associated Lie group. For the case of the symplectic group, elements of the Lie algebra are of the form (2.31) where S is any symmetric matrix. The dimension of the Lie algebra in this case, therefore, is just the dimensionality of the set of all $2n \times 2n$ symmetric matrices. This number is easily computed. There are 2n independent entries on the diagonal of a $2n \times 2n$ symmetric matrix, and $[(2n)^2 - 2n]/2$ independent entries above the diagonal. Finally, all the entries below the diagonal are given in terms of the entries above the diagonal by the symmetry condition. Therefore, the dimension of the symplectic group Lie algebra, which will be written as dim Sp(2n), is given by the relation

$$\dim Sp(2n) = 2n + [(2n)^2 - 2n]/2 = n(2n + 1). \tag{2.36}$$

2.5 Exponential representation

The discussion so far has shown that symplectic matrices sufficiently near the identity element can be written as exponentials of elements in the symplectic group Lie algebra. The next question to ask is what can be said about representing symplectic matrices in general.

To study this question, it is useful to employ polar decomposition. Let M be any real nonsingular matrix. Then M can be written uniquely in form

$$M = P\mathcal{O}. \tag{2.37}$$

Here P is a real positive definite symmetric matrix, and $\mathcal{O}$ is a real orthogonal matrix.[6] (A matrix $\mathcal{O}$ is orthogonal if $\mathcal{O}\tilde{\mathcal{O}} = \tilde{\mathcal{O}}\mathcal{O} = I$.) Now suppose that M is symplectic. Using (2.8), the symplectic condition can be written in the form

$$M = J^{-1} \tilde{M}^{-1} J. \tag{2.38}$$

Then, upon inserting the polar decomposition (2.37) into (2.38), one finds the relation

$$P\mathcal{O} = J^{-1} P^{-1} J J^{-1} \mathcal{O} J. \tag{2.39}$$

Exercise 2.20: Verify Eq. (2.39).

Next, observe that the matrix $J^{-1} P^{-1} J$ is real, symmetric, and positive definite, and observe that the matrix $J^{-1} \mathcal{O} J$ is real and orthogonal.

<u>Exercise 2.21</u>: Verify the two observations above.

Consequently, because polar decomposition is unique, Eq. (2.39) implies the relations

$$P = J^{-1} P^{-1} J \tag{2.40a}$$

$$O = J^{-1} O J. \tag{2.40b}$$

Using the fact that P is symmetric and O is orthogonal, Eqs. (2.40) can also be written in the form

$$P = J^{-1} \tilde{P}^{-1} J \tag{2.41a}$$

$$O = J^{-1} \tilde{O}^{-1} J. \tag{2.41b}$$

It follows that each of the matrices P and O are themselves symplectic.

<u>Exercise 2.22</u>: Verify Eqs. (2.41), and the claim that O and P are symplectic.

The next thing to do is to work with the matrices O and P. Consider first the matrix O. Since O is real orthogonal and has determinant +1 (O is symplectic), it can be written in the form

$$O = \exp(A), \tag{2.42}$$

where A is a unique real antisymmetric matrix,

$$\tilde{A} = -A. \tag{2.43}$$

Upon inserting the representation (2.42) into the condition (2.40b), one finds the condition

$$O = \exp(A) = \exp(J^{-1}AJ). \tag{2.44}$$

<u>Exercise 2.23</u>: Verify Eq. (2.44) using (2.40b) and the definition (2.25).

Since the matrix $(J^{-1}AJ)$ is real antisymmetric if the matrix A is, and since A is unique, it follows from (2.44) that A has the property

$$J^{-1}AJ = A \text{ or } AJ = JA. \tag{2.45}$$

Using (2.43), the condition (2.45) can also be written in the form

$$\tilde{A}J + JA = 0. \tag{2.46}$$

Now compare (2.46) with Eq. (2.30). According to the argument applied earlier, the matrix A can be written in the form

$$A = JS^c, \tag{2.47}$$

where S^c is a real symmetric matrix. Further, since A commutes with J, see Eq. (2.45), it follows that S^c commutes with J.

$$S^c J = JS^c. \tag{2.48}$$

In summary, it has been shown that $\mathcal{O}$ can be written in the form

$$\mathcal{O} = \exp(JS^c), \tag{2.49}$$

where S^c is a real symmetric matrix which commutes with J.

Exercise 2.24: Verify that $(J^{-1}AJ)$ is real antisymmetric if A is.

It remains to see what can be said about the matrix P. Since P is real, symmetric, and positive definite, it can be written in the form

$$P = \exp(B), \tag{2.50}$$

where B is unique, real, and symmetric,

$$\tilde{B} = B. \tag{2.51}$$

Now insert the representation (2.50) into the condition (2.40a) to obtain the result

$$P = \exp(B) = \exp(-J^{-1}BJ). \tag{2.52}$$

Since the matrix $(J^{-1}BJ)$ is real symmetric if the matrix B is, and since B is unique, it follows from (2.52) that B has the property

$$J^{-1}BJ = -B \quad \text{or} \quad BJ + JB = 0. \tag{2.53}$$

Using (2.51), the condition (2.53) can be re-expressed in the form

$$\tilde{B}J + JB = 0. \tag{2.54}$$

Consequently, B can also be written in the form

$$B = JS^a, \tag{2.55}$$

where S^a is a real symmetric matrix. However, in this case Eq. (2.51) implies the condition

$$JS^a + S^aJ = 0. \tag{2.56}$$

That is, S^a anticommutes with J. In summary, it has been shown that P can be written in the form

$$P = \exp(JS^a), \tag{2.57}$$

where S^a is a real symmetric matrix which anticommutes with J.

Now combine Eqs. (2.37), (2.49), and (2.57). The result is that any symplectic matrix can be written in the form

$$M = \exp(JS^a)\,\exp(JS^c). \tag{2.58}$$

It has been shown that the most general symplectic matrix can be written as the product of two exponentials of elements in the symplectic group Lie algebra, and each of the elements is of a special type.

It is interesting to examine the properties of commuting and anticommuting with J in a bit more detail. Let S be any symmetric matrix. Form the matrices S^a and S^c by the rules

$$S^a = (S - J^{-1}SJ)/2 \tag{2.59a}$$

$$S^c = (S + J^{-1}SJ)/2. \tag{2.59b}$$

It is easily verified that S^a and S^c are symmetric, and anticommute and commute respectively with J as the notation suggests.

Exercise 2.25: Verify the claims just made.

Also, it is obvious by construction that

$$S = S^a + S^c. \tag{2.60}$$

That is, any symmetric matrix can be uniquely decomposed into a sum of two symmetric matrices which anticommute and commute with J respectively.

2.6 Basis for Lie algebra

The last topic to be discussed is that of a suitable basis for the symplectic group Lie algebra. For simplicity, the discussion will be limited to the cases Sp(2) and Sp(4). These two cases, along with Sp(6), are those of primary interest for accelerator applications.

In selecting a basis for the Lie algebra of the symplectic group, it is convenient to begin with the observation that matrices of the form JS^c constitute a Lie algebra all by themselves. That is, the commutator of any two matrices of the form JS^c is again a matrix of the form JS^c. By contrast, the commutator of a matrix of the form JS^c with that of the form JS^a is again a matrix of the form JS^a. Finally, the commutator of two matrices of the form JS^a is a matrix of the form JS^c. It is therefore convenient to begin with the matrices of the form JS^c.

Exercise 2.26: Check the assertions just made about various commutators.

In the 2×2 case of Sp(2), the most general symmetric matrix S is of the form

$$S = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}, \tag{2.61}$$

and J is simply the matrix

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.62}$$

Requiring that J commute with S gives the restrictions

$$\beta = 0, \quad \alpha = \gamma. \tag{2.63}$$

Consequently, the most general S^c in the 2×2 case is just a multiple of the identity,

$$S^c = \alpha I, \tag{2.64}$$

and JS^c is simply a multiple of J,

$$JS^c = \alpha J. \tag{2.65}$$

It is therefore convenient to select J itself as one of the basis elements of the Lie algebra.

Exercise 2.27: Verify that the requirement that J commute with S does indeed give the restrictions (2.63).

Next study the matrix S^a. Requiring that J anticommute with the S of (2.61) gives only the restriction

$$\gamma = -\alpha. \tag{2.66}$$

Exercise 2.28: Verify this fact.

Consequently, S^a is of the general form

$$S^a = \alpha \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \beta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad (2.67)$$

JS^a is of the general form

$$JS^a = \alpha \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} + \beta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (2.68)$$

Upon combining the results of Eqs. (2.65) and (2.68), one sees that a convenient choice of basis elements for the Lie algebra of Sp(2) is given by the matrices

$$B_1 = J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad (2.69a)$$

$$B_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad (2.69b)$$

$$B_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad (2.69c)$$

They satisfy the commutation rules

$$[B_1, B_2] = 2B_3 \qquad (2.70a)$$

$$[B_2, B_3] = -2B_1 \qquad (2.70b)$$

$$[B_3, B_1] = 2B_2. \qquad (2.70c)$$

<u>Exercise 2.29</u>: Verify the commutation rules (2.70) for Sp(2).

Suppose the basis elements B_i are used to evaluate (2.58). Making this calculation, one finds that the most general 2×2 symplectic matrix can be written in the form

$$M = \exp(b_2 B_2 + b_3 B_3)\ \exp(b_1 B_1), \tag{2.71}$$

where the b_i are arbitrary real coefficients. [Note that there are indeed three coefficients as predicted by Eq. (2.36) evaluated for n=1.] Thus, Eq. (2.71) gives a complete parameterization of the 2×2 symplectic group.

<u>Exercise 2.30</u>: Evaluate $\exp(b_2 B_2 + b_3 B_3)$ and $\exp(b_1 B_1)$ using (2.25) to find the results

$$\exp(b_2 B_2 + b_3 B_3) = I \cosh[(b_2^2 + b_3^2)^{1/2}]$$
$$+ [(b_2 B_2 + b_3 B_3)/(b_2^2 + b_3^2)^{1/2}] \sinh[(b_2^2 + b_3^2)^{1/2}], \tag{2.72}$$

$$\exp(b_1 B_1) = I \cos b_1 + B_1 \sin b_1 . \tag{2.73}$$

It is interesting to note that, according to Eq. (2.72), $\exp(b_2 B_2 + b_3 B_3)$ has the topology of two dimensional Euclidean space E^2 since b_2 and b_3 can each range from $\pm\infty$ without any duplication of results. By contrast, $\exp(b_1 B_1)$, according to (2.73), has the topology of a circle C since it is periodic in b_1 with period 2π. It follows that Sp(2) has the product topology $E^2 \times C$, and is therefore infinitely connected.

The case of Sp(4) is somewhat more complicated. The most general 4×4 real symmetric matrix S can be written in the block form

$$S = \begin{pmatrix} A & B \\ \tilde{B} & C \end{pmatrix}, \tag{2.74}$$

where the matrices A, B, and C are 2×2 and real, and the matrices A and C are themselves symmetric,

$$\tilde{A} = A\ ,\ \tilde{C} = C. \tag{2.75a,b}$$

Requiring that J commute with S gives the restrictions

$$\tilde{B} = -B \tag{2.76a}$$

$$C = A. \tag{2.76b}$$

<u>Exercise 2.31</u>: Verify the restrictions (2.76).

Thus, the most general S^c is of the form

$$S^c = \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \tag{2.77}$$

with the restrictions (2.75a) and (2.76a).
Correspondingly, JS^c is of the form

$$JS^c = \begin{pmatrix} -B & A \\ -A & -B \end{pmatrix} \quad . \tag{2.78}$$

The restriction (2.75a) requires that A be of the form

$$A = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + c\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.79}$$

where a, b, and c are arbitrary coefficients. The restriction (2.76a) requires that B be of the form

$$B = d\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad , \tag{2.80}$$

where d is also an arbitrary coefficient. Consequently, the Lie algebra spanned by matrices of the form JS^c is four dimensional. A convenient choice of basis elements is given by the matrices

$$B_0 = J = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \tag{2.81a}$$

$$B_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{2.81b}$$

$$B_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} \tag{2.81c}$$

$$B_3 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} . \tag{2.81d}$$

They satisfy the commutation rules

$$\begin{aligned} [B_0, B_i] &= 0 \quad , i = 0, 1, 2, 3 \\ [B_1, B_2] &= 2B_3 \\ [B_2, B_3] &= 2B_1 \\ [B_3, B_1] &= 2B_2 . \end{aligned} \tag{2.82}$$

<u>Exercise 2.32</u>: Verify the commutation rules (2.82).

The commutation rules (2.82) will be recognized as a variant of the commutation rules of the unitary group U(2). Indeed, let V be the unitary transformation

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} I & iI \\ iI & I \end{pmatrix} . \tag{2.83}$$

Then one finds from (2.78) the result

$$V^{-1}(JS^c)V = \begin{pmatrix} -B+iA & 0 \\ 0 & -B-iA \end{pmatrix} . \tag{2.84}$$

Matrices of the form $-B+iA$ with A and B real and obeying (2.75a) and (2.76a) span the space of $n \times n$ anti-Hermitian matrices in the general $n \times n$ case. Consequently, upon exponentiation, the matrices of the form $-B+iA$ generate the unitary group U(n), and the matrices $-B-iA$ generate the complex conjugate representation $\bar{U}(n)$. Therefore, the

Lie algebra spanned by the matrices JS^c is reducible, and is a variant of the Lie algebra of U(n) in the general case.

To study the matrix S^a in the 4×4 case, require that J anti-commute with the S of (2.74). One now finds the restrictions

$$\tilde{B} = B \quad , \tag{2.85a}$$

$$C = -A \quad . \tag{2.85b}$$

<u>Exercise 2.33</u>: Verify the restrictions (2.85).

Thus, the most general S^a is of the form

$$S^a = \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \tag{2.86}$$

with A and B real and subject to the restrictions (2.75a) and (2.85a). Correspondingly, JS^a is of the form

$$JS^a = \begin{pmatrix} B & -A \\ -A & -B \end{pmatrix} \quad . \tag{2.87}$$

As before, the most general matrix A satisfying (2.75a) can be written in the form

$$A = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + c\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \tag{2.88}$$

where a, b, and c are arbitrary coefficients. Similarly, the most general B satisfying (2.85a) can be written in the form

$$B = d\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + c\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + f\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.89}$$

Consequently, the vector space spanned by matrices of the form JS^a is six dimensional. Since the Lie algebra generated by the matrices JS^c is four dimensional, the complete Lie algebra generated by both the matrices JS^c and JS^a is ten dimensional. This result is in accord with Eq. (2.36) evaluated for n=2. That is, the Lie algebra of Sp(4) is ten dimensional.

After some agony, one finds that a possible and perhaps convenient choice for the six matrices required to provide a basis for matrices of the form JS^a is as follows:

$$F_1 = \begin{bmatrix} 0 & -1 & 0 & -1 \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{bmatrix} \tag{2.90a}$$

$$F_2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \tag{2.90b}$$

$$F_3 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix} \tag{2.90c}$$

$$G_1 = \begin{bmatrix} 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \tag{2.90d}$$

$$G_2 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \qquad (2.90e)$$

$$G_3 = \begin{bmatrix} -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} . \qquad (2.90f)$$

With this choice of basis, and using (2.58), the most general 4×4 symplectic matrix can be written in the form

$$M = \exp(\sum_1^3 f_i F_i + g_i G_i) \exp(\sum_0^3 b_i B_i). \qquad (2.91)$$

Exercise 2.34: Work out the commutation rules of the B's, F's, and G's for Sp(4).
Answer: See Table (2.1).

Exercise 2.35: Show that every matrix of the form $M = \exp(JS^a)$ is symplectic and has all its eigenvalues on the positive real axis. Show that every matrix of the from $M = \exp(JS^c)$ is symplectic and has all its eigenvalues on the unit circle.

Exercise 2.36: Show that the topology of Sp(2n) is that of $E^m \times U(n)$ with $m = n(n+1)$. Sp(2n) is therefore infinitely connected in general.

Research problem 2.1: Work out a suitable basis for the Lie algebra of Sp(6). According to the work done so far, elements of the form JS^c must generate a U(3). This requires 9 elements. Also, by (2.36), the Lie algebra of Sp(6) is 21 dimensional. Thus there must be 12 elements of the form JS^a.

	B_0	B_1	B_2	B_3	F_1	F_2	F_3	G_1	G_2	G_3
B_0		0	0	0	$2G_1$	$2G_2$	$2G_3$	$-2F_1$	$-2F_2$	$-2F_3$
B_1			$2B_3$	$-2B_2$	0	$2F_3$	$-2F_2$	0	$2G_3$	$-2G_2$
B_2				$2B_1$	$-2F_3$	0	$2F_1$	$-2G_3$	0	$2G_1$
B_3					$2F_2$	$-2F_1$	0	$2G_2$	$-2G_1$	0
F_1						$-4B_3$	$4B_2$	$-4B_0$	0	0
F_2							$-4B_1$	0	$-4B_0$	0
F_3								0	0	$-4B_0$
G_1									$-4B_3$	$4B_2$
G_2										$-4B_1$
G_3										

Table 2.1. Commutation rules for the Lie algebra Sp(4)

3. LIE ALGEBRAIC STRUCTURE OF CLASSICAL MECHANICS

3.1 Definition and properties of Poisson bracket

Let $H(q,p,t)$ be the Hamiltonian for some dynamical system and let f be any dynamical variable. That is, let $f(q,p,t)$ be any function of the phase space variables q,p and the time t. Consider the problem of computing the total time rate of change of f along a trajectory generated by H. According to the chain rule, this derivative is given by the expression

$$df/dt = \partial f/\partial t + \sum_i [(\partial f/\partial q_i)\, \dot{q}_i + (\partial f/\partial p_i)\, \dot{p}_i]. \qquad (3.1)$$

However, the $\dot{q}$'s and $\dot{p}$'s are given by Hamilton's equations of motion (1.11). Consequently, the expression for df/dt can also be written in the form

$$df/dt = \partial f/\partial t + \sum_i [(\partial f/\partial q_i)(\partial H/\partial p_i) - (\partial f/\partial p_i)(\partial H/\partial q_i)]. \qquad (3.2)$$

The second quantity appearing on the right of (3.2) occurs so often that it is given a special symbol and a special name in honor of Poisson. Let f and g be any two functions of the variables q,p,t. Then the Poisson bracket of f and g, denoted by the symbol $[f,g]$ is defined by the equation

$$[f,g] = \sum_i [(\partial f/\partial q_i)(\partial g/\partial p_i) - (\partial f/\partial p_i)(\partial g/\partial q_i)] . \qquad (3.3)$$

With this new notation, Eq. (3.2) can be written in the compact form

$$df/dt = \partial f/\partial t + [f,H]. \qquad (3.4)$$

Note that the Poisson bracket symbol $[,]$ is the same as that used earlier for a commutator. This is somewhat awkward, but unfortunately there are not always enough convenient symbols to go around.

Exercise 3.1: Evaluate the so-called fundamental Poisson brackets $[q_i,q_j]$, $[p_i,p_j]$, $[q_i,p_j]$.

The Poisson bracket operation has several remarkable properties. Upon calculation one finds the relations

1. Distributive property

$$[(af+bg),h] = a[f,h] + b[g,h] \qquad (3.5)$$

for arbitrary constants a,b.

2. Antisymmetry condition

$$[f,g] = -[g,f]. \qquad (3.6)$$

3. Jacobi identify

$$[f,[g,h]] + [g,[h,f]] + [h,[f,g]] = 0. \qquad (3.7)$$

4. Derivation with respect to ordinary multiplication

$$[f,gh] = [f,g]h + g[f,h] . \qquad (3.8)$$

Exercise 3.2: Verify Eqs. (3.5), (3.6), and (3.8). You may also try to verify (3.7). However, this is more easily done later on after the development of additional notation.

At this point it is useful to make some additional definitions of a mathematical nature: An algebra A over a field of numbers F is defined as a linear vector space supplemented by a rule for multiplying may two vectors to yield a third vector. This multiplication rule must satisfy certain properties. Indicating multiplication by a "o", we require that to every ordered pair of elements $x,y \in A$, there corresponds a third unique element of A, denote by $x \circ y$, and called the product of x and y. The product satisfies

1. $(cx) \circ y = x \circ (cy) = c(x \circ y)$

2. $(x + y) \circ z = x \circ z + y \circ z$

3. $x \circ (y + z) = x \circ y + x \circ z$

for any $x,y, z \in A$ and $c \in F$.

An example of an algebra is the set of all $N \times N$ matrices. The set of all $N \times N$ matrices forms an N^2 dimensional vector space. It also forms an algebra if we use for the "o" operation ordinary matrix multiplication. Note that in this case multiplication is associative, that is, the multiplication rule satisfies the property

$$(x \circ y) \circ z = x \circ (y \circ z).$$

A second example of an algebra is the set of all 3-vectors with the multiplication rule $\vec{a} \circ \vec{b} = \vec{a} \times \vec{b}$. Here "x" denotes the usual cross product. This algebra is not associative.

Exercise 3.3: Verify that this algebra is not associative.

A Lie algebra L is an algebra in which the multiplication rule (sometimes now called a Lie product) satisfies two further properties. For convenience, multiplication of x and y will now be denoted by the symbol $[x,y]$,

$$[x,y] = x \circ y.$$

In using this customary notation, however, it should be understood that the bracket [,] does not necessarily refer to the Poisson bracket. Rather, in this context, it refers to the Lie product abstractly, and independently of any particular realization.
The two additional properties for a Lie product are

4. $[x,y] = -[y,x]$ (antisymmetry)

5. $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$ (Jacobi condition).

Exercise 3.4: Verify that the algebra of 3-vectors with multiplication defined by $[\vec{a},\vec{b}] = \vec{a} \times \vec{b}$ is a Lie algebra.

Exercise 3.5: Verify that the set of all $N \times N$ matrices with the multiplication rule defined by $[A,B] = AB - BA$ forms a Lie algebra. In particular, check the Jacobi condition.

The reader will recognize that the general definition of a Lie algebra as just given is a bit more complicated than that used previously for matrices. However, exercise (3.5) shows that the matrix case is a special case of the general definition.

Now the stage is set for a more subtle conclusion. Observe that the set of all functions of the variables q,p,t forms a linear vector space. That is, any linear combination of two such functions is again a function. Now define the Lie product of any two functions to be the Poisson bracket (3.3). Equations (3.5) and (3.6) show that conditions 1 through 4 of a Lie algebra are satisfied. And Eq. (3.7) shows that condition 5 is satisfied. Consequently, the set of functions of the variables q,p,t forms a Lie algebra! This Lie algebra will be called the Poisson bracket Lie algebra of dynamical variables.

Exercise 3.6: Determine the dimensionality of the Poisson bracket Lie algebra of dynamical variables.
Answer: The set of functions of q,p,t is an infinite dimensional vector space.

At this point it is convenient to introduce a more compact notation for the phase space variables $(q_1 \ldots q_n)$, $(p_1 \ldots p_n)$. To do this, introduce the 2n variables $(z_1 \ldots z_{2n})$ by this rule

$$z_i = q_i \quad , \quad i = 1, \quad n \tag{3.9a}$$

$$z_{n+i} = p_i \quad , \quad i = 1, \quad n. \tag{3.9b}$$

That is, the first n z's are the q's and the last n z's are the p's.

With the definition (3.9), it is easily verified that the fundamental Poisson brackets $[z_i, z_j]$ are given by the relation

$$[z_i, z_j] = J_{ij}, \tag{3.10}$$

where J is the fundamental $2n \times 2n$ matrix given by (2.1) and used in defining symplectic matrices.

Exercise 3.7: Verify Eq. (3.10).

Also, suppose functions f and g of the variables q,p,t are written more compactly as f(z,t), g(z,t). Then the general Poisson bracket (3.3) can be written more compactly in the form

$$[f,g] = \sum_{i,j} (\partial f/\partial z_i)\, J_{ij}\, (\partial g/\partial z_j). \tag{3.11}$$

Suppose further that the 2n quantities $(\partial f/\partial z_i)$ are viewed as the components of a vector conveniently written as $\partial_z f$, etc. Then the right-hand side of (3.11) can be viewed as a combination of two vectors and a matrix which can be written even more compactly using matrix and scalar product notation,

$$[f,g] = (\partial_z f,\ J\, \partial_z g). \tag{3.12}$$

Exercise 3.8: Verify Eqs. (3.11) and (3.12) starting from the definition (3.3).

Exercise 3.9: Verify the Jacobi identity (3.7).

3.2 Equations, constants, and integrals of motion

It has already been shown that any dynamical variable f(z,t) of a dynamical system governed by a Hamiltonian H obeys the equation of motion

$$df/dt = \partial f/\partial t + [f,H]. \tag{3.13}$$

Exercise 3.10: Show that the dynamical variables z_i obey the equations of motion

$$z_i = (J\, \partial_z H)_i. \tag{3.14}$$

A dynamical variable f is called a constant of motion if its total time derivative vanishes. In view of (3.13), a constant of motion satisfies the equation

$$\partial f/\partial t + [f,H] = 0. \tag{3.15}$$

<u>Exercise 3.11</u>: Suppose that the dynamical variables f and g are constants of motion. Verify Poisson's theorem which states that the quantity $[f,g]$ is then also a constant of motion.
<u>Hint</u>: Use the Jacobi identity.

It can be shown in general that any Hamiltonian dynamical system with n degrees of freedom has 2n functionally independent constants of motion.

Suppose that a constant of motion f does not explicitly depend on the time t,

$$\partial f/\partial t = 0. \tag{3.16}$$

A constant of motion which does not explicitly depend on the time will be called an <u>integral of motion</u>. Evidently, an integral of motion obeys the equation

$$[f,H] = 0. \tag{3.17}$$

<u>Exercise 3.12</u>: Suppose that f and g are integrals of motion. Show that $[f,g]$ is then also an integral of motion.

<u>Exercise 3.13</u>: Suppose that the Hamiltonian H for a dynamical system does not depend explicitly on the time t. Show that then H is an integral of motion.

The question of the existence of integrals of motion is quite complicated. Observe that if $f(z)$ is an integral of motion, then any given trajectory must remain for all time on a general hypersurface in phase space defined by an equation of the form

$$f(z) = \text{constant}.$$

If there are several functionally independent integrals of motion, then the general trajectory is further restricted to lie in the intersection of several hypersurfaces for all time. Thus, the greater the number of integrals, the more can be said about the behavior of a dynamical system.

Consider a time independent Hamiltonian $H(z)$. A point z^o in phase space for which the vector $\partial_z H$ is zero is called a <u>critical point</u>. Evidently, according to Eq. (3.14), a critical point is some kind of equilibrium point. Now suppose some small region R of phase space contains <u>no</u> critical points. Then, it can be shown that provided R is small enough, the Hamiltonian $H(z)$ has 2n-1 functionally independent integrals of motion in the region R. Furthermore, n of these integrals can be arranged to be in involution. (Two functions f and g are said to be in involution if their Poisson bracket $[f,g]$ is zero.)

The result just stated is of limited use unless all trajectories starting in R happen to remain in R. In general, and contrary to the impression given by most textbooks, most dynamical Hamiltonian systems do not have global integrals of motion. If a time independent

Hamiltonian dynamical system with n degrees of freedom has n global integrals of motion in involution, the system is said to be completely integrable. In general, only the soluble problems found in textbooks fall into this category. Most Hamiltonian dynamical systems, including the majority encountered in real life, are not completely integrable and are therefore sufficiently complicated to be in some sense insoluble. In particular, the behavior of most Hamiltonian systems is sufficiently complicated that the trajectories are not generally confined to lie on hypersurfaces in phase space.[7-9]

3.3 Lie operators

Let $f(z,t)$ be some function, and let $g(z,t)$ be any other function. Associated with each f is a Lie operator which acts on general functions g. The Lie operator associated with the function f will be denoted by the symbol $:f:$ and it is defined in terms of Poisson brackets by the rule

$$:f:g = [f,g]. \tag{3.18}$$

In an analogous way, powers of $:f:$ are defined by taking repeated Poisson brackets. For example, $:f:^2$ is defined by the relation

$$:f:^2 g = [f, [f,g]]. \tag{3.19}$$

Finally, $:f:$ to the zero power is defined to be the identity operator,

$$:f:^0 g = g. \tag{3.20}$$

Evidently, a Lie operator, as well as its powers, is a linear operator because of Eq. (3.5). For the same reason, the sum of two Lie operators is again a Lie operator. Specifically, one finds the relation

$$a:f: + b:g: = :(af + bg): \tag{3.21}$$

for any two scalars a,b and any two functions f,g. Therefore, the set of Lie operators forms a linear vector space.

A Lie operator is also a derivation with respect to the operation or ordinary multiplication. That is, a Lie operator satisfies the product rule analogous to that for differentiation: Let g and h be any two functions. Then, according to Eq. (3.8), $:f:$ obeys the rule

$$:f:(gh) = (:f:g)h + g(:f:h). \tag{3.22}$$

Exercise 3.14: Starting from (3.22), show that $:f:^n$ obeys the Leibniz rule

$$:f:^n(gh) = \sum_{m=0}^{n} \binom{n}{m}(:f:^m g)(:f:^{n-m} h), \tag{3.23}$$

where $\binom{n}{m}$ is the binomial coefficient defined by

$$\binom{n}{m} = \frac{n!}{(m!)(n-m)!} \quad . \tag{3.24}$$

In addition to being a derivation with respect to ordinary multiplication, a Lie operator is also a derivation with respect to Poisson bracket multiplication. Suppose g and h are any two functions. Then the Jacobi identity (3.7) can be written in the form

$$[f,[g,h]] = [[f,g],h] + [g,[f,h]], \tag{3.25}$$

or equivalently, using Lie operator notation,

$$:f:[g,h] = [:f:g,h] + [g,:f:h] \ . \tag{3.26}$$

<u>Exercise 3.15</u>: Verify Eqs. (3.25) and (3.26).

<u>Exercise 3.16</u>: Write and verify the analog of the Leibniz rule of exercise (3.14) for the case of $:f:^n[g,h]$.

Since the set of Lie operators forms a linear vector space, it is of interest to inquire whether the vector space can be given a multiplication rule which will convert it into a Lie algebra. The answer is yes, as is nearly obvious, since Lie operators are linear operators and linear operators are quite similar to matrices. The Lie product of two Lie operators :f: and :g: is simply taken to be their commutator. Denoting the Lie product of two Lie operators by the symbol :f:, :g: , the Lie product is defined by the rule

$$\{:f:,:g:\} = :f::g: - :g::f:. \tag{3.27}$$

Note that there are now two Lie algebras which have to be kept in mind. First, there is the Lie algebra of functions of z,t with the Lie product defined to be the Poisson bracket. Second, there is the Lie algebra of Lie operators with the Lie product defined to be the commutator.

There is, however, one point which has been overloooked. Namely, is the right-hand side of (3.27) a Lie operator? To answer this question, it is useful to view the Jacobi identity (3.7) for Poisson brackets form yet another perspective. For any function h, the Jacobi identity can be written in the form

$$[f,[g,h]] - [g,[f,h]] = [[f,g],h]. \tag{3.28}$$

However, using Lie operator notation, this same equation can be written in the form

$$:f::g:h - :g::f:h = :[f,g]:h, \tag{3.29}$$

or more compactly, using (3.27),

$$\{:f:,:g:\}h = :[f,g]:h. \tag{3.30}$$

But, since h is an arbitrary function, Eq. (3.29) can also be viewed as the operator identity

$$\{:f:,:g:\} = :[f,g]:. \tag{3.31}$$

Evidently, the commutator of two Lie operators :f: and :g: is again a Lie operator, and is in fact the Lie operator associated with the function [f,g].

Put another way, Eq. (3.31) shows that there is a close connection between the Lie algebra of functions and the Lie algebra of Lie operators. Specifically, the Lie product (commutator) of two Lie operators is the Lie operator of the Lie product (Poisson bracket) of the two associated functions. Mathematicians have a word for such a situation. They would say that the two Lie algebras are homomorphic.

Exercise 3.17: Verify Eqs. (3.28), (3.29), and (3.30).

Exercise 3.18: Verify that the Lie product defined by (3.27) satisfies all the properties required to make the set of Lie operators into a Lie algebra. See exercise (3.5).

3.4 Lie Transformations

Since powers of :f: have been defined, it is also possible to deal with power series in :f:. Of particular importance is the power series exp(:f:). This particular object is called the Lie transformation associated with :f: or f. The Lie transformation is also a linear operator, and is formally defined as expected by the exponential series

$$\exp(:f:) = \sum_{n=0}^{\infty} :f:^n/n!. \tag{3.32}$$

In particular, the action of exp(:f:) on any function g is given by the rule

$$\exp(:f:)\, g = g + [f,g] + [f,[f,g]]/2! + \cdots. \tag{3.33}$$

Exercise 3.19: Let q and p be the phase-space coordinates for a system having one degree of freedom. Let f be the function

$$f = -\lambda p^2/2 .$$

Show that

$$\exp(:f:)\, p = p$$

$$\exp(:f:)\, q = q + \lambda p .$$

Here λ is an arbitrary parameter.
Hint: Observe that the series (3.33) terminates in this case.

Exercise 3.20: Repeat exercise (3.19) for the case $f = \lambda q^2$.

Exercise 3.21: Repeat exercise (3.19) for the case $f = \lambda q^3$.

Exercise 3.22: Repeat exercise (3.19) for the case $f = -\lambda pq$. Now you must sum an infinite series.

Answer:

$$\exp(:f:)\, q = (e^{\lambda})\, q$$

$$\exp(:f:)\, p = (e^{-\lambda})\, p.$$

Exercise 3.23: Repeat exercise (3.19) for the case $f = -\lambda(p^2 + q^2)/2$,
Answer:

$$\exp(:f:)\, q = q \cos \lambda + p \sin \lambda$$

$$\exp(:f:)\, p = -q \sin \lambda + p \cos \lambda .$$

The fact that :f: is a derivation with respect to ordinary multiplication, see Eq. (3.22), implies that the Lie transformation exp(:f:) is an isomorphism with respect to ordinary multiplication. This is another remarkable property of the exponential function! That is, suppose g and h are any two functions. Then the Lie transformation exp(:f:) has the property

$$\exp(:f:)\,(gh) = (\exp(:f:)g)\,(\exp(:f:)h). \qquad (3.34)$$

In words, Eq. (3.34) says that one can either let a Lie transformation act on the product of two functions, or act on each function separately and then take the product of the results. Both operations give the same net result.

Exercise 3.24: Verify Eq. (3.34) for the case $f = \lambda q^2$ and $g = h = p$.

Exercise 3.25: Prove Eq. (3.34), using the definition (3.32) and the Leibniz rule (3.23), by expansion and resummation of various series.

The isomorphism property of exp(:f:) described by (3.34) often facilitates computations involving Lie transformations. Let the symbol z stand, as usual, for the collection of quantities $z_1 \ldots z_{2n}$. Similarly, let the symbol exp(:f:)z stand for the collection of quantities $\exp(:f:)z_1 \ldots \exp(:f:)z_{2n}$. Now let g(z) be any function. Then it follows from (3.34) that

$$\exp(:f:)\, g(z) = g[\exp(:f:)z]. \tag{3.35}$$

That is, the action of a Lie transformation on a function is to perform a Lie transformation on its arguments.

To see the truth of (3.35), suppose first that g were a polynomial in the quantities $z_1 \ldots z_{2n}$. But a polynomial is just a sum of monomials of the form

$$z_1^{m_1}\, z_2^{m_2} \ldots z_{2n}^{m_{2n}} \quad .$$

It follows from (3.34) that

$$\exp(:f:)\, z_1^{m_1} \ldots z_{2n}^{m_{2n}} = [\exp(:f:)z_1]^{m_1} \cdots [\exp(:f:)z_{2n}]^{m_{2n}} \, . \tag{3.36}$$

Also, as commented earlier, exp(:f:) is a linear operator. Therefore a Lie transformation has the advertised property (3.35) when acting on polynomials. But the set of polynomials is dense in the complete set of functions on any bounded domain. Consequently, (3.35) holds in general by continuity.

The last observation to be made is that since :f: is also a derivation with respect to Poisson bracket multiplication, the Lie transformation exp(:f:) must also be an isomorphism with respect to Poisson bracket multiplication. That is, suppose g and h are any two functions. Then the Lie transformation exp(:f:) has the property

$$\exp(:f:)\,[g,h] = [\exp(:f:)\, g, \exp(:f:)\, h]. \tag{3.37}$$

This property will be essential for subsequent discussions of symplectic maps and charged particle beam transport.

> Exercise 3.26: Derive (3.37) from the definition (3.32) and the results of exercise (3.16). Mutatis mutandis, what is required is a repeat of exercise (3.25).

4. SYMPLECTIC MAPS

4.1 Definitions

Let $z_1 \ldots z_{2n}$ be a set of canonical variables for some Hamiltonian dynamical system. Suppose a transformation is made to some new set of variables $\bar{z}_1(z,t) \ldots \bar{z}_{2n}(z,t)$.[10] Such a transformation will be called a mapping, and will be denoted by the symbol $\mathcal{M}$,

$$\mathcal{M}: \quad z \to \bar{z}(z,t). \tag{4.1}$$

Also, let $M(z,t)$ be the Jacobian matrix of the map $\mathcal{M}$. It is defined by the equation

$$M_{ab}(z,t) = \partial \bar{z}_a / \partial z_b . \tag{4.2}$$

The map $\mathcal{M}$ is said to be symplectic if its Jacobian matrix M is a symplectic matrix for all values of z and t,

$$\tilde{M}\, J\, M = J, \quad \text{or} \quad M\, J\, \tilde{M} = J. \tag{4.3}$$

Note that in general M depends on z and t. However, the particular combinations $\tilde{M} J M$ or $M J \tilde{M}$ must be z and t independent. Therefore, a symplectic map must have very special properties.

To appreciate the significance of a symplectic mapping, consider the Poisson brackets of the various $\bar{z}$'s with each other. Using Eq. (3.11), one finds the result

$$[\bar{z}_a, \bar{z}_b] = \sum_{c,d} (\partial \bar{z}_a / \partial z_c)\, J_{cd}\, (\partial \bar{z}_b / \partial z_d). \tag{4.4}$$

By using the definition (4.2) of the Jacobian matrix M, Eq. (4.4) can also be written in the form

$$\begin{aligned} [\bar{z}_a, \bar{z}_b] &= \sum_{c,d} M_{ac}\, J_{cd}\, M_{bd} \\ &= \sum_{c,d} M_{ac}\, J_{cd}\, \tilde{M}_{db} \\ &= (M\, J\, \tilde{M})_{ab} \,. \end{aligned} \tag{4.5}$$

Finally, upon using the symplectic condition (4.3), one finds the result

$$[\bar{z}_a, \bar{z}_b] = (M\, J\, \tilde{M})_{ab} = J_{ab} = [z_a, z_b]. \tag{4.6}$$

Consequently, the necessary and sufficient condition for a map $\mathcal{M}$ to be symplectic is that it preserve the fundamental Poisson brackets (3.10). This statement is equivalent, in turn, to the condition that the map $\mathcal{M}$ must preserve the Poisson bracket Lie algebra of all dynamical variables.

Symplectic mappings also have a geometrical apsect. Let z^o be some point in phase space, and suppose it is sent to the point $\bar{z}^o$ under the action of a symplectic map $\mathcal{M}$. Also, let dz and δz be two small vectors originating at z^o. Under the action of $\mathcal{M}$, they are sent to two vectors $d\bar{z}$ and $\delta \bar{z}$ originating at the point $\bar{z}^o$. See Fig. 4.1. According to the chain rule, one has the relation

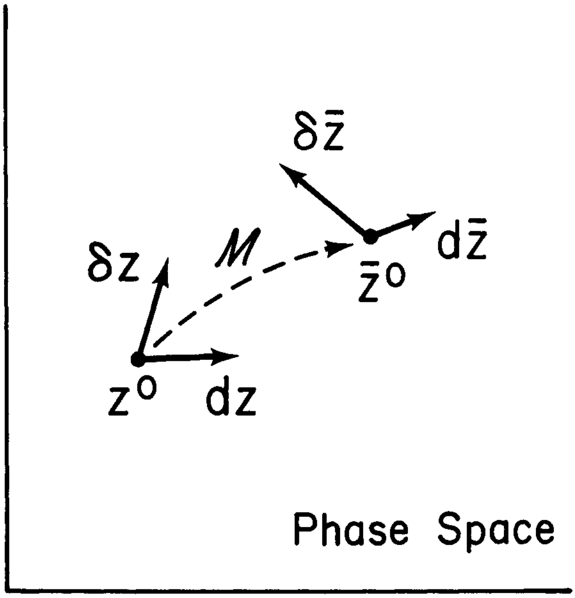


Fig. 4.1: The action of a symplectic map M on phase space. The general point z^o is mapped to the point $\bar{z}^o$, and the small vectors dz and δz are mapped to the small vectors $d\bar{z}$ and $\delta \bar{z}$. The figure is only schematic since in general phase space has a large number of dimensions.

$$d\bar{z}_a = \sum_b (\partial \bar{z}_a/\partial z_b)\, dz_b, \tag{4.7}$$

or more compactly, using (4.2),

$$d\bar{z} = M\, dz. \tag{4.8}$$

Similarly, the vectors δz and $\delta \bar{z}$ are related by the equation

$$\delta \bar{z} = M\, \delta z. \tag{4.9}$$

Now use the two vectors $\delta \bar{z}$, $d\bar{z}$ and the matrix J to form the quantity $(\delta \bar{z}, J d\bar{z})$. This quantity is called the <u>fundamental symplectic two-form</u>. (Given any two vectors, a two-form is a rule for computing an associated number. The rule must be such that the answer depends linearly on each vector.) Suppose the relations (4.8) and (4.9) are inserted into the two-form $(\delta \bar{z}, J d\bar{z})$. Then, using matrix manipulation and the symplectic condition (4.3), one finds the relation

$$(\delta \bar{z},\ J\, d\bar{z}) = (M\, \delta z,\ J\, M\, dz) = (\delta z,\ \tilde{M}\, J\, M\, dz) = (\delta z,\ J\, dz). \tag{4.10}$$

That is, the value of the fundamental symplectic two-form is unchanged by a symplectic map. Evidently, a necessary and sufficient condition for a map to be symplectic is that it preserve the fundamental symplectic two-form at all points of phase space and for all time.

> <u>Exercise 4.1</u>: Consider a two-dimensional phase space consisting of the variables q,p. Evaluate the quantity $(\delta z, J\, dz)$ and show that it is related to the area formed by the small parallelogram with sides δz and dz.

There is a third aspect of symplectic mappings which should already be familiar. In the usual treatments of Classical Mechanics, an important topic is that of canonical transformations. Canonical transformations are usually defined as those transformations which either

a. preserve the Hamiltonian form of the equations of motion for all Hamiltonian dynamical systems, or
b. preserve the fundamental Poisson brackets.

In case b, according to the previous discussion, canonical transformations and symplectic maps are the same thing. In case a, it can be shown that the most general canonical transformation is a map $\mathcal{M}$ whose Jacobian matrix satisfies the condition

$$\tilde{M}\, J\, M = \lambda J, \tag{4.11}$$

where λ is some real nonzero constant independent of z and t.[11] Furthermore, it can be shown that $\mathcal{M}$ in this case consists of a symplectic map followed or preceded by a simple scaling of phase-space variables. Therefore, in either case, the central object of interest is a symplectic map.

4.2 Group Properties

Let $\mathcal{M}$ be a symplectic mapping of z to $\bar{z}$, and suppose it has an inverse $\mathcal{M}^{-1}$,

$$\mathcal{M}: z \to \bar{z} \tag{4.12a}$$

$$\mathcal{M}^{-1}: \bar{z} \to z. \tag{4.12b}$$

According to (4.8), the relation between a small change dz in z, and the associated small change $d\bar{z}$ in $\bar{z}$ is given by the Jacobian matrix M of $\mathcal{M}$. Since M is symplectic, it has an inverse M^{-1}. Therefore, Eq. (4.8) can be inverted to give the relation

$$dz = M^{-1}\, d\bar{z}. \tag{4.13}$$

But now, comparison of (4.12b) and (4.13) shows that the Jacobian matrix of $\mathcal{M}^{-1}$ is M^{-1}. Note also that the local existence of $\mathcal{M}^{-1}$ did not really have to be assumed, but follows instead from the implicit function theorem since M^{-1} is known to exist from the symplectic condition. Finally, the matrix M^{-1} is symplectic since the inverse of a symplectic matrix is also a symplectic matrix. It follows that $\mathcal{M}^{-1}$ is a symplectic map. What has been shown is that if $\mathcal{M}$ is a symplectic map, then $\mathcal{M}^{-1}$ exists (at least locally) and is also a symplectic map.

Next suppose that $\mathcal{M}^{(1)}$ is a symplectic mapping of z to $\bar{z}$ and $\mathcal{M}^{(2)}$ is a symplectic mapping of $\bar{z}$ to another set of variables $\bar{\bar{z}}$. Now consider the composite mapping $\mathcal{M} = \mathcal{M}^{(2)}\, \mathcal{M}^{(1)}$ which sends z to $\bar{\bar{z}}$.

$$\mathcal{M} = \mathcal{M}^{(2)}\, \mathcal{M}^{(1)} \tag{4.14}$$

$$\mathcal{M}^{(1)} : z \to \bar{z} \tag{4.15a}$$

$$\mathcal{M}^{(2)} : \bar{z} \to \bar{\bar{z}} \tag{4.15b}$$

$$\mathcal{M}^{(2)}\mathcal{M}^{(1)} : z \to \bar{\bar{z}}. \tag{4.15c}$$

According to the chain rule, the Jacobian matrix M of the composite mapping $\mathcal{M}$ is the product of the Jacobian matrices of $\mathcal{M}^{(2)}$ and $\mathcal{M}^{(1)}$,

$$M = M^{(2)}\, M^{(1)}. \tag{4.16}$$

However, the matrices $M^{(2)}$ and $M^{(1)}$ are symplectic since they are the Jacobian matrices of symplectic maps. It follows from (4.16) and the group property for symplectic matrices that M is also a symplectic matrix. Consequently, the composite mapping M is also a symplectic map. What has been shown is that if $M^{(1)}$ and $M^{(2)}$ are symplectic maps, so is their product $M^{(2)} M^{(1)}$.

It is also obvious that the identity mapping which sends each z into itself is a symplectic map, for the Jacobian matrix of this map is evidently the identity matrix, and the identity matrix is symplectic.

The previous discussion has shown that the set of symplectic maps has properties very analogous to the group properties of the group of symplectic matrices. As defined earlier, the concept of a group applied only to matrices. However, it is clear that the concept of a group can be enlarged to include the possibility of general mappings. When this is done, the set of all symplectic maps is entitled to be called a group.

4.3 Relation to Hamiltonian Flows

Let H(z,t) be the Hamiltonian for some dynamical system. Consider a large Euclidean space with 2n+1 axes labeled by the variables $z_1 \ldots z_{2n}$ and the time t. This construction will be called state space. See Fig. 4.2. Suppose the 2n quanties $z_1(t^o) \ldots z_{2n}(t^o)$ are specified at some initial time t^o. Then the quantities $z_1(t) \ldots z_{2n}(t)$ at some other time t are uniquely determined by the initial conditions $z_1(t^o) \ldots z_{2n}(t^o)$ and Hamilton's equations of motion (3.14). The set of all trajectories in state space for all possible initial conditions will be called a Hamiltonian flow.

Let t^o be some initial time, and let t^f be some other final time. Also, let $z^{(0)}$ denote the set of quantities $z_1(t^o) \ldots z_{2n}(t^o)$, and let $z^{(f)}$ denote the corresponding set $z_1(t^f) \ldots z_{2n}(t^f)$. Then it is a remarkable fact that the relation between the quantities $z^{(0)}$ and $z^{(f)}$ is a symplectic map.

Theorem 4.1: Let H(z,t) be the Hamiltonian for some dynamical system, and let $z^{(0)}$ denote a set of initial conditions at some initial time t^o. Also, let $z^{(f)}$ denote the coordinates at some time t^f of the trajectory with initial conditions $z^{(0)}$. Finally, let M denote the mapping from $z^{(0)}$ to $z^{(f)}$ obtained by following the Hamiltonian flow specified by H,

$$M: \; z^{(0)} \rightarrow z^{(f)}. \tag{4.17}$$

Then the mapping M is symplectic.

Proof: Suppose the flow takes place for a time interval of duration T so that t^o and t^f are related by the equation

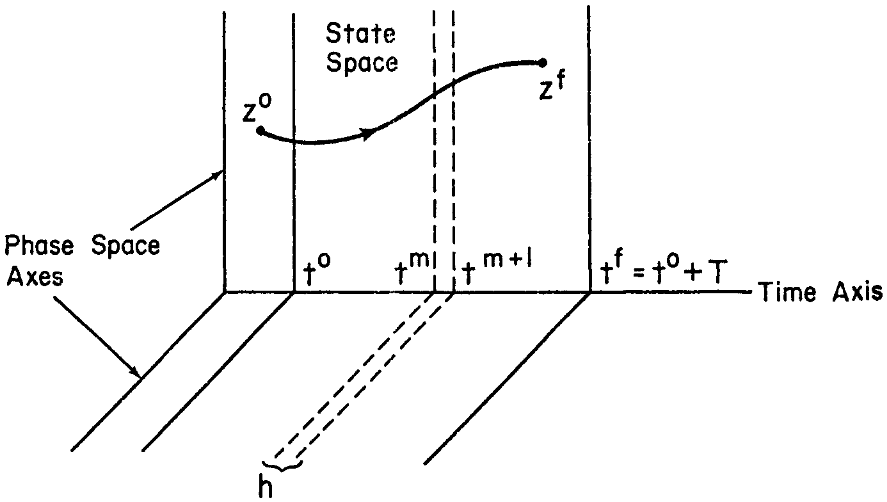


Fig. 4.2: A trajectory in state space. Under the Hamiltonian flow specified by a Hamiltonian H, the general phase-space point z^o is mapped into the phase-space point z^f. The mapping is symplectic for any Hamiltonian.

$$t^f = t^o + T. \tag{4.18}$$

Divide the interval T into N small steps each of duration h. Evidently, T, N, and h are related by the equation

$$T = Nh. \tag{4.19}$$

Also, define intermediate times t^m at each step by the rule

$$t^m = t^o + mh, \quad m = 0,1, \cdots N. \tag{4.20}$$

Suppose that the mapping $\mathcal{M}$ is viewed as a composite of mappings betwen adjacent times t^m and t^{m+1}. That is, $\mathcal{M}$ is written in the form

$$\mathcal{M} = \mathcal{M}^{t^f \leftarrow t^{N-1}} \cdots \mathcal{M}^{t^{m+1} \leftarrow t^m} \cdots \mathcal{M}^{t^1 \leftarrow t^o} \tag{4.21}$$

with the notation that $\mathcal{M}^{t^{m+1} \leftarrow t^m}$ denotes the mapping between the quantities

$$z^{(m)} = z_1(t^m) \cdots \ z_{2n}(t^m)$$

and

$$z^{(m+1)} = z_1(t^{m+1}) \cdots \ z_{2n}(t^{m+1}). \tag{4.22}$$

Corresponding to the relation (4.21), the Jacobian matrix M of the mapping $\mathcal{M}$ can be written using the chain rule in the product form

$$M = M^{t^f \leftarrow t^{N-1}} \cdots M^{t^{m+1} \leftarrow t^m} \cdots M^{t^1 \leftarrow t^0}, \tag{4.23}$$

where, as the notation is meant to indicate, $M^{t^{m+1} \leftarrow t^m}$ is the Jacobian matrix for the map $\mathcal{M}^{t^{m+1} \leftarrow t^m}$.

Next it will be shown that each matrix in the product (4.23) is symplectic at least through terms of order h. According to Taylor's series, the relation between $z^{(m+1)}$ and $z^{(m)}$ can be written in the form

$$\begin{aligned} z_a^{(m+1)} &= z_a(t^{m+1}) = z_a(t^m+h) \\ &= z_a(t^m) + h\,\dot{z}_a(t^m) + O(h^2) \\ &= z_a^{(m)} + h(J\,\partial_z H)_a + O(h^2). \end{aligned} \tag{4.24}$$

Here use has also been made of the equations of motion (3.14). Suppose Eq. (4.24) is used to compute the associated Jacobian matrix. The result of this computation is the relation

$$M_{ab}^{t^{m+1}\leftarrow t^{m}} = \partial z_a^{(m+1)} / \partial z_b^{(m)}$$

$$= \delta_{ab} + h \sum_c J_{ac} \, \partial^2 H/\partial z_c \partial z_b + O(h^2). \quad (4.25)$$

Using matrix notation, Eq. (4.25) can be written more compactly in the form

$$M^{t^{m+1}\leftarrow t^{m}} = I + h\, J\, S + O(h^2), \quad (4.26)$$

where S is the symmetric matrix

$$S_{cb} = \partial^2 H/\partial z_c \partial z_b. \quad (4.27)$$

Now compare Eq. (4.26) with Eqs. (2.28) and (2.31). Evidently, the Jacobian matrix (4.26) is a symplectic matrix at least through terms of order h.

The desired proof is almost complete. Since symplectic matrices form a group, the product matrix M given by (4.23) differs from a symplectic matrix by terms at most of order Nh^2 because each of the N terms in the product differs from a symplectic matrix by terms at most of order h^2. Now take the limit $h \to 0$ and $N \to \infty$. In this limit terms proportional to Nh^2 vanish since, using (4.19),

$$Nh^2 = (T/h)h^2 = Th, \quad (4.28)$$

and the quantity Th vanishes as h goes to zero. It follows that M is a symplectic matrix, and $\mathcal{M}$ is a symplectic map.

What has been shown is that the problem of describing and following Hamiltonian flows, which is one of the fundamental aspects of classical mechanics, is equivalent to the problem of calculating and representing symplectic maps.

Exercise 4.2: Let H be the harmonic oscillator Hamiltonian given by

$$H = (z_2^2 + \omega^2 z_1^2)/2. \quad (4.29)$$

Find the map $\mathcal{M}$ produced by the Hamiltonian flow governed by H. Verify that $\mathcal{M}$ is a symplectic map.

Answer: The mapping $\mathcal{M}$ is given by

$$\mathcal{M}: \quad z^o \to z^f$$

with

$$z_1^f = z_1^o \cos\phi + (z_2^o/\omega)\sin\phi$$

$$z_2^f = -\omega z_1^o \sin\phi + z_2^o \cos\phi \tag{4.30}$$

where

$$\phi = \omega(t^f - t^o). \tag{4.31}$$

The reader should find the associated Jacobian matrix M and verify that it is symplectic.

Exercise 4.3: Consider the behavior of light rays in an optical system as shown in Fig. (1.4) and described in exercise (1.14). Let $\vec{p}$ be a two-component vector with entries p_x and p_y, and let $\vec{q}$ be a two-component vector with entries $q_x=x$ and $q_y=y$. Evidently, a ray leaving the initial point P^i is characterized by the initial quantities $\vec{q}^i$ and $\vec{p}^i$. The quantity $\vec{q}^i$ specifies the initial point of origin of the ray in the object plane, and, according to (1.29), $\vec{p}^i$ describes the initial direction of the ray. Similarly, $\vec{q}^f$ and $\vec{p}^f$ characterize the ray as it arrives at the final point P^f in the image plane. Show that the relationship between the initial quantities $\vec{q}^i$, $\vec{p}^i$ and the final quantities $\vec{q}^f$, $\vec{p}^f$ is given by a symplectic map.

4.4 Liouville's theorem and the Poincare invariants

It has been seen that Hamiltonian flows generate symplectic maps, and that symplectic maps have special properties. It follows that Hamiltonian flows must have special properties. One of these properties is described by Liouville's theorem.

Consider an ensemble of noninteracting systems with each member of the ensemble governed by the same Hamiltonian $H(z,t)$. At some initial instant t^i, let each member of the ensemble be characterized by a point in phase space corresponding to its initial conditions. Suppose, further that all the points of the ensemble at the initial instant t^i occupy a certain region R^i of phase space. The volume V^i initially occupied by the ensemble in phase space is given by the integral

$$V^i = \int_{R^i} dz_1^i \dots dz_{2n}^i. \qquad (4.32)$$

Now follow all the trajectories of the members of the ensemble through state space to some later instant t^f. The members of the ensemble will then occupy some final region R^f of phase space, and the volume of this region will be given by the integral

$$V^f = \int_{R^f} dz_1^f \dots dz_{2n}^f . \qquad (4.33)$$

Evidently, the relation between the quantities z^i and z^f for each member of the ensemble is given by a common symplectic map $\mathcal{M}$. Also, the relation between the differential quantities dz^i and dz^f is given by the Jacobian matrix M belonging to $\mathcal{M}$,

$$dz^f = M\, dz^i. \qquad (4.34)$$

It follows from the standard rules for changing variables of integration that the volume V^f is also given by the relation

$$V^f = \int_{R^i} |\det M|\, dz_1^i \dots dz_{2n}^i . \qquad (4.35)$$

But, since M is a symplectic matrix, it must have determinant +1. Therefore, comparison of (4.35) and (4.32) shows that the two volumes V^f and V^i are the same,

$$V^f = V^i . \qquad (4.36)$$

This is Liouville's theorem.

There is a slighly different phrasing of Liouville's theorem which is also worth mentioning. By construction, the number of ensemble points in V^f and V^i is the same. Therefore, since V^f equals V^i, one may also say that the density of points in phase space is preserved by Hamiltonian flows.

The volume invariant of Liouville's theorem is actually the last in a hierarchy of invariants called the Poincare invariants. The first invariant in the series consists of a certain 2-dimensional integral over a 2-dimensional hypersurface in phase space. The next consists of a 4-dimensional integral over a 4-dimensional hypersurface, etc. The last consists of a 2n dimensional integral which is just the volume of Liouville's theorem.

A complete and proper discussion of all the Poincare invariants requires the use of the calculus of differential forms. However, the first in the series of invariants is easily discussed using the fundamental symplectic two-form $(\delta z, Jdz)$ introduced earlier.

Suppose the points of an ensemble at the instant t^i lie on a 2-dimensional hypersurface R_2^i. Let this surface be parameterized by the two quantities α, β. That is, there are 2n relations of the form

$$z_j = f_j(\alpha, \beta) \tag{4.37}$$

which describe the surface. Next consider the integral over the surface R_2^i given by

$$I_2^i = \int_{R_2^i} (\delta z^i, Jdz^i). \tag{4.38}$$

The integral (4.38) is to be understood as follows: Let dz^i be the vector formed using (4.37) when only α is allowed to vary

$$dz_j^i = (\partial f_j/\partial\alpha)\, d\alpha \tag{4.39a}$$

or, in vector notation,

$$dz^i = \partial_\alpha f\, d\alpha. \tag{4.39b}$$

Similarly, let δz^i be the vector formed when only β is allowed to vary,

$$\delta z^i = \partial_\beta f\, d\beta\,. \tag{4.39c}$$

Then the integral (4.38) can also be written in the form

$$I_2^i = \int_{R_2^i} (\partial_\beta f,\ J\,\partial_\alpha f)\, d\alpha d\beta. \tag{4.40}$$

Now follow, as before, the trajectories of the members of the ensemble through state space to some later instant t^f. Then the points will lie on some other 2-dimensional hypersurface R_2^f, and one can form the associated integral

$$I_2^f = \int_{R_2^f} (\delta z^f,\ J\, dz^f). \tag{4.41}$$

However, by a change of variables using (4.34) and its counterpart for δz, the integral (4.41) can also be written in the form

$$I_2^f = \int_{R_2^i} (M\delta z^i,\ J\ M\ dz^i). \tag{4.42}$$

But, as found earlier in Eq. (4.10) using the symplectic condition, one has the relation

$$(M\ \delta z^i,\ J\ M\ dz^i) = (\delta z^i,\ J\ dz^i). \tag{4.43}$$

It follows that the two-dimensional integral based on the fundamental symplectic two-form is conserved,

$$I_2^f = I_2^i. \tag{4.44}$$

One final point is worth mentioning. In constructing the general higher order Poincare invariants, the calculus of differential forms is used to make general 2m-forms for $m = 2, 3, \ldots, n$ from the fundamental symplectic two-form. The invariance of all these forms, including the last of the hierarchy ($m=n$) which is just the volume element, follows from the invariance (4.10) of the fundamental symplectic two-form. This invariance is in turn equivalent to the symplectic condition (4.3). Thus, the symplectic condition is really the fundamental condition from which everything else follows.

Exercise 4.4: Suppose a "burst" of protons is injected into a uniform electric field $\vec{E} = E_o \hat{e}_z$. Assume the burst is initially concentrated at x and y = 0 and v_x and $v_y = 0$, but is uniformly spread in z and v_z about the values z = 0 and $v_z = v_z^o$ within intervals $\pm\Delta z$ and $\pm\Delta v_z$. Thus the problem is essentially that of one-dimensional motion along the z axis. The initial distribution is shown schematically in figure 4.3.

Find the distribution at later times, and verify Liouville's theorem. Do not assume Δz and Δv_z are infinitesimal. Neglect Coulomb interactions between particles.

5. LIE ALGEBRAS AND SYMPLECTIC MAPS

5.1 Lie transformations as symplectic maps

Let f(z,t) be any dynamical variable, and let exp(:f:) be the Lie transformation associated with f. This Lie transformation can be used to define a map $\mathcal{M}$ which produces new variables $\bar{z}(z,t)$ by the rule

$$\bar{z}_a(z,t) = \exp(:f:)\ z_a,\quad a = 1, 2, \ldots, 2n. \tag{5.1}$$

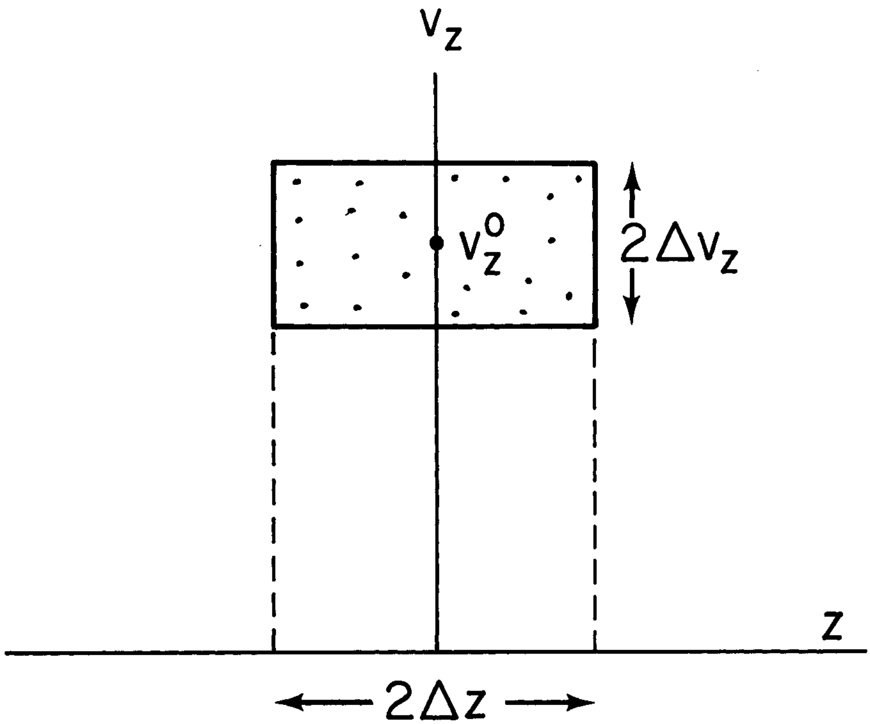


Figure (4.3). Initial phase-space distribution for exercise 4.4.

The relations (5.1) can also be expressed more compactly by writing

$$\bar{z} = \mathcal{M} z \tag{5.2}$$

with

$$\mathcal{M} = \exp(:f:). \tag{5.3}$$

Consider the Poisson brackets of the various $\bar{z}$'s with each other. Using the definition (5.1) and the ismorphism condition (3.37), one finds the result

$$\begin{aligned}[\bar{z}_a, \bar{z}_b] &= [\exp(:f:)\ z_a,\ \exp(:f:)\ z_b] \\ &= \exp(:f:)\ [z_a, z_b] \\ &= \exp(:f:)\ J_{ab} = J_{ab}.\end{aligned} \tag{5.4}$$

It follows from (5.4) that $\mathcal{M}$ is a symplectic map! What has been shown is that every Lie transformation may be viewed as a symplectic map.

Exercise 5.1: Verify Eq. (5.4).

The symplectic map defined by (5.3) has the particular property that f is an invariant function for the map. That is, one has the relation

$$f(\bar{z},t) = f(z,t). \tag{5.5}$$

To see the truth of this assertion, apply Eq. (3.35) to the case where $g = f$. One finds, using the notation of (5.1), the result

$$\exp(:f:)\ f(z,t) = f(\bar{z},t). \tag{5.6}$$

However, using the expression (3.33), one also obtains the result

$$\exp(:f:)\ f(z,t) = f(z,t) \tag{5.7}$$

since the Poisson bracket $[f,f]$ is zero by the antisymmetry condition. Comparison of (5.6) and (5.7) shows that Eq. (5.5) is indeed correct. Note that in all these calculations, the time t plays no essential role and may be regarded simply as a parameter.

Suppose the symplectic map exp(-:f:) is applied to both sides of (5.1). One finds the result

$$\exp(-:f:)\ \bar{z}_a = \exp(-:f:)\ \exp(:f:)\ z_a. \tag{5.8}$$

Consider first the problem of evaluating the right-hand side of (5.8). Observe that the Lie operators :f: and -:f: commute. Indeed, using Eq. (3.31), one finds the result

$$\{:f:, -:f:\} = :[f,-f]: = :0: = 0. \tag{5.9}$$

One might therefore imagine that the two operators in the product $\exp(-:f:)\exp(:f:)$ cancel each other to give a net identity operation. This is indeed the case. Consequently, when read from right to left, Eq. (5.8) may be rewritten in the form

$$z_a = \exp(-:f:)\, \bar{z}_a \,. \tag{5.10}$$

What has been shown is that if $\mathcal{M}$ is given by the Lie transformation relation (5.3) then $\mathcal{M}^{-1}$ is given by the relation

$$\mathcal{M}^{-1} = \exp(-:f:). \tag{5.11}$$

<u>Exercise 5.2</u>: Suppose f and g are in <u>involution</u>. That is,

$$[f,g] = 0. \tag{5.12}$$

Show from the power series definition (3.32) that in this case

$$\exp(:f:)\exp(:g:) = \exp(:f+g:)\,. \tag{5.13}$$

In attempting to evaluate (5.10) explicitly, the reader may be at somewhat of a loss as to exactly how to proceed since what appears to be called for is a set of Poisson brackets with respect to the variables z. However, in view of the invariance property (5.5), one may equally well compute all Poisson brackets with respect to the variables $\bar{z}$.

5.2 <u>Factorization theorem</u>

Note what has been accomplished so far. Section 2.4 showed that matrices of the form JS with S symmetric produce a Lie algebra. It also showed that any symplectic matrix sufficiently near the identity can be written in the form exp(JS). Similarly, section 3.3 showed that the set of Lie operators :f: forms a Lie algebra. And section 5.1 has just shown that Lie transformations exp(:f:) are symplectic maps. What remains to be studied is the question of whether any symplectic map $\mathcal{M}$ can be written in exponential form.

The answer to this question is given by the factorization theorem:

<u>Theorem 5.1</u>: Let $\mathcal{M}$ be an analytic symplectic mapping which maps the origin into itself. That is, the relation between the $\bar{z}$'s and z's is assumed to be expressible in a power series of the form

$$\bar{z}_a = \sum_b L_{ab} z_b + \text{higher order terms in } z. \tag{5.14}$$

Under these conditions the map $\mathcal{M}$ can be written as a product of Lie transformations,

$$\mathcal{M} = \exp(:f_2^c:)\exp(:f_2^a:)\exp(:f_3:)\exp(:f_4:)\ \ldots \tag{5.15}$$

Moreover, the functions f_m are homogeneous polynomials of degree m in the variable $z_1 \ldots z_{2n}$. Finally, the polynomials f_2^c and f_2^a are quadratic polynomials of the form

$$f_2^c = -(1/2) \sum_{ij} S_{ij}^c \, z_i z_j \qquad (5.16a)$$

$$f_2^a = -(1/2) \sum_{ij} S_{ij}^a \, z_i z_j \qquad (5.16b)$$

where S^c and S^a are real symmetric matrices which commute and anti-commute with J respectively.

Partial Proof: Let M(z) be the Jacobian matrix of the map $\mathcal{M}$. Then from Eq. (5.14), it is evident that

$$M(0) = L. \qquad (5.17)$$

Also, since M is symplectic for all values of z, see Eq. (4.3), it follows that L is a symplectic matrix. Consequently, using the results summarized in Eqs. (2.37) and (2.58), the matrix L can be written in the form

$$L = P\mathcal{O} = \exp(JS^a) \exp(JS^c). \qquad (5.18)$$

As indicated in Eq. (5.16a), let f_2^c be a quadratic polynomial defined in terms of the matrix S^c appearing in the decomposition (5.18). Now consider the Lie operator $:f_2^c:$. Suppose this Lie operator acts on the various z's. One finds the result

$$\begin{aligned} :f_2^c: z_k &= -(1/2) \sum_{ij} S_{ij}^c \, [z_i z_j, z_k] \\ &= -(1/2) \sum_{ij} S_{ij}^c \, \{[z_i, z_k] z_j + [z_j, z_k] z_i\} \\ &= -(1/2) \sum_{ij} S_{ij}^c \, \{J_{ik} \, z_j + J_{jk} \, z_i\} \\ &= \sum_i (JS^c)_{ki} \, z_i. \end{aligned} \qquad (5.19)$$

Here use has been made of the antisymmetry of J and the symmetry of S^c. Using matrix and vector notation, Eq. (5.19) can also be written in the more compact form

$$:f_2^c: z = (JS^c) \, z. \qquad (5.20)$$

From this form it is easy to see that one has the general relation

$$:f_2^c:^m \, z = (JS^c)^m \, z. \tag{5.21}$$

Finally, it follows from (5.21) that one also has the relation

$$\exp(:f_2^c:) \, z = \exp(JS^c) \, z = \mathcal{O} z. \tag{5.22}$$

Exercise 5.3: Verify Eqs. (5.19) through (5.22).

Exercise 5.4: Verify the analogous relation

$$\exp(:f_2^a:) \, z = \exp(JS^a) \, z = Pz. \tag{5.23}$$

By using matrix and vector notation, and the representation (5.18), Eq. (5.14) can be written in the more compact form

$$\bar{z} = P \, \mathcal{O} \, z + \text{higher order terms in } z. \tag{5.24}$$

Now apply the Lie transformation $\exp(-:f_2^c:)$ to both sides of (5.24). One finds the result

$$\exp(-:f_2^c:) \, \bar{z} = P \, \mathcal{O} \exp(-:f_2^c:) \, z + \text{higher order terms in } z. \tag{5.25}$$

However, it follows from (5.22) that one also has the relation

$$\exp(-:f_2^c:) \, z = \exp(-JS^c) \, z = \mathcal{O}^{-1} \, z. \tag{5.26}$$

Consequently, Eq. (5.25) can be rewritten in the form

$$\exp(-:f_2^c:) \, \bar{z} = P \, z + \text{higher order terms in } z. \tag{5.27}$$

Next apply the Lie transformation $\exp(-:f_2^a:)$ to both sides of (5.27). By making arguments analogous to those of the previous paragraph, one finds the result

$$\exp(-:f_2^a:) \exp(-:f_2^c:) \, \bar{z} = z + \text{higher order terms in } z. \tag{5.28}$$

Exercise 5.5: Verify Eqs. (5.24) through (5.28).

Suppose f_3 is some cubic polynomial in the z's. Then one finds the result

$$\exp(-:f_3:) \, z_b = z_b + \underbrace{:-f_3: \, z_b}_{\substack{\text{quadratic} \\ \text{terms}}} + (1/2!)\underbrace{:-f_3:^2 \, z_b}_{\substack{\text{cubic} \\ \text{terms}}} \ldots , \tag{5.29}$$

where the degrees of the various terms have been indicated. Consider now the effect of multiplying both sides of (5.28) by the Lie transformation $\exp(-:f_3:)$. Using (5.29), one finds the result

$$\exp(-:f_3:)\ \exp(-:f_2^a:)\ \exp(-:f_2^c:)\ \bar{z} = z$$
$$-\ :f_3:\ z + \text{higher order terms of } z. \qquad (5.30)$$

At this point, one might hope that by a suitable choice of f_3, all quadratic terms from the right-hand side of (5.30) could be eliminated. This can indeed be shown to be the case.[12] Consequently, there exists an f_3 (which is in fact unique) such that (5.30) takes the form

$$\exp(-:f_3:)\ \exp(-:f_2^a:)\ \exp(-:f_2^c:)\ \bar{z} = z$$
$$+ \text{ cubic and higher order terms in } z. \qquad (5.31)$$

Similarly, there exist unique higher order polynomials f_4, f_5, etc. which can be used successively to remove degree by degree each of the higher order terms on the right-hand side of (5.31). Consequently, one has the final relation

$$\exp(-:f_4:)\ \exp(-:f_3:)\ \exp(-:f_2^a:)\ \exp(-:f_2^c:)\ \bar{z} = z. \qquad (5.32)$$

Now multiply both sides of (5.32) by the $\mathcal{M}$ of Eq. (5.15). One finds the relation

$$\bar{z} = \mathcal{M}z \qquad (5.33)$$

which is the desired result.

Exercise 5.6: Suppose f_m and g_n are homogeneous polynomials of degrees m and n respectively. Show that $[f_m,g_n]$ is then also a homogeneous polynomial, and determine its degree.
Answer: Degree of $[f_m,g_n] = (m+n-2)$.

Exercise 5.7: Verify Eq. (5.29) and the assertion about the degree of the various terms. Verify the derivation of (5.33) from (5.32).

Theorem (5.1) is a key result. Recall that in theorem (4.1) it was shown that Hamiltonian flows produce symplectic maps. Now, thanks to theorem (5.1), it is possible to describe the most general analytic symplectic map (which sends the origin into itself) simply in terms of various homogeneous polynomials. Finally, it can be shown that the restriction of preserving the origin can be removed by including Lie transformations of the form $\exp(:f_1:)$ where f_1 is a suitably chosen polynomial linear in the z's. Consequently, any analytic symplectic map can be represented as a product of Lie transformations generated by polynomials.

At this point, two comments are appropriate. First, suppose the factored product representation (5.15) is truncated at any point. Then the resulting expression is still a symplectic map because each term in the product is a symplectic map. Also, if the truncation consists of dropping all terms in the product (5.15) beyond $\exp(:f_m:)$ for some m, then according to exercise 5.6 the power series expansion (5.14) for the truncated map agrees with that of the original map through terms of degree (m-1). Consequently, a truncated map provides a symplectic approximation to the exact map.

Second, suppose Eq. (5.15) is decomposed, as shown below, into those factors involving only quadratic polynomials, and the remaining factors involving cubic and higher degree polynomials.

$$\mathcal{M} = \overbrace{\exp(:f_2^c:)\exp(:f_2^a:)}^{\text{"Gaussian optics"}} \quad \overbrace{\underbrace{\exp(:f_3:)}_{\substack{\text{second order}\\ \text{effects or}\\ \text{aberrations}\\ \text{due to}\\ \text{sextupoles}}}\underbrace{\exp(:f_4:)}_{\substack{\text{third order effects}\\ \text{or aberrations due to}\\ \text{octupoles or iterated}\\ \text{sextupoles}}}\ \ldots}^{\text{Aberrations or nonlinear corrections}}$$

It will be demonstrated in subsequent sections that dropping all terms beyond those involving the quadratic polynomials leads to a lowest order approximation for $\mathcal{M}$ which is equivalent to the paraxial Gaussian optics approximation in the case of light optics, and the usual linear matrix approximation in the case of charged particle beam optics. Moreover, the remaining factors $\exp(:f_3:)$ $\exp(:f_4:)\ldots$represent aberrations or nonlinear corrections to the lowest order approximation. In particular, in the case of charged particle beam optics, the factor $\exp(:f_3:)$ describes various chromatic effects and the effects due to sextupoles. Similarly, the factor $\exp(:f_4:)$ describes higher order chromatic effects, the effects due to iterated sextupoles, and the effects due to octupoles. Finally in some cases, f_3, f_4, etc. also describe what may be called "kinematic" nonlinearities in the equations of motion. They arise, for example, from the fact that the Hamiltonians (1.22) and (1.23) are intrinsically nonlinear even in the absence of electric and magnetic fields.

5.3 Symplectic map for time independent Hamiltonian

Suppose the Hamiltonian of theorem (4.1) does not explicitly depend on the time. Then the symplectic map (4.17) obtained by following the Hamiltonian flow specified by H can be written immediately in the form

$$\mathcal{M} = \exp\{-(t^f - t^o) :H:\}. \tag{5.34}$$

To see the truth of (5.34), let $\mathcal{M}$ act on z to give the result

$$\bar{z} = z = \sum_{m=0}^{\infty} (1/m!)\,(t^f - t^o)^m :-H:^m z. \tag{5.35}$$

However, Taylor's theorem gives the result

$$\bar{z} = z(t^f) = z(t^o) + \sum_{m=1}^{\infty} (1/m!)(t^f-t^o)^m \ (d/dt)^m z(t)\Big|_{t^o} \ . \quad (5.36)$$

Also, Hamilton's equations of motion for the z's can be written in the form

$$\begin{aligned} \dot{z} &= [z,H] = [-H,z] = :-H:z \\ \ddot{z} &= [-H,\dot{z}] = :-H:\dot{z} = :-H:^2 z \\ (d^3 z)/(dt)^3 &= :-H:^3 z, \text{ etc.} \end{aligned} \quad (5.37)$$

Upon inserting the results of (5.37) into (5.36), one obtains the desired result (5.35).

Exercise 5.8: Evaluate $\mathcal{M}z$ with $\mathcal{M}$ given by (5.34) for the H of Eq. (4.29). See exercise (3.23).

Exercise 5.9: Suppose that the Hamiltonian H is not necessarily time independent, but does have the property that the Lie operators :H(z,t): for various times all commute. That is, one has the relation

$$\{:H(z,t):, :H(z,t'):\} = 0 \text{ for all } t,t'. \quad (5.38)$$

Show that in this case the symplectic map obtained by following the Hamiltonian flow specified by H can be written in the form

$$\mathcal{M} = \exp\left(-\int_{t^o}^{t^f} :H: dt\right). \quad (5.39)$$

5.4 A calculus for Lie transformations and noncommuting operators

A. Introductory Background

Section 4.3 showed that Hamiltonian flows produce symplectic maps, and section 5.2 showed that symplectic maps which send the origin into itself can be written in the factored product form (5.15). In addition, Eq. (5.34) gives an explicit representation of the symplectic map in the case of a time independent Hamiltonian. In subsequent sections these results will be applied to charged particle beam transport, light optics, and orbits in circular machines. The purpose of this section is to provide a collection of formulas for the manipulation of Lie transformations and noncommuting operators in general. These formulas have been derived elsewhere, and some will be stated without proof.[12] Some formulas will be used to compute the product of two symplectic maps when each is written in factored product form. Others will be used to combine various exponents in a factored product decomposition into a single exponent. Still others will be used to produce factored product decompositions.

Work with noncommuting quantities is often facilitated by the concept of an adjoint Lie operator. Let :f: be some Lie operator, and let :g: be any other Lie operator. The adjoint of the Lie operator :f:, which will be denoted by the symbol #:f:#, is a kind of super operator which acts on other Lie operators according to the rule[13]

$$\#{:}f{:}\#{:}g{:} = \{:f:,:g:\}. \tag{5.40}$$

Here, the right-hand side of (5.40) denotes the commutator as in Eq. (3.27). Powers of #:f:# can also be defined by repeated applications of (5.40). For example, $\#{:}f{:}\#^2$ is defined by the relation

$$\#{:}f{:}\#^2 \;{:}g{:} = \{:f:,\{:f:,:g:\}\}. \tag{5.41}$$

Finally, #:f:# to the zero power is defined to be the identity operator,

$$\#{:}f{:}\#^0 \;{:}g{:} = {:}g{:} \;. \tag{5.42}$$

Exercise 5.10: As usual, let the symbols { , } denote the operation of commutation,

$$\{\#{:}f{:}\#,\#{:}g{:}\#\} = \#{:}f{:}\#\#{:}g{:}\# - \#{:}g{:}\#\#{:}f{:}\#. \tag{5.43}$$

As a test of your grasp of what is going on, prove the relations

$$\{\#{:}f{:}\#,\#{:}g{:}\#\} = \#\{:f:,:g:\}\# = \#{:}[f,g]{:}\#. \tag{5.44}$$

Hint: Use the Jacobi condition for commutators.

To simplify notation in some cases where no confusion can arise, the set of colons in the symbols #:f:# for the adjoint of the Lie operator :f: will be omitted. That is, the abbreviated symbol #f# will be used to serve for the complete symbol #:f:#.

Exercise 5.11: Prove the relation

$$\#f\#^n{:}g{:} = {:}({:}f{:}^n g){:} \;. \tag{5.45}$$

Exercise 5.12: Prove the relation

$$\#f\#({:}g{:}{:}h{:}) = (\#f\#{:}g{:}){:}h{:} + {:}g{:}(\#f\#{:}h{:}) \;. \tag{5.46}$$

That is, #f# is also a derivation with respect to the multiplication of Lie operators.

Armed with this new notation, it is now possible to state some useful results. Suppose :f: and :g: are two Lie operators. Then, it can be shown that

$$\exp(:f:)\ :g:\ \exp(-:f:) = \exp(\#f\#)\ :g: \ . \tag{5.47}$$

Here, as the notation suggests,

$$\exp(\#f\#) = \sum_{m=0}^{\infty} \#f\#^m/m! \quad . \tag{5.48}$$

Exercise 5.13: Verify (5.47) term by term for at least the first few terms by comparing power series expansions.

Now, using (5.47), it follows that

$$\exp(\#f\#)\ :g: = :\exp(:f:)g: \quad . \tag{5.49}$$

Consequently, Eq. (5.47) can also be written in the form

$$\exp(:f:)\ :g:\ \exp(-:f:) = :\exp(:f:)g: \quad . \tag{5.50}$$

Next observe that because #f# is a derivation, see Eq. (5.46), it follows from (5.50) that a similar relation holds for powers of :g:,

$$\exp(:f:):g:^m\ \exp(-:f:) = :\exp(:f:)g:^m \quad . \tag{5.51}$$

Moreover, Eq. (5.51) in turn implies the relation

$$\exp(:f:)\ \exp(:g:)\exp(-:f:) = \exp(:\exp(:f:)g:). \tag{5.52}$$

Exercise 5.14: Verify Eqs. (5.49) through (5.52) starting with Eq. (5.47).

A second useful result is the analog of formulas (2.34) and (2.35) for Lie operators. Suppose :f: and :g: are any two Lie operators. Then one has the Campbell-Baker-Hausdorff formula

$$\begin{aligned}\exp(s:f:)\ \exp(t:g:) &= \exp(s:f: + t:g: \\ &+ (st/2)\{:f:,:g:\} + (s^2t/12)\{:f:,\{:f:,:g:\}\} \\ &+ (st^2/12)\{:g:,\{:g:,:f:\}\} + \ldots).\end{aligned} \tag{5.53}$$

Moreover, using Eqs. (3.31) and (5.45), Eq. (5.53) can also be written in the form

$$\exp(s:f:)\ \exp(t:g:) = \exp(:h:) \tag{5.54a}$$

with

$$\begin{aligned}h &= sf + tg + (st/2)[f,g] \\ &+ (s^2t/12):f:^2g + (st^2/12):g:^2f + \ldots \ .\end{aligned} \tag{5.54b}$$

B. Concatenation Formulas

As an application of the formulas developed so far, consider the problem of computing the product of two symplectic maps when each is expressed in factored product form. This problem arises, for example, in the case that one knows the effect of two beam elements separately, and one wants to know the net effect when one beam element is followed by another.

Let $\mathcal{M}^f$ and $\mathcal{M}^g$ denote the symplectic maps given by the expressions

$$\mathcal{M}^f = \exp(:f_2:)\ \exp(:f_3:)\ \exp(:f_4:), \tag{5.55a}$$

$$\mathcal{M}^g = \exp(:g_2:)\ \exp(:g_3:)\ \exp(:g_4:). \tag{5.55b}$$

Also, let $\mathcal{M}^h$ be the product of $\mathcal{M}^f$ and $\mathcal{M}^g$,

$$\mathcal{M}^h = \mathcal{M}^f\ \mathcal{M}^g. \tag{5.56}$$

The problem is to find polynomials h_2, h_3, etc., such that

$$\mathcal{M}^h = \exp(:h_2:)\ \exp(:h_3:)\ \exp(:h_4:)\ \ldots\ . \tag{5.57}$$

That is, the problem is to express $\mathcal{M}^h$ as given by (5.56) in the factored product form (5.57). For simplicity, only expressions for h_2, h_3, and h_4 will be found explicitly.

Before proceeding further, it is necessary to establish a few simple facts.

Exercise 5.15: Suppose g_2 is a quadratic polynomial written in the form

$$g_2 = -(1/2) \sum_{ij} S_{ij}\ z_i z_j\ , \tag{5.58}$$

where S is some symmetric matrix. Suppose further that f_m is some homogeneous polynomial of degree m. Show that then $\exp(:g_2:)\ f_m$ is also a homogeneous polynomial of degree m.

Solution:

$$\exp(:g_2:)\ f_m(z) = f_m(M^g\ z) \tag{5.59}$$

where M^g is the linear transformation defined by the equation

$$M^g = \exp(JS). \tag{5.60}$$

See Eqs. (3.35), (5.16), and (5.22).

Simply from its definition, $\mathcal{M}^h$ can be written in the form

$$\mathcal{M}^h = \exp(:f_2:)\,\exp(:f_3:)\,\exp(:f_4:)\times$$
$$\exp(:g_2:)\,\exp(:g_3:)\,\exp(:g_4:). \qquad (5.61)$$

Next, by suitable insertions of various Lie transformations and their inverses, Eq. (5.61) can be rewritten in the form

$$\mathcal{M}^h = \exp(:f_2:)\,\exp(:g_2:)\,\exp(-:g_2:)\,\exp(:f_3:)\,\exp(:g_2:)\times$$
$$\exp(-:g_2:)\,\exp(:f_4:)\,\exp(:g_2:)\,\exp(:g_3:)\,\exp(:g_4:). \qquad (5.62)$$

Exercise 5.16: Verify Eq. (5.62) starting from (5.61).

Evidently, comparison of (5.62) and (5.57) shows that h_2 is determined by the equation

$$\exp(:h_2:) = \exp(:f_2:)\,\exp(:g_2:)\quad . \qquad (5.63)$$

Indeed, using the notation of Eq. (5.61) of exercise (5.15), one has the result[14]

$$M^h = M^g\,M^f. \qquad (5.64)$$

Exercise 5.17: Verify Eq. (5.64).

Next observe that Eq. (5.63) contains the factors $\exp(-:g_2:)$ $\exp(:f_3:)$ $\exp(:g_2:)$. Here is where Eq. (5.52) comes into play. Employing Eq. (5.52), one finds the result

$$\exp(-:g_2:)\,\exp(:f_3:)\,\exp(:g_2:) = \exp(:\exp(-:g_2:)f_3:). \qquad (5.65)$$

Also, according to exercise (5.15), one has the relation

$$\exp(-:g_2:)\,f_3(z) = f_3\,[\,(M^g)^{-1}\,z\,] \qquad (5.66)$$

where $(M^g)^{-1}$ is the matrix

$$(M^g)^{-1} = \exp(-JS) = -J\,\tilde{M}^g\,J. \qquad (5.67)$$

Here use has also been made of (2.8).

Exercise 5.18: Verify Eqs. (5.65) through (5.67).

In order to simplify further expressions, introduce the notation

$$f_m^t(z) = f_m[(M^g)^{-1} z] \qquad (5.68)$$

which indicates that the homogeneous polynomial $f_m(z)$ of degree m has been transformed to the new homogeneous polynomial $f_m[(M^g)^{-1} z]$. With this notation, Eq. (5.65) can be written in the more compact form

$$\exp(-:g_2:)\ \exp(:f_3:)\ \exp(:g_2:) = \exp(:f_3^t:) \quad . \qquad (5.69)$$

Similarly, one finds for the factor $\exp(-:g_2:)\ \exp(:f_4:)\ \exp(:g_2:)$, which also occurs in (5.62), the result

$$\exp(-:g_2:)\ \exp(:f_4:)\ \exp(:g_2:) = \exp(:f_4^t:). \qquad (5.70)$$

Exercise 5.19: Verify Eq. (5.70).

Putting together the work done so far, one finds that Eq. (5.62) can also be written in the form

$$\mathcal{M}^h = \exp(:h_2:)\ \exp(:f_3^t:)\ \exp(:f_4^t:)\ \exp(:g_3:)\ \exp(:g_4:). \qquad (5.71)$$

Again, by a suitable insertion of a Lie transformation and its inverse, Eq. (5.71) can be written in the form

$$\mathcal{M}^h = \exp(:h_2:)\ \exp(:f_3^t:)\ \exp(:g_3:)\ \exp(-:g_3:)\times$$
$$\exp(:f_4^t:)\ \exp(:g_3:)\ \exp(:g_4:). \qquad (5.72)$$

Now consider the factor $\exp(-:g_3:)\ \exp(:f_4^t:)\ \exp(:g_3:)$. Again using Eq. (5.52), this factor can be written in the form

$$\exp(-:g_3:)\ \exp(:f_4^t:)\ \exp(:g_3:) = \exp(:\exp(-:g_3:)f_4^t:). \qquad (5.73)$$

However, using the results of exercise (5.6), one finds the relation

$$\exp(-:g_3:)\ f_4^t = f_4^t + \text{polynomials of degree 5 and greater.} \qquad (5.74)$$

Therefore, if one is not interested in computing the h_m of degree 5 and higher, Eq. (5.72) can also be written in the form

$$\mathcal{M}^h = \exp(:h_2:)\ \exp(:f_3^t:)\ \exp(:g_3:)\ \exp(:f_4^t:)\ \exp(:g_4):\cdots . \qquad (5.75)$$

At this point Eqs. (5.53) and (5.54) can be brought into play. Using these equations, one finds for the product $\exp(:f_3^t:)\ \exp(:g_3:)$ occurring in (5.75) the result

$$\exp(:f_3^t:)\ \exp(:g_3:) = \exp(:f_3^t: + g_3: + (1/2)\ :[f_3^t,g_3]: + \cdots). \tag{5.76}$$

Note that $[f_3^t,g_3]$ is a homogeneous polynomial of degree 4, and that the remaining terms not shown on the right-hand side of (5.76) are polynomials of still higher degree. Consequently, Eq. (5.76) can also be written in the form

$$\exp(:f_3^t:)\ \exp(:g_3:) = \exp(:f_3^t: + :g_3:)\ \exp(:[f_3^t,g_3]/2:\ \cdots). \tag{5.77}$$

Finally, again using (5.53) and (5.54), it is clear that Eq. (5.72) for $\mathcal{M}^h$ can be written in the form

$$\mathcal{M}^h = \exp(:h_2:)\ \exp(:f_3^t: + :g_3:)\ \exp(:[f_3^t,g_3]/2: + :f_4^t: + :g_4:)\ \ldots\ . \tag{5.78}$$

Comparison of Eqs. (5.57) and (5.78) shows that the polynomials h_3 and h_4 are given by the expressions

$$h_3 = f_3^t + g_3 \tag{5.79}$$

$$h_4 = [f_3^t,g_3]/2 + f_4^t + g_4. \tag{5.80}$$

With further work, it is possible to find the polynomials h_5, h_6, etc.

C. Formulas for Combining Exponents

Sometimes, as will be shown later, it is useful to be able to write the product of two Lie transformations as a single Lie transformation. This is what the Campbell-Baker-Hausdorff formula (5.53) attempts to do. In general, there are no known convenient expressions for all the terms on the right-hand side of (5.54b). However, it is possible to sum the series completely with respect to s and the first few powers in t.[12] One such result can be written in the form

$$h = sf + s:f:[1 - \exp(-s:f:)]^{-1}\ (tg) + O(t^2). \tag{5.81}$$

Here the operator expression involving :f: is to be interpreted as the infinite series

$$\begin{aligned} s:f:[1 - \exp(-s:f:)]^{-1} &= s:f:[1 - \sum_{m=0}^{\infty} (-s:f:)^m/m!]^{-1} \\ &= s:f:[-\sum_{m=1}^{\infty} (-s:f:)^m/m!]^{-1} \qquad (5.82) \\ &= 1 + (s/2):f: + (s^2/12):f:^2 + \cdots \end{aligned}$$

Equations (5.54a) and (5.81) may be combined to give the result

$$\exp(s{:}f{:})\exp(t{:}g{:})=\exp[s{:}f{:}+{:}\{s{:}f{:}[1-\exp(-s{:}f{:})]^{-1}(tg)\}{:}+O(t^2)]. \tag{5.83}$$

D. Factorization Formulas

Equation (5.83) gives a formula for combining two exponentials into one grand exponential. Sometimes, as in, for example the construction of a factored product decomposition, it is useful to be able to turn the process around. Define a quantity h by writing

$$s{:}f{:}[1 - \exp(-s{:}f{:})]^{-1}\, g = h. \tag{5.84}$$

Observe that Eq. (5.84) may be solved for the quantity g to give the relation

$$g = \{[1 - \exp(-s{:}f{:})]/[s{:}f{:}]\}\, h. \tag{5.85}$$

Here the operator expression appearing on the right of (5.85) is interpreted to be the series

$$[1 - \exp(-s{:}f{:})]/[s{:}f{:}] = -\sum_{m=1}^{\infty} (-s)^m{:}f{:}^m/[m!\,s{:}f{:}]$$

$$= -\sum_{m=1}^{\infty} (-s)^{m-1}{:}f{:}^{m-1}/m!. \tag{5.86}$$

Now insert Eq. (5.84) into Eq. (5.83). One finds, upon reading right to left, the result

$$\exp[s{:}f{:} + t{:}h{:} + O(t^2)] = \exp(s{:}f{:})\exp(t{:}g{:}). \tag{5.87}$$

Finally, the term of $O(t^2)$ can be taken from the left to the right-hand side of (5.87) to produce the relation

$$\exp[s{:}f{:} + t{:}h{:}] = \exp(s{:}f{:})\exp(t{:}g{:})\exp[{:}O(t^2){:}]. \tag{5.88}$$

Equation (5.88) gives a formula for writing the sum of two exponentials as a product of two exponentials.

It is worth remarking, in closing, that the operation described by Eq. (5.85), which is required for evaluating (5.88), can be written in a more compact form. First, observe the formal integral identity

$$[1 - \exp(-s{:}f{:})]/[s{:}f{:}] = \int_0^1 d\tau\, \exp(-\tau s{:}f{:}). \tag{5.89}$$

Using this identity, Eq. (5.85) can be written in the form

$$g = \int_0^1 d\tau \exp(-\tau s{:}f{:})\, h. \tag{5.90}$$

But now Eq. (3.35) can be employed to give the final integral formula

$$g(2) = \int_0^1 d\tau \;\; h \exp(-\tau s{:}f{:}) z \; . \tag{5.91}$$

Exercise 5.20: Verify the expansions (5.82) and (5.86).

Exercise 5.21: Verify the integral identity (5.89).

6. APPLICATIONS TO CHARGED PARTICLE BEAM TRANSPORT AND LIGHT OPTICS

Section 4.3 showed that Hamiltonian flows produce symplectic maps, and section 5.2 showed that symplectic maps which send the origin into itself can be written in the form

$$\mathcal{M} = \exp({:}f_2^c{:}) \exp({:}f_2^a{:}) \exp({:}f_3{:}) \cdots . \tag{6.1}$$

Thus, the result of a Hamiltonian flow which sends the origin into itself can be completely described by a set of polynomials f_2, f_3, etc. The purpose of this section is to give examples of the computation of the first few polynomials for quadrupole, sextupole, dipole, and drift beam elements. These calculations serve as the beginning of a complete catalog of Lie operators for the commonly used beam elements. Further results may be found elsewhere, and additional work is still in progress.[15] A brief discussion will also be given of Lie operators corresponding to various simple optical elements in the case of light optics.

6.1 The perfect quadrupole

Consider the case of a perfect magnetic quadrupole of length ℓ. By "perfect" it is meant that the magnetic field in the body of the magnet is assumed to have only a quadrupole component, and end effects and fringe fields are neglected. Then, employing the coordinates shown in Fig. (1.1) and utilizing the results of exercises 1.5 and 1.10, one finds for the Hamiltonian K the result

$$K = -[(p_t^2/c^2) - m_o^2 c^2 - p_y^2 - p_z^2]^{1/2} - (qa_2/2)(z^2 - y^2). \tag{6.2}$$

The design orbit for a quadrupole passes through the center of the quadrupole, and has a certain design energy. Consequently, the design orbit can be characterized by writing the equations

$$p_y = 0 \quad , \; y = 0 \qquad (6.3a,b)$$

$$p_z = 0 \quad , \; z = 0 \qquad (6.4a,b)$$

$$p_t = p_t^o \quad , \; t(x) = x(v_x^o)^{-1} + \text{constant.} \qquad (6.5a,b)$$

The meaning of Eqs. (6.3) and (6.4) should be evident. Equations (6.5) may require some explanation. Applying Hamilton's equations, one has the relations

$$p_t' = dp_t/dx = -\partial K/\partial t = 0, \qquad (6.6a)$$

$$t' = dt/dx = \partial K/\partial p_t = -(p_t/c^2)[(p_t^2/c^2) - m_o^2c^2 - p_y^2 - p_z^2]^{-1/2}. \qquad (6.6b)$$

It follows from Eq. (6.6a) that p_t has a constant value. Hence, Eq. (6.5a) is correct. Inserting this constant value p_t^o (corresponding to a particular value for the design energy) and Eqs. (6.3) and (6.4) into Eq. (6.6b) gives the result that on the design orbit

$$t' = dt/dx = (v_x^o)^{-1}, \qquad (6.7a)$$

where v_x^o is the velocity of the design orbit given by the relation

$$v_x^o = [(p_t^o/c)^2 - m_o^2c^2]^{1/2}[-p_t^o/c^2]^{-1} \quad . \qquad (6.7b)$$

Integration of (6.7a) gives the advertized result (6.5b).

According to Eq. (6.5), following the Hamiltonian flow generated by K along the design orbit does not lead to a mapping of the origin of phase space into itself since the time along the orbit changes, and p_t usually differs from zero. This situation can be remedied by working with a different set of canonical variables. Define "new" variables T, Y, Z, P_T, P_Y, P_Z by the relations

$$t = T + x(v_x^o)^{-1} \qquad (6.8a)$$

$$y = Y \qquad (6.8b)$$

$$z = Z \qquad (6.8c)$$

$$p_t = p_t^o + P_T \qquad (6.8d)$$

$$p_y = P_Y \qquad (6.8e)$$

$$p_z = P_Z . \qquad (6.8f)$$

It is readily verified that this change of variables is a canonical transformation and arises from the transformation function

$$F_2(q,P) = yP_Y + zP_Z + [t - x(v_x^o)^{-1}][P_T + p_t^o], \tag{6.9}$$

where v_x^o and p_t^o are to be regarded as constant numbers. For example, one finds the relations[16]

$$T = \partial F_2/\partial P_T = t - x(v_x^o)^{-1} \tag{6.10a}$$

$$p_t = \partial F_2/\partial t = P_T + p_t^o \tag{6.10b}$$

which are in agreement with Eqs. (6.8a) and (6.8d).

In terms of these new variables, it is evident that the design orbit can be taken to be given by the equations

$$T = X = Y = P_T = P_X = P_Y = 0. \tag{6.11}$$

Note that the variable T measures deviations in transit time from that of the design orbit, and its conjugate momentum P_T measures deviations in p_t from the design orbit value p_t^o.

Let K be regarded as the "old" Hamiltonian, and let H denote the "new" Hamiltonian for the new variables. [Note that in this context H is not to be confused with the Hamiltonians of exercises (1.3) or (1.4).] Then one has the relation[16]

$$H = K + \partial F_2/\partial x \tag{6.12}$$

since in the current procedure x plays the role of the independent "time-like" variable. Carrying out the prescription (6.12), one finds the result

$$H = -[(p_t^o + P_T)^2/c^2 - m_o^2c^2 - p_y^2 - p_z^2]^{1/2} - (qa_2/2)(z^2 - y^2)$$
$$-(v_x^o)^{-1}(P_T + p_t^o). \tag{6.13}$$

Here the quantities Y, Z, P_Y, P_Z have been replaced by their lower case counterparts in view of Eqs. (6.8b), (6.8c), (6.8e), and (6.8f) in order to simplify notation. From Eqs. (6.11) it follows that the mapping $\mathcal{M}$ produced by following the Hamiltonian flow generated by H does indeed send the origin of phase space into itself. Moreover, according to section (5.3), the mapping $\mathcal{M}$ is given explicitly by the expression

$$\mathcal{M} = \exp(-\ell :H:), \tag{6.14}$$

where, as assumed earlier, ℓ is the length of the quadrupole.

The problem at this point is to take the M given by (6.14) and re-express it in the factored product form (6.1). The first step in solving this problem is to expand H in a power series about the design orbit to give an expression of the form

$$H = H_0 + H_1 + H_2 + H_3 + \cdots \tag{6.15}$$

where H_n denotes a homogeneous polynomial of order n. One finds for the first few polynomials the results

$$H_0 = m_o^2\, c^2/p^o \tag{6.16a}$$

$$H_1 = 0 \tag{6.16b}$$

$$H_2 = (P_T^2/2)(m_o^2/p^{o3}) + (p_y^2 + p_z^2)/(2p^o) - (qa_2/2)(z^2 - y^2) \tag{6.16c}$$

$$H_3 = [P_T^3/2][(m_o^2)/(p^{o4}v_x^o)] + [P_T(p_y^2 + p_z^2)/(2p^{o2}v_x^o)] \ . \tag{6.16d}$$

Here p^o denotes the magnitude of the design mechanical momentum, and is related to p_t^o by the equation

$$p_t^o = -[m_o^2c^4 + c^2p^{o2}]^{1/2} \ . \tag{6.17}$$

See exercise (1.13).

<u>Exercise 6.1</u>: Verify Eqs. (6.16). Note that the vanishing of H_1 is equivalent to the condition that M send the origin of phase space into itself.

Next insert the expansion (6.15) into the expression (6.14) for M and imagine that the result is written in factored product form. Doing so gives the relation

$$\begin{aligned} M &= \exp\{-\ell(:H_2: \ + :H_3: + \cdots)\} \\ &= \exp(:f_2:)\exp(:f_3:) \cdots \ . \end{aligned} \tag{6.18}$$

Note that as expected the constant part of the Hamiltonian plays no role since $:H_o: = 0$.

The calculation of f_2 and f_3 can now be carried out with the aid of Eqs. (5.88) and (5.90). One finds the results

$$f_2 = -\ell H_2 \tag{6.19}$$

$$f_3 = -\ell \int_o^1 d\tau \exp(\tau\ell :H_2:)H_3 \ . \tag{6.20}$$

Note that Eq. (5.88) involves uncalculated terms of order t^2 where t is an expansion parameter. In the present context, these terms are those in the expansion which involve multiple Poisson brackets containing any number of H_2's and two H_3's. However, according to exercise (5.6), these terms are all polynomials of order 4. Therefore, they contribute to f_4, but not to f_2 or f_3. Similarly, the terms H_4 and higher in H do not contribute to f_2 and f_3.

Exercise 6.2: Verify Eqs. (6.19) and (6.20) and the assertions just made.

To evaluate f_3 as given by Eq. (6.20), and to gain additional insight, it is useful to study in detail the first factor in M given by $\exp(:f_2:) = \exp(-\ell :H_2:)$. Observe that H_2 can be written as the sum of three terms which are mutually in involution (have vanishing Poisson brackets),

$$H_2 = H_2^T + H_2^y + H_2^z \tag{6.21}$$

where

$$H_2^T = (P_T^2/2)(m_o^2/p^{o^3}) \tag{6.22a}$$

$$H_2^y = p_y^2/(2p^o) + (qa_2/2)y^2 \tag{6.22b}$$

$$H_2^z = p_z^2/(2p^o) - (qa_2/2)z^2 \tag{6.22c}$$

Consequently, using exercise 5.2, the quantity of interest can also be written in the form

$$\exp(-\ell :H_2:) = \exp(-\ell :H_2^T:)\exp(-\ell :H_2^y:)\exp(-\ell :H_2^z:). \tag{6.23}$$

Each of the three factors appearing in (6.23) can now be evaluated individually.

Consider first the factor $\exp(-\ell :H_2^T:)$. Following exercise (3.19), it is easily verified that $\exp(-\ell :H_2^T:)$ leaves the variables y, p_y, z, p_z in peace, and transforms the variables P_T and T according to the rules

$$\exp(-\ell :H_2^T:)T = T + (\ell m_o^2/p^{o^3})P_T \tag{6.24a}$$

$$\exp(-\ell :H_2^T:)P_T = P_T \quad . \tag{6.24b}$$

The above relations can be written more compactly in the form

$$\exp(-\ell:H_2^T:)\begin{pmatrix} T \\ P_T \end{pmatrix} = \begin{pmatrix} 1 & \ell m_o^2/P^{o^3} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} T \\ P_T \end{pmatrix}. \quad (6.25)$$

Physically, they describe the dependence of transit time on particle energy.

Next consider the factor $\exp(-\ell:H^y:)$. It affects only the coordinates y and p_y. Furthermore, the action of powers of $(-\ell:H_2^y:)$ on y and p_y is easily evaluated. One finds the result

$$(-\ell:H_2^y:)y = -\ell(2p^o)^{-1}[p_y^2,y] = (\ell/p^o)p_y \quad (6.26a)$$

$$(-\ell:H_2^y:)p_y = -(\ell qa_2/2)[y^2,p_y] = -q\ell a_2 y \quad (6.26b)$$

$$\begin{aligned}(-\ell:H_2^y:)^2 y &= (-\ell:H_2^y:)(-\ell:H_2^y:)y \\ &= -(\ell:H_2^y:)(\ell/p^o)p_y = -(q\ell^2 a_2/p^o)y \end{aligned} \quad (6.26c)$$

$$\begin{aligned}(-\ell:H_2^y:)^2 p_y &= (-\ell:H_2^y:)(-\ell:H_2^y:)p_y \\ &= (-\ell:H_2^y:)(-q\ell a_2 y) = -(q\ell^2 a_2/p^o)p_y, \text{ etc.}\end{aligned} \quad (6.26d)$$

Thus, as in exercise 3.23, the infinite series describing the effect of $\exp(-\ell:H_2^y:)$ on y and p_y can be summed. Writing the result in matrix form as before, one finds the relation

$$\exp(-\ell:H_2^y:)\begin{pmatrix} y \\ p_y \end{pmatrix} = \begin{pmatrix} \cos k\ell & (kp^o)^{-1}\sin k\ell \\ -kp^o \sin k\ell & \cos k\ell \end{pmatrix}\begin{pmatrix} y \\ p_y \end{pmatrix}, \quad (6.27)$$

where k denotes the quantity

$$k = (qa_2/p^o)^{1/2} . \quad (6.28)$$

Observe that Eq. (6.27) gives just the result expected in y and p_y space for the effect of a horizontally focussing quadrupole described in the linear matrix approximation.[17]

Exercise 6.3: Verify Eq. (6.27). Summation of the series is facilitated by the identity

$$\exp(w) = \cosh w + \sinh w$$

which decomposes the exponential series into even and odd parts.

Finally, consider the factor $\exp(-\ell :H_2^z:)$. It affects only z and p_z. According to Eqs. (6.22b) and (6.22c), the forms of H_2^y and H_2^z differ only in the sign of the quantity a_2. Consequently, the effect of $\exp(-\ell:H_2^z:)$ on z and p_z can be inferred by analogy with Eq. (6.27). One finds the result

$$\exp(-\ell:H_2^z:)\begin{pmatrix} z \\ \\ p_z \end{pmatrix} = \begin{pmatrix} \cosh k\ell & (kp^o)^{-1} \sinh k\ell \\ \\ kp^o \sinh k\ell & \cosh k\ell \end{pmatrix} \begin{pmatrix} z \\ \\ p_z \end{pmatrix} . \quad (6.29)$$

Note that Eq. (6.29) gives just the vertical defocussing action expected in linear matrix approximation for the quadrupole in question.[17]

Exercise 6.4: Verify Eq. (6.29).
Hint: Just write the expression analogous to (6.27) and make the analytic continuation $k \to ik$ (or, if you wish, $k \to -ik$).

So far, it has been shown that the factor $\exp(:f_2:)$ reproduces the usual linear matrix approximation to charged particle beam optics. The next term to compute is f_3.

Observe that H_3 as given by (6.16d) can be written as a sum of terms in the form

$$H_3 = H_3^T + H_3^{Ty} + H_3^{Tz} \quad (6.30)$$

where

$$H_3^T = [P_T^3/2][(m_o^2)/(p^{o^4} v_x^o)] \quad (6.31a)$$

$$H_3^{Ty} = P_T\, p_y^2/(2p^{o^2} v_x^o) \quad (6.31b)$$

$$H_3^{Tz} = P_T\, p_z^2/(2p^{o^2} v_x^o). \quad (6.31c)$$

Moreover, using the fact that some of the terms in (6.31) are in involution with some of the terms in (6.22), it can easily be verified that the quantity exp $-\ell(:H_2: + :H_3)$ can be written in the form

$$\exp\{-\ell(:H_2: + :H_3:)\} =$$

$$\exp(-\ell:H_2^T:)\exp(-\ell:H_3^T:)\exp\{-\ell(:H_2^y:+:H_3^{Ty}:)\} \times$$

$$\exp\{-\ell(:H_2^z:+:H_3^{Tz}:)\}. \tag{6.32}$$

This facilitates the computation of f_3 since the only terms that need to be factored now are $\exp\{-\ell(:H_2^y:+:H_3^{Ty}:)\}$ and its z counterpart.

<u>Exercise 6.5</u>: Verify Eq. (6.32).

Suppose the factorization of $\exp\{-\ell(:H_2^y:+:H_3^{Ty}:)\}$ is written in the form

$$\exp\{-\ell(:H_2^y: + :H_3^{Ty}:)\} = \exp(:f_2^y:)\exp(:f_3^{Ty}:)\exp(:f_4:)\ldots\ . \tag{6.33}$$

Then it follows from (5.88) and (5.90) that f_2^y and f_3^{Ty} are given by the expressions

$$f_2^y = -\ell H_2^y \tag{6.34}$$

$$f_3^{Ty} = -\ell \int_0^1 d\tau\ \exp(\tau\ell:H_2^y:)\ H_3^{Ty}\ . \tag{6.35}$$

Moreover, there is the simplification

$$\exp(\tau\ell:H_2^y:)H_3^{Ty} = P_T/(2p^{o^2} v_x^o)\{\exp(\tau\ell:H_2^y:)p_y\}^2\ . \tag{6.36}$$

Here use has been made of the isomorphism property (3.34) and (3.35). It is therefore only necessary to evaluate the quantity $\exp(\tau\ell:H_2^y:)p_y$. This can be done easily by reference to (6.27). One finds the result

$$\exp(\tau\ell:H_2^y:)p_y = ykp^o\ \sin(k\tau\ell) + p_y\ \cos(k\tau\ell). \tag{6.37}$$

Consequently, f_3^{Ty} is given by the relation

$$f_3^{Ty} = -\ell P_T/(2p^{o^2} v_x^o) \int_0^1 d\tau\{ykp^o\ \sin(k\tau\ell) + p_y\ \cos(k\tau\ell)\}^2. \tag{6.38}$$

The integration can easily be carried out to give the explicit result

$$f_3^{Ty} = -\ell P_T/(4p^{o^2} v_x^o)\{p_y^2[1 + (2k\ell)^{-1} \sin(2k\ell)]$$
$$+ y^2(kp^o)^2[1 -(2k\ell)^{-1} \sin(2k\ell)]$$
$$+ yp_y(p^o/\ell)[1 - \cos 2k\ell]\} . \qquad (6.39)$$

Exercise 6.6: Verify Eqs. (6.37) through (6.39).

In a similar fashion, the factor $\exp\{-\ell(:H_2^z: + :H_3^{Tz}:)\}$ can be handled by writing

$$\exp\{-\ell(:H_2^z: + :H_3^{Tz}:)\} = \exp(:f_2^z:)\exp(:f_3^{Tz}:)\exp(:f_4:)\cdots \quad . (6.40)$$

As observed before, H_2^y and H_2^z are of the same form except for the sign of a_2. Also, inspection of Eqs. (6.31b) and (6.31c) show that H_3^{Ty} and H_3^{Tz} are of the same form. Consequently, the expression for f_3^{Tz} can be inferred by analogy from that for f_3^{Ty} as given by (6.39). One finds the result

$$f_3^{Tz} = -\ell P_T/(4p^{o^2} v_x^o)\{p_y^2[1 + (2k\ell)^{-1} \sinh(2k\ell)]$$
$$- y^2(kp^o)^2[1 - (2k\ell)^{-1} \sinh(2k\ell)]$$
$$+ yp_y(p^o/\ell)[1 - \cosh(2k\ell)]\}. \qquad (6.41)$$

Exercise 6.7: Verify (6.41) using the method of exercise (6.4).

Let us now summarize what has been accomplished. Upon combining Eqs. (6.18), (6.32), (6.33), and (6.40), one finds for M the result

$$M = \exp(-\ell:H_2^T:)\exp(-\ell:H_3^T:)\exp(:f_2^y:)\exp(:f_3^{Ty}:)\times$$
$$\exp(:f_2^z:)\exp(:f_3^{Tz}:). \qquad (6.42a)$$

It is easily verified that the various factors in (6.42a) can be rearranged to give the result

$$M = \exp(-\ell:H_2:)\exp\{:(-\ell H_3^T + f_3^{Ty} + f_3^{Tz}):\}. \quad (6.42b)$$

Consequently, the quantity f_3 appearing in (6.18) is given by the relation

$$f_3 = -\ell H_3^T + f_3^{Ty} + f_3^{Tz}. \qquad (6.43)$$

This completes the calculation of f_3.

<u>Exercise 6.8</u>: Verify that (6.42a) can be rearranged to give (6.42b). That is, show that all rearrangements involve commuting Lie operators of functions which are in involution.

Observe that all the terms in $f_2 = -H_2$ and f_3 are independent of the time. This means that the map $\mathcal{M}$ leaves P_T unchanged. This is to be expected, of course, since a static quadrupole field does not change the energy of a particle. Observe also that all terms in f_3 contain at least one factor of P_T. This means that f_3 is zero at the design energy (or, equivalently, at the design momentum). Consequently f_3, in the case of a perfect quadrupole, describes purely chromatic effects.

6.2 The perfect sextupole

Consider next the case of a perfect sextupole of length ℓ. Again employing the coordinates shown in Fig. (1.1), and using the results of exercises 1.5 and 1.11, one finds in this case for the Hamiltonian K the result

$$K = -[(P_t^2/c^3) - m_o^2c^2 - p_y^2 - p_z^2]^{1/2} - qa_s(yz^2 - y^3/3). \qquad (6.44a)$$

A moment's thought shows that the design orbit for a sextupole is the same as that for a quadrupole, and hence it is advantageous to make the same canonical transformation (6.8). When this is done, the Hamiltonian K will be replaced by the new Hamiltonian H given by the expression

$$H = -[p_t^o + P_T)^2/c^2 - m_o^2c^2 - p_y^2 - p_z^2]^{1/2}$$
$$-qa_s(yz^2 - y^3/3) - (v_x^o)^{-1}(P_T + p_t^o). \qquad (6.44b)$$

Finally, H can be expanded about the design orbit as in (6.15) to give for the first few homogeneous polynomials H_n the result

$$H_0 = m_o^2c^2/p^o \qquad (6.45a)$$

$$H_1 = 0 \qquad (6.45b)$$

$$H_2 = (P_T^2/2)(m_o^2/p^{o^3}) + (p_y^2 + p_z^2)/(2p^o) \qquad (6.45c)$$

$$H_3 = [P_T^2/2][(m_o^2/p^{o^4} v_x^o)]$$

$$+ P_T(p_y^2 + p_z^2)/(2p^{o^2} v^o_x) \qquad (6.45d)$$

$$- qa_s(yz^2 - y^3/3).$$

As before, the symplectic map $\mathcal{M}$ produced by following through the sextupole the Hamiltonian flow generated by H is given by the expression

$$\mathcal{M} = \exp\{-\ell(:H_2: + :H_3: + \ldots\} . \qquad (6.46)$$

Eventually, we will want to factor $\mathcal{M}$. However, before doing so, it is interesting to examine the expression for $\mathcal{M}$ in the limiting case of a very short but very strong sextupole. That is, consider the limit $\ell \to 0$, $a_s \to \infty$ in such a way that $\lim \ell a_s = \alpha$. In this limit, one finds

$$\lim(\ell H) = -q\alpha(yz^2 - y^3/3). \qquad (6.47)$$

Therefore, the mapping $\mathcal{M}$ in this case is given by the simple expression

$$\mathcal{M} = \exp\{q\alpha:(yz^2 - y^3/3):\} . \qquad (6.48)$$

It is evident that the $\mathcal{M}$ of (6.48) leaves the quantities P_T, T, y, and z unchanged. Moreover, it is easily checked that

$$q\alpha:(yz^2 - y^3/3):p_y = -q\alpha y^2 \qquad (6.49a)$$

$$\{q\alpha:(yz^2 - y^3/3):\}^2 p_z = 0, \text{ etc.} \qquad (6.49b)$$

$$q\alpha:(yz^2 - y^3/3):p_z = 2q\alpha yz \qquad (6.50a)$$

$$\{q\alpha:(yz^2 - y^3/3):\}^2 p_z = 0, \text{ etc.} \qquad (6.50b)$$

Consequently, one finds the relations

$$\mathcal{M} p_y = p_y - q\alpha y^2 \qquad (6.51a)$$

$$\mathcal{M} p_z = p_z + 2q\alpha yz. \qquad (6.51b)$$

This is just the usual thin lens or impulsive "kick" approximation for the action of a short sextupole. What this brief calculation has demonstrated is that the kick approximation does indeed produce a symplectic map. Therefore its use in many applications is a very reasonable procedure.

<u>Exercise 6.9</u>: Evaluate $\mathcal{M}$ for a quadrupole in the kick approximation.

We turn now to the task of factoring $\mathcal{M}$. The required calculations are similar to those already performed for the case of the quadrupole. First, one finds that H_2 can be decomposed as in (6.21) with the result

$$H_2^T = (P_T^2/2)(m_o^2/p^{o^3}) \tag{6.52a}$$

$$H_2^y = p_y^2/(2p^o) \tag{6.52b}$$

$$H_2^z = p_z^2/(2p^o). \tag{6.52c}$$

Next, the actions of $\exp(-\ell{:}H_2^T{:})$, $\exp(-\ell{:}H_2^y{:})$, and $\exp(-\ell{:}H_2^z{:})$ are easily found. The action of $\exp(-\ell{:}H_2^T{:})$ is the same as in (6.25), and those of $\exp(-\ell{:}H_2^y{:})$ and $\exp(-\ell{:}H_2^z{:})$ are the $a_2 \to 0$ limits of (6.27) and (6.29) respectively. One finds, for example, the result

$$\exp(-\ell{:}H_2^y{:})\begin{pmatrix} y \\ p_y \end{pmatrix} = \begin{pmatrix} 1 & (\ell/p^o) \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y \\ p_y \end{pmatrix} \tag{6.53}$$

which is just the expression expected for the effect of a drift of length ℓ in the linear matrix approximation. Similarly, $\exp(-\ell{:}H_2^z{:})$ gives the same transformation in z,p_z space.

The cubic portion H_3 of the Hamiltonian can be decomposed in the form

$$H_3 = H_3^T + H_3^{Ty} + H_3^{Tz} + H_3^{yz} \ . \tag{6.54}$$

The first three components in (6.54) are the same as those given in (6.31); and the fourth, H_3^{yz}, is given by the expression

$$H_3^{yz} = -qa_s(yz^2 - y^3/3). \tag{6.55}$$

Now observe that many of the components of H_2 as given by (6.52) are in involution with the components of H_3. It follows that $\mathcal{M}$ as given by (6.46) can also be written in the form

$$\mathcal{M} = \exp(-\ell{:}H_2^T{:})\exp\{-\ell{:}(H_2^y + H_2^z + H_3^{yz}){:}\}\times$$
$$\exp\{-\ell{:}(H_3^T + H_3^{Ty} + H_3^{Tz}){:}\} \ . \tag{6.56}$$

Therefore, the only factorization that needs to be carried out is that for the quantity $\exp\{-\ell:(H_2^y + H_2^z + H_3^{yz}):\}$.

<u>Exercise 6.10</u>: Verify the factorization (6.56) starting from (6.46).

Suppose the desired factorization is written in the form

$$\exp\{-\ell:(H_2^y + H_2^z + H_3^{yz}):\} = \exp(-\ell:H_2^y:)\exp(-\ell:H_2^z:)\times$$
$$\exp(:f_3^{yz}:)\exp(:f_4:)\cdots \quad . \tag{6.57}$$

Then, following the usual procedure, the quantity f_3^{yz} is given by the expression

$$f_3^{yz} = -\ell \int_0^1 d\tau \exp(\tau\ell:H_2^y:)\exp(\tau\ell:H_2^z:)H_3^{yz}. \tag{6.58}$$

The integrand may be simplified using the isomorphism property (3.34) and (3.35) to give the result

$$\exp(\tau\ell:H_2^y:)\exp(\tau\ell:H_2^z:)H_3^{yz} =$$
$$-qa_s\{[\exp(\tau\ell:H_2^y:)y][\exp(\tau:H_2^z:)z]^2 - [\exp(\tau\ell:H_2^y:)y]^3/3\}. \tag{6.59}$$

Moreover, from (6.53), and its z counterpart, it follows that

$$\exp(\tau\ell:H_2^y:)y = y - (\tau\ell/p^o)p_y \tag{6.60a}$$
$$\exp(\tau\ell:H_2^z:)z = z - (\tau\ell/p^o)p_z. \tag{6.60b}$$

Consequently, the expression for f_3^{yz} can also be written in the form

$$f_3^{yz} = \ell qa_s \int_0^1 d\tau\{[y-(\tau\ell/p^o)p_y][z-(\tau\ell/p^o)p_z]^2-[y-(\tau\ell/p^o)p_y]^3/3\}. \tag{6.61}$$

The integration can now easily be carried out to give the explicit result

$$f_3^{yz} = \ell qa_s\{(yz^2 - y^3/3) - (\ell/p_o)(z^2p_y$$
$$- y^2p_y + 2yzp_z)/2 + (\ell/p^o)^2(yp_z^2 - yp_y^2$$
$$+ 2zp_yp_z)/3 - (\ell/p^o)^3(p_yp_z^2 - p_y^3/3)/4\}. \tag{6.62}$$

Observe that all terms in (6.62) except the first vanish in the short and strong sextupole limit as expected.

Combining (6.57) with (6.56) gives the final result

$$f_3 = -\ell H_3^{T} - \ell H_3^{Ty} - \ell H_3^{Tz} + f_3^{yz} \ . \qquad (6.63)$$

This completes the calculation of f_3 for the sextupole. Note that the first three terms in (6.63) describe chromatic effects since, according to (6.31), they all depend on P_T. However, the term f_3^{yz} is independent of P_T, and describes nonlinear effects due to the sextupole magnetic field.

Exercise 6.11: Verify (6.62) starting with (6.58).

6.3 The normal entry and exit dipole

For the sake of simplicity, the discussion of dipole magnets will be limited to the case of normal entry and exit as illustrated in Fig. (1.3). Also, the discussion of fringe field effects will be omitted although they are important.

The treatment of the normal entry and exit dipole is most easily carried out in cylindrical coordinates. It begins with the Hamiltonian (1.23) and the potentials of exercise 1.12. That is, the Hamiltonian with ϕ as an independent variable is given by the expression

$$K = \rho[p_t^2/c^2 - m_o^2c^2 - p_\rho^2 - p_z^2]^{1/2} - q\rho^2 B/2. \qquad (6.64)$$

Let p^o be the momentum of the design orbit. Then the design orbit in the dipole is a circular arc with radius ρ_o given by the familiar relation

$$\rho_o = p^o/(qB). \qquad (6.65)$$

In terms of these quantities, the design orbit is given by the equations

$$p_\rho = 0 \quad , \ \rho = \rho_o \qquad (6.66a,b)$$

$$p_z = 0 \quad , \ z = 0 \qquad (6.67a,b)$$

$$p_t = p_t^o \quad , \ t(\phi) = \phi(v_\phi^o)^{-1} + \text{constant}. \qquad (6.68a,b)$$

Here p_t^o and p^o are related by Eq. (6.17) as before; and the quantity v_ϕ^o is specified by the condition

$$dt/d\phi \Big|_{\substack{\text{design}\\ \text{orbit}}} = \partial K/\partial p_t \Big|_{\substack{\text{design}\\ \text{orbit}}} = (v_\phi^o)^{-1}. \tag{6.69}$$

Upon evaluation (6.69), one finds the result

$$v_\phi^o = [(p_t^o/c)^2 - m_o^2 c^2]^{1/2}[\rho_o p_t^o/c^2]^{-1} \quad . \tag{6.70}$$

As was the case with the quadrupole and sextupole, it is convenient to introduce new coordinates which facilitate expansion about the design orbit. They are defined in this instance by the relations

$$t = T + \phi(v_\phi^o)^{-1} \tag{6.71a}$$

$$\rho = \rho_o + Y \tag{6.71b}$$

$$z = Z \tag{6.71c}$$

$$p_t = p_t^o + P_T \tag{6.71d}$$

$$p_\rho = P_Y \tag{6.71e}$$

$$p_z = P_Z \quad . \tag{6.71f}$$

Evidently the design orbit corresponds to all the new variables having zero values. Simple calculation shows that the indicated change of variables can be obtained from the transformation function

$$F_2 = (\rho - \rho_o)P_Y + zP_Z + [t - \phi(v_\phi^o)]^{-1}[P_T + p_t^o] \,. \tag{6.72}$$

Now let H be the Hamiltonian for the new variables. It is given by the relation

$$H = K + \partial F_2/\partial\phi \tag{6.73}$$

or, more explicitly, by the relation

$$H = (\rho_o + y)[(p_t^o + P_T)^2/c^2 - m_o^2 c^2 - p_y^2 - p_z^2]^{1/2}$$
$$- q(\rho_o + y)^2 B/2 - (v_\phi^o)^{-1}(P_T + p_t^o) \quad . \tag{6.74}$$

Here, as before, lower case letters have again been introduced for simplicity. Note, however, that in this application y is not a Cartesian coordinate but instead describes deviations in ρ from the design value ρ_o.

The next step is to expand H in a power series about the design orbit as in (6.15). Upon expanding (6.74), one finds for the first few terms the result

$$H_0 = -\rho_o [(p^o/2) + (m_o^2 c^2/p^o)] \tag{6.75a}$$

$$H_1 = 0 \tag{6.75b}$$

$$H_2 = -\rho_o p_y^2/(2p^o) - qy^2B/2 - \rho_o p_z^2/(2p^o)$$

$$+ yP_T p_t^o/(p^o c^2) - P_T^2 \rho_o m_o^2/(2p^{o^3}) \tag{6.75c}$$

$$H_3 = P_T^3 \rho_o p_t^o m_o^2/(2p^{o^5} c^2) + P_T(p_y^2 + p_z^2) p_t^o/(2p^{o^3} c^2)$$

$$- yP_T^2 m_o^2/(2p^{o^3}) - y(p_y^2 + p_z^2)/(2p^o). \tag{6.75d}$$

Here use has been made of (6.65).

Note that H_1 is zero as expected. Observe also that the last term in H_3 is independent of P_T. It therefore produces effects similar to that of a sextupole. Indeed, terms of this sort from the various dipoles in a lattice make important contributions to the chromaticity in the case of a small ring.[19] A remark is also in order about the quadratic terms. The first two terms in H_2 are of the same sign and similar to those in (6.22b). They produce the horizontal focussing associated with a normal entry and exit magnet. The third term shows there is no focussing in the vertical direction. The fourth term involving the product yP_T describes the dispersion produced by a bend. Finally, the last term describes the dependence of transit time on energy.

Exercise 6.12: Verify Eqs. (6.75).

We are ready to consider the symplectic map $\mathcal{M}$ produced by following the Hamiltonian flow specified by H. Since ϕ now plays the role of an independent variable, the map $\mathcal{M}$, in analogy to (5.34), can be written in the form

$$\mathcal{M} = \exp\{-(\phi^f - \phi^i):H:\}. \tag{6.76}$$

Here ϕ^i is the initial value of ϕ at the entry to the dipole, and ϕ^f is the final value upon exit. But according to Figs. (1.1) and (1.3), ϕ decreases as a particle moves through a dipole. Therefore, the change in ϕ can be written in the form

$$\phi^f - \phi^i = -\theta,$$

where θ is a positive quantity equal to the bend angle as illustrated in Fig. (1.3). With this explanation, the symplectic map M can be written in the form

$$M = \exp\{\theta(:H_2: + :H_3: + \ldots)\} \quad . \tag{6.77}$$

All that remains to be done is to factor M. As a first step in this process, suppose H_2 as given by (6.75c) is written as a sum of three terms mutually in involution,

$$H_2 = H_2^T + H_2^z + H_2^{yT} \tag{6.78}$$

$$H_2^T = -P_T^2 \rho_o m_o^2/(2p^{o^3}) \tag{6.79a}$$

$$H_2^z = -\rho_o p_z^2/(2p^o) \tag{6.79b}$$

$$H^{yT} = -\rho_o p_y^2/(2p^o) + yP_T/(\rho_o v_\phi^o) - qy^2B/2 \ . \tag{6.79c}$$

Here, in simplifying (6.79c), use has been made of the relation

$$(v_\phi^o)^{-1} = (\rho_o p_t^o/c^2)[(p_t^o/c)^2 - m_o^2c^2]^{-1/2} = (\rho_o p_t^o)/(p^o c^2). \tag{6.80}$$

The evaluation of $\exp(\theta:H_2^T:)$ and $\exp(\theta:H_2^z:)$ is straight forward, and introduces no new features beyond those already encountered in earlier sections. However, the calculation of $\exp(\theta:H_2^{yT}:)$ requires a bit more work since H_2^{yT} depends on the three variables y, p_y, and P_T. Evidently, $\exp(\theta:H_2^{yT}:)$ leaves z, p_z, and P_T unaffected. Moreover, the action of powers of $:H_2^{yT}:$ on the variables y, p_y, and T is easily calculated. One finds the result

$$:H_2^{yT}: \begin{bmatrix} y \\ p_y \\ T \\ P_T \end{bmatrix} = G \begin{bmatrix} y \\ p_y \\ T \\ P_T \end{bmatrix} \tag{6.81}$$

where G denotes the matrix

$$G = \begin{bmatrix} 0 & \rho_o/P^o & 0 & 0 \\ -qB & 0 & 0 & (\rho_o v_\phi^o)^{-1} \\ -(\rho_o v_\phi^o)^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} . \tag{6.82}$$

It follows that the action of $\exp(:H_2^{yT}:)$ is given by the relation

$$\exp(\theta :H^{yT}:) \begin{bmatrix} y \\ P_y \\ T \\ P_T \end{bmatrix} = \exp(\theta G) \begin{bmatrix} y \\ P_y \\ T \\ P_T \end{bmatrix} . \tag{6.83}$$

The determination of $\exp(\theta G)$ is facilitated by an algebraic property of the matrix G. A short calculation shows that G has the characteristic polynomial $P(\lambda)$ given by the equation

$$P(\lambda) = \det(G - \lambda I) = \lambda^2(\lambda^2 + 1). \tag{6.84}$$

Here use has been made of (6.65). Consequently, by the Cayley-Hamilton theorem, the matrix G has the special property[18]

$$G^4 = -G^2. \tag{6.85}$$

The property (6.85) makes it easy to sum the exponential series for exp(G). One finds the result

$$\begin{aligned} \exp(\theta G) &= \cosh(\theta G) + \sinh(\theta G) \\ &= [I + (\theta G)^2/2! + (\theta G)^4/4! + (\theta G)^6/6! + \cdots] \\ &+ [(\theta G) + (\theta G)^3/3! + (\theta G)^5/5! + (\theta G)^7/7! + \cdots] \\ &= [I + G^2(\theta^2/2! - \theta^4/4! + \theta^6/6! + \cdots)] \\ &+ [\theta G + G^3(\theta^3/3! - \theta^5/5! + \theta^7/7! + \cdots)] \\ &= I + \theta G + G^2(1 - \cos\theta) + G^3(\theta - \sin\theta). \end{aligned} \tag{6.86}$$

Upon inserting the explicit form (6.82) for G into (6.86), one finds the desired expression

$$\exp(\theta G) = \begin{bmatrix} \cos\theta & (\rho_o/p^o)\sin\theta & 0 & (p^o v^o_\phi)^{-1}(1-\cos\theta) \\ -(p^o/\rho_o)\sin\theta & \cos\theta & 0 & (\rho_o v^o_\phi)^{-1}\sin\theta \\ -(\rho_o v^o_\phi)^{-1}\sin\theta & (1-\cos\theta)/(p^o v^o_\phi) & 1 & (p^o\rho_o)^{-1}(v^o_\phi)^{-2}(\sin\theta-\theta) \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.87}$$

Note that (6.87) describes the expected effect of dispersion in the median plane of the bend.

All tools have now been assembled for the calculation f_3. Because no new techniques are involved, the actual work will be left as an exercise for the reader.

<u>Exercise 6.13</u> Verify Eqs. (6.84) through (6.87).

<u>Exercise 6.14</u>: Calculate f_3 for the normal entry and exit dipole.

6.4 The straight section drift

The calculation of results for a drift have been saved for last because they are particularly easy. Consider the quadrupole Hamiltonian (6.13) in the case of zero field strength. Then one finds for H an expansion of the form (6.15) with the homogeneous polynomials H_0 through H_4 given explicitly by the expressions

$$H_0 = m_o^2c^2/p^o \tag{6.88a}$$

$$H_1 = 0 \tag{6.88b}$$

$$H_2 = (P_T^2/2)(m_o^2/p^{o^3}) + (p_y^2 + p_z^2)/(2p^o) \tag{6.88c}$$

$$H_3 = [P_T^3/2][(m_o^2)/(p^{o^4} v_x^o)] + [P_T(p_y^2 + p_z^2)/(2p^{o^2} v_x^o)] \tag{6.88d}$$

$$H_4 = P_T^4 m_o^2 c^2(5m_o^2c^2 + 4p^{o^2})/(8p^{o^7} c^4) \tag{6.88e}$$

$$+P_T^2(p_y^2 + p_z^2)(3m_o^2c^2 + 2p^{o^2})/(4p^{o^5} c^2) + (p_y^2 + p_z^2)^2/8p^{o^3}).$$

Observe that all the terms in Eqs. (6.88) are mutually in involution.

It follows that the symplectic map $\mathcal{M}$ for a drift can be factored immediately to give the result

$$\begin{aligned} \mathcal{M} &= \exp\{-\ell:(H_2 + H_3 + H_4 + \cdots):\} \\ &= \exp(-\ell:H_2:)\exp(-\ell:H_3:)\exp(-\ell:H_4:) \cdots . \end{aligned} \tag{6.89}$$

Note that H_4 contains the term $(p_y^2 + p_z^2)^2$. This is an example of a "kinematic" nonlinearity which occurs even apart from chromatic effects and the effects of electric and magnetic fields.

Exercise 6.15: Verify Eqs. (6.88) and (6.89).

6.5 Application to light optics

According to section 1.3 and exercise 4.3, the passage of light rays through an optical system can be described by a symplectic map. When applied to optics, the factorization theorem indicates that the effect of any collection of lenses, prisms, and mirrors can be characterized by a set of homogeneous polynomials. It is easy to verify that the polynomials f_2 reproduce Gaussian optics, and the higher order polynomials f_3, f_4, etc. describe departures from Gaussian optics, and are related to aberrations in the case of an imaging system. Thus, from a Lie algebraic perspective, the fundamental problem of geometrical optics is to study what polynomials correspond to various optical elements, and to study what polynomials correspond to various desired optical properties.

Exercise 6.16: Show that $\mathcal{M}$ given by

$$\mathcal{M} = \exp\{-\ell/(2n):(\vec{p}^{\,i})^2:\}$$

corresponds to transit by a distance ℓ through a medium of refractive index n in the Gaussian approximation.

Answer:

$$p_\alpha^f = \mathcal{M} p_\alpha^i = p_\alpha^i$$

$$q_\alpha^f = \mathcal{M} q_\alpha^i = q_\alpha^i + (\ell/n)p^i .$$

Exercise 6.17: Show that $\mathcal{M}$ given by

$$\mathcal{M} = \exp\{(n_2 - n_1)/(2r):(\vec{q}^{\,i})^2:\}$$

corresponds to refraction in the Gaussian approximation for rays passing through a spherical interface of radius r from a medium having index of refraction n_1 to a second medium having index n_2.

<u>Answer</u>:

$$q_\alpha^f = \mathcal{M} q_\alpha^i = q_\alpha^i$$

$$p_\alpha^f = \mathcal{M} p_\alpha^i = p_\alpha^i + \{(n_2 - n_1)/r\} q_\alpha^i .$$

To simplify discussion, suppose, as is often the case, that the optical system under consideration is <u>axially symmetric</u> about some axis and is also <u>symmetric</u> with respect to <u>reflections</u> through some plane containing the axis of symmetry. Then axial symmetry requires that the various f_n be functions only of the variables $\vec{p}^2$, $\vec{q}^2$, $\vec{p}\cdot\vec{q}$, and $\vec{p}\times\vec{q}$; and reflection symmetry rules out the variable $\vec{p}\times\vec{q}$. It follows that all the f_n with odd n must vanish since it is impossible to construct an odd order homogeneous polynomial using only the variables $\vec{p}^2$, $\vec{p}\cdot\vec{q}$, and $\vec{q}^2$. Consequently, in any case having the assumed symmetries, the optical symplectic map $\mathcal{M}$ must be of the general form

$$\mathcal{M} = \exp(:f_2:)\exp(:f_4:)\exp(:f_6:) \cdots . \tag{6.90}$$

Of course in the general case of no particular symmetries, the odd degree polynomials f_3, f_5, etc., can also occur, and all the polynomials f_n can in principle depend on the components of the vectors $\vec{p}$ and $\vec{q}$ in an unrestricted fashion.

It can be shown that the polynomials f_4, f_6, etc. in the case of an imaging system are related to third-order, fifth-order, etc. aberrations respectively. Consider the case of f_4. According to the previous discussion of symmetry, f_4 in the cases of present interest can depend only on the variables $\vec{p}^2$, $\vec{p}\cdot\vec{q}$, and $\vec{q}^2$. Consequently, f_4 must be of the general form

$$f_4 = A(\vec{p}^2)^2 + B\,\vec{p}^2(\vec{p}\cdot\vec{q}) + C(\vec{p}\cdot\vec{q})^2 + D\,\vec{p}^2\vec{q}^2 + E(\vec{p}\cdot\vec{q})\vec{q}^2 + F(\vec{q}^2)^2 . \tag{6.91}$$

Here the quantities A through F are arbitrary coefficients whose values depend on the particular optical system under consideration. When employed in (6.90), the last term $F(\vec{q}^2)^2$ has no effect on the quality of an image since $[F(\vec{q}^2)^2, q_\alpha]=0$. (It does, however affect the arrival direction $\vec{p}$ of a ray at the image plane, and therefore may be important if the optical system under study is to be used for some other purpose as part of a larger optical system.) The remaining terms do affect the image. Specifically, one finds in the imaging case the following one to one correspondence between the terms in the expansion (6.91) and the classical Seidel third-order monochromatic aberrations:

TABLE OF ABBERATIONS

Term	Seidel Aberration
$A(\vec{p}^{\,2})^2$	Spherical Aberration
$B\vec{p}^{\,2}(\vec{p}\cdot\vec{q})$	Coma
$C(\vec{p}\cdot\vec{q})^2$	Astigmatism
$D\ \vec{p}^{\,2}\vec{q}^{\,2}$	Curvature of Field
$E(\vec{p}\cdot\vec{q})\vec{q}^{\,2}$	Distortion

In reflecting upon what has been accomplished so far, it is evident that all that has been assumed is that the optical map $\mathcal{M}$ is symplectic (a consequence of Fermat's principle) and has certain symmetries. As a consequence of these assumptions, Gaussian optics was obtained as a first approximation, and it was found that for an imaging system only five well-defined kinds of monochromatic aberrations can occur in third order. The use of Lie algebraic methods seems to be an optimal way of arriving at and understanding these basic results.

To proceed further, it is necessary to have a catalog of Lie operators corresponding to various simple optical elements. The first simple optical element to be considered is transit through a slab of thickness ℓ composed of a homogeneous medium having constant refractive index n. In this case, the associated symplectic map $\mathcal{M}$ is found to be given by the expression

$$\mathcal{M} = \exp\{\ell{:}(n^2 - \vec{p}^{\,2})^{1/2}{:}\} = \exp\{(-\ell)/(2n){:}\vec{p}^{\,2}{:}\}\times$$
$$\exp(-\ell)/(8n^3){:}(\vec{p}^{\,2})^2{:}\}\cdots \quad . \tag{6.92}$$

<u>Exercise 6.18</u>: Verify (6.92) starting with (1.30) and (6.14) and making a power series expansion.

Next consider a lens. Figure 6.1 shows a lens with planar entrance and exit faces. It is composed of two media having indices of refraction n_1 and n_2 separated by a curved interface. Suppose a cartesian coordinate system is located in the exit face (right face) of the lens with the z axis along the optical axis and the x and y axes lying in the face of the lens. Then the shape of the curved interface is taken to be given by the equation

$$z = -\alpha(x^2 + y^2) + \beta(x^2 + y^2)^2 + \gamma(x^2 + y^2)^3 + \cdots \,. \tag{6.93}$$

For the case of a spherical lens with radius of curvature r, the shape of the surface is described by the relation

$$z = -r + (r^2 - x^2 - y^2)^{1/2} \quad . \tag{6.94}$$

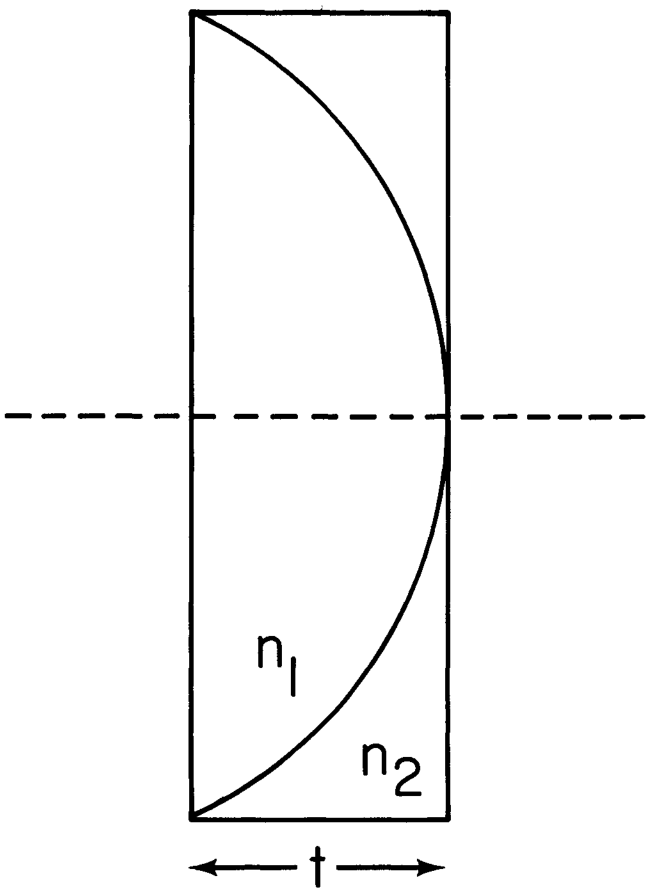


Fig. 6.1: A lens with planar entrance and exit faces. It is composed of two media having indices of refraction n_1 and n_2, and separated by a curved interface.

In this case the quantities α and β have the explicit values

$$\alpha = 1/(2r)$$
$$\beta = -\alpha^3 = -1/(8r^3). \qquad (6.95)$$

Evidently, the most general lens with axial symmetry (including the case of a corrector plate) can be obtained by joining together lenses of the type shown in Fig. 6.1 and their reversed counterparts, perhaps with slabs of constant index material in between.

Calculations show that the symplectic map $\mathcal{M}$ for this lens, when light passes from left to right, is given through fourth order by the expression

$$\mathcal{M} = \exp\{(-t)/(2n_1):\vec{p}^{\,2}:\} \exp\{(-t)/(8n_1^3):(\vec{p}^{\,2})^2:\}\times$$
$$\exp\{\alpha(n_2-n_1):\vec{q}^{\,2}:\} \exp(:f_4:). \qquad (6.96)$$

Each of the terms appearing in (6.96) has a simple interpretation. Proceeding from left to right, comparison with (6.92) shows that the first two terms in (6.96) simply correspond to propagation through a slab of thickness t and refractive index n_1. Here, as shown in Fig. 6.1, t is the thickness of the lens. The third term, $\exp\{\alpha(n_2-n_1):\vec{q}^{\,2}:\}$, simply produces the refraction expected for a thin lens in Gaussian approximation. See exercise 6.17 . Finally, the last term, $\exp(:f_4:)$, describes third order departures from Gaussian optics produced at the refracting interface.

In general, f_4 has an expansion of the form (6.91) because the lens shown in Fig. 6.1 and described by (6.93) has the required symmetry properties. Moreover, calculation shows that the coefficients A through F have the explicit values

$$A = 0 \quad , \quad B = 0 \quad , \quad C = 0$$
$$D = \alpha(n_2-n_1)/(2n_1n_2)$$
$$E = 2\alpha^2(n_1-n_2)/n_1 \qquad (6.97)$$
$$F = \alpha^3(n_1-n_2)\{n_1[(\beta/\alpha^3)+2] - 2n_2\}/n_1.$$

Observe that three of the six possibly nonzero terms are in fact zero. This fact may be interpreted as a good omen indicating that the Lie algebraic description is ideally suited in some sense to the task of characterizing the optical properties of a lens. However, Eqs. (6.97) should not be interpreted to mean that the lens of Fig. 6.1 has no spherical aberration, coma, or astigmatism. Strictly speaking, as described earlier, aberrations are not properties of a lens by itself, but rather are properties of an optical system which is imaging in the

Gaussian approximation. As will be seen shortly, when the lens in question is preceeded and followed by transit in order to produce an imaging system, then the complete system generally does have these and other aberrations.

To see how Lie transformations may be concatenated to describe a compound optical system, consider the simple case of an imaging system consisting of transit in air (n=1) over a distance d_1, refraction by the lens of Fig. 6.1 with $n_1 = n$ and $n_2 = 1$, and final transit in air over a distance d_2. Using the results already obtained, the symplectic map M for this sytem is given by the product of transit, refraction, and transit:

$$\begin{aligned} M = {} & \exp\{(-d_1/2):\vec{p}^2:\} \exp\{(-d_1/8):(\vec{p}^2)^2:\} \times \\ & \exp\{(-t)/(2n):\vec{p}^2:\} \exp\{(-t)/(8n^3):(\vec{p}^2)^2:\} \times \\ & \exp\{\alpha(1-n):\vec{q}^2:\} \exp(:f_4:) \times \\ & \exp\{(-d_2/2):\vec{p}^2:\} \exp\{(-d_2/8):(\vec{p}^2)^2:\} \quad . \end{aligned} \tag{6.98}$$

Because the optical system is assumed to be imaging in the Gaussian approximation, the quantities d_1, d_2, t, α, and n are taken to be related by the familiar imaging condition

$$1/(d_1 + t/n) + 1/d_2 = 1/f. \tag{6.99}$$

Here f is the focal length of the lens defined by the relation

$$1/f = (n-1)/r = 2\alpha(n-1). \tag{6.100}$$

Using the tools developed previously, the expression (6.98) can be re-expressed in the factored product form

$$M = M_G \exp(:f_4^*:) \cdots . \tag{6.101}$$

Here the quantity M_G denotes the Gaussian portion of the map, and is given by the relation

$$\begin{aligned} M_G = {} & \exp\{(d_1/2):\vec{p}^2:\} \times \\ & \exp\{(-t)/(2n):\vec{p}^2:\} \exp\{\alpha(1-n):\vec{q}^2:\} \times \\ & \exp\{(-d_2/2):\vec{p}^2:\} \end{aligned} \tag{6.102}$$

as would be expected from Gaussian optics. The determination of the function f_4^* requires somewhat more effort. In general, f_4^* has an expansion of the same form as that in (6.91). The explicit determination of the coefficient can be carried out using Eqs. (5.70) and (5.80).

Denoting the coefficients occurring in f_4^* by the symbols A^* through F^*, one finds the results

$$\begin{aligned}
A^* &= -m^4(d_1 + tn^{-3})/8 + (d_2^2 D - d_2^3 E + d_2^4 F) - d_2/8 \\
B^* &= -m^3(d_1 + tn^{-3})/(2f) + (-2d_2 D + 3d_2^2 E - 4d_2^3 F) \\
C^* &= -m^2(d_1 + tn^{-3})/(2f^2) + (-2d_2 E + 4d_2^2 F) \qquad (6.103) \\
D^* &= -m^2(d_1 + tn^{-3})/(4f^2) + (D - d_2 E + 2d_2^2 F) \\
E^* &= -m(d_1 + tn^{-3})/(2f^3) + (E - 4d_2 F) \\
F^* &= -(d_1 + tn^{-3})/(8f^4) + F.
\end{aligned}$$

Here the quantities D, E,and F are those given by (6.97).

Evidently all the coefficients A^* through F^* are nonzero in the general case. Thus an imaging system made with a single lens having one flat face generally suffers form all five Seidel aberrations.

By looking at the terms appearing in the coefficients A^* through F^*, it is possible to determine the various sources of aberration. For example, consider A^* which describes spherical aberration. The first term in A^* arises from transit in air over the distance d_1 and transit through the thickness of the lens, the second set of terms is produced by the lens itself, and the last term is caused by the final transit over the distance d_2. From this result it is clear that simple transit, when combined with even perfect Gaussian refraction, is itself a source of aberration.

Further discussion of the use of Lie algebraic methods in light optics is outside the scope of this section. Additional results may be found elsewhere.[20]

Exercise 6.19: Show that $\mathcal{M}_G$ as given by (6.102) has the action

$$\mathcal{M}_G \begin{bmatrix} \vec{q} \\ \vec{p} \end{bmatrix} = \begin{bmatrix} m & 0 \\ -1/f & 1/m \end{bmatrix} \begin{bmatrix} \vec{q} \\ \vec{p} \end{bmatrix}$$

where m is the magnification defined by the ratio

$$m = -d_2/(d_1 + t/n).$$

Exercise 6.20: Verify the expressions for A^* through F^*.

6.6 Work in progress

The preparation of a catalog of Lie operators as outlined in sections 6.1 through 6.4 is being continued to embrace all common beam elements including active elements such as accelerating cavities and bunchers. In addition, the catalog will contain all fourth order polynomial terms f_4 as well as the f_2 and f_3 terms previously described. Consequently the catalog will include octupoles as well as sextupoles.

Second, this catalog is being incorporated in a computer program which is being written to concatenate various beam elements, using the calculus for Lie transformations described in Section 5.1, in order to find the complete behavior for any collection of beam elements. The goal of this effort is to produce a charged particle beam transport code whose function is similar to the widely used program TRANSPORT,[21] but which will routinely work to one higher order in nonlinear effects so as to treat octupole elements, iterated sextupoles, and all other sources of third-order nonlinearities. Because the Lie algebraic methods automatically take into account the symplectic nature of M, it is expected that storage requirements for the program under development will be approximatley 34 times less than what would be required if one were to try to extend the methods of TRANSPORT to one order higher. The methods of TRANSPORT, when extended to the next higher order, would require the use of 83×83 matrices in the general case. This requires $(83)^2 = 6889$ storage locations for each matrix. By contrast, the Lie algebraic methods require only the storage of the coefficients in the polynomials f_2, f_3, and f_4. This requires only 203 storage locations for each symplectic map in the general case. It is also hoped that the program will be fast enough so that it can be run iteratively to produce self-consistent space-charge forces in the case of intense beams. Finally, it is hoped that the description of nonlinearities in terms of polynomials rather than in terms of large matrices will provide additional insight into methods for both controlling and exploiting nonlinear behavior.

7. APPLICATIONS TO ORBITS IN CIRCULAR MACHINES AND COLLIDING BEAMS

The purpose of this chapter is to apply the results developed in previous chapters to the case of orbits in circular machines and colliding beam machines. By "circular" it is meant that each orbit passes through the same set of beam elements over and over again. Attention will be devoted to the existence of closed orbits and their stability. Stability will be discussed both in the linear approximation and in the general case where nonlinear structure resonances must be considered. Finally, the beam-beam interaction will be treated using methods analogous to those developed for structure resonances.

7.1 Existence of closed orbits

A perfectly constructed lattice for a "circular" machine has at least one closed orbit, namely the design orbit associated with the design momentum. In this section, it is shown that, under very general conditions, a perfect lattice also has a unique closed orbit for every other momentum value. Moreover, an efficient method is presented for its computation. Finally, as a byproduct of the discussion, it is shown that the same conclusions hold for a perturbed lattice. This is comforting, because lattices as actually constructed are, of course, unavoidably imperfect.

The proof of the existence and the computation of the off-momentum closed orbits are facilitated by topological methods that have their basis in the work of Poincare. These methods are both elegant and powerful, and therefore worthy of a brief explanation. Basically, they consist of converting the problem of studying orbits into an equivalent mathematical problem of studying a related symplectic map. This symplectic map is in turn studied by fixed point methods.

The first step is the construction of a Poincare surface of section. Consider an imaginary plane which intersects the design orbit at right angles somewhere in some straight section. Then it is obvious that every other orbit will also intersect this plane. Indeed, every orbit intersects this plane each time it goes around the lattice.

In our subsequent discussion, we will assume that the total momentum p remains a constant along an orbit. This is equivalent to assuming that all accelerating sections and bunchers have been turned off, and the effect of synchrotron radiation, if any, is ignored. Then every orbit with a given fixed total momentum p is completely determined by the values of its two transverse coordinates and two transverse momenta at the point of intersection. Indeed, the orbit equations can be written in the form

$$u_i' = g_i(\vec{u},\theta) \qquad i = 1,\ldots 4. \qquad (7.1)$$

Here $\vec{u}$ denotes a four-component vector whose entries are the two coordinates and two momenta transverse to the beam axis. The quantity θ denotes some generalized angle which increases by 2π in going around the lattice, and a prime denotes differentiation with respect to θ. Consequently, the values of the four quantities u_i at the point of intersection may be viewed as a complete set of initial conditions for the four first-order orbit differential equations. Finally, it is only necessary to recall that the solution of a set of first-order differential equations is uniquely and completely specified by initial conditions. The whole situation may be summarized by saying that there is a certain four-dimensional hypersurface in phase space which cuts across every phase-space trajectory for the orbits under study, and therefore is appropriately called a <u>surface of section</u>. In addition, each orbit is uniquely specified in terms of any point (four coordinates) at which it crosses the surface.

> <u>Exercise 7.1</u>: Supply the necessary reasoning to show that the equations of motion for the transverse coordinates and momenta can be written in the form (7.1). Show that in fact the equations of motion come from a Hamiltonian.
>
> <u>Hint</u>: Imagine applying theorem (1.1) to each beam element in the lattice. In each element, select θ to be proportional to q_1, perhaps with a proportionality constant which varies from element to element. Observe that, according to (1.21c), p_t (or equivalently the total momentum p) is a constant since under the assumptions made, K is time independent for each beam element. Therefore, as far as the equations of motion for the remaining variables are concerned, p_t may be treated as a constant parameter. Also, Eq. (1.21a) may be ignored since it is unnecessary to know the transit time in order to determine the orbit. The remaining Eqs. (1.21b) and (1.21d), with suitable scaling from element to element to reflect the varying proportionality between q_1 and θ, then lead to Eqs. (7.1).

The next step is to observe that orbits in the lattice generate a <u>symplectic mapping</u> M of the surface of section into itself. Consider a point on the surface of section. Since any such point requires four numbers for its specification, it is convenient to denote these four numbers collectively by a four-component vector $\vec{a}$. Now use the coordinates of $\vec{a}$ as initial conditions, and follow the orbit with these initial conditions once around the lattice until it again crosses the surface of section at some point $\vec{b}$. The mapping M, called the <u>Poincare map</u>, is simply defined by the relation

$$\vec{b} = M\vec{a}. \tag{7.2}$$

That is, M describes the effect of <u>one circuit around the lattice</u>.

Note that since the equations of motion are in general nonlinear, the relation between $\vec{b}$ and $\vec{a}$ defined by the Poincare map M is also nonlinear. Finally, the symplectic nature of the map follows from theorem (4.1) with the independent variable θ playing the role of the time variable. That is, the Poincare map M is the result of following a Hamiltonian flow. (See exercise 7.1).

Much of what one wants to know about orbits is equivalent to a knowledge of M. For example, the key question of the long term behavior of orbits for a large number of turns is equivalent to a knowledge of M^n for large n. More particularly of interest for the current discussion, the determination of a closed orbit is equivalent to the discovery of a point $\vec{f}$, called a fixed point, which is sent into itself under the action of M,

$$M\vec{f} = \vec{f}. \tag{7.3}$$

That is, the set of initial conditions $\vec{f}$ for a closed orbit must, by definition, be mapped into itself by one circuit around the lattice.

Another concept needed is that of the linear part of a map. Let $\vec{a}$ be an arbitrary point in the surface of section and let $\vec{a} + \vec{\varepsilon}$, where $\vec{\varepsilon}$ is a small vector, be a point near $\vec{a}$. Now consider the point $M(\vec{a} + \vec{\varepsilon})$. According to Eq. (7.2) this point should be near $\vec{b}$ since $\vec{a} + \vec{\varepsilon}$ is near $\vec{a}$. In fact, there is a power series expansion in $\vec{\varepsilon}$ of the form

$$M(\vec{a} + \vec{\varepsilon}) = \vec{f} + M_a\vec{\varepsilon} + O(\vec{\varepsilon}^2) \tag{7.4}$$

where M_a is the 4×4 Jacobian matrix of M at the point $\vec{a}$. (See section 4.1.) Evidently, M_a describes the linearized behavior of M in the vicinity of $\vec{a}$. For this reason, the Jacobian matrix M_a can also be viewed as the linear part of M at $\vec{a}$.

The linear part of M can be conveniently calculated from the variational equations associated with the main orbit Eqs. (7.1). Variational equations describe all orbits near a particular orbit. In contrast to the main orbit equations, the variational equations are linear by construction. Let $\vec{u}_*(\theta)$ denote a particular orbit of interest. Then orbits near this orbit can be written in the form

$$\vec{u} = \vec{u}_*(\theta) + \varepsilon\vec{w} \tag{7.5}$$

where ε is a small quantity. By definition, $\vec{u}_*(\theta)$ satisfies the Eq. (7.1). Inserting the prescription (7.5) into the equation of motion (7.1), and retaining terms of lowest order in ε, shows that $\vec{w}$ must satisfy the variational equation

$$\vec{w}' = A_*(\theta)\vec{w}. \tag{7.6}$$

Here A_* is a 4 × 4 theta-dependent matrix defined by

$$A_{*ij}(\theta) = \partial g_i(\vec{u},\theta)/\partial u_j \Big|_{\vec{u} = \vec{u}_*(\theta)}. \tag{7.7}$$

It is evident that the variational Eq. (7.6) is linear in $\vec{w}$, and that the θ dependence of A_*, as indicated by the subscripted star, depends in general on the orbit $\vec{u}_*(\theta)$ about which variations are being studied.

Because the variational equations are linear, their solution for all initial conditions can be obtained by a finite amount of computation: Let $\theta = 0$ denote some arbitrary point in the lattice. Consider the first-order linear <u>matrix</u> variational equation defined by

$$B_*'(\theta) = A_*(\theta)\, B_*(\theta) \tag{7.8}$$

with the initial condition

$$B_*(0) = I. \tag{7.9}$$

Here B_* is a 4 × 4 matrix, I denotes the 4 × 4 identity matrix, and A_* is the same matrix as defined in (7.7). The solution to (7.8), with the initial condition (7.9), is uniquely defined and can be obtained if desired by a finite amount of numerical integration. Indeed, the integration of (7.8) is equivalent to the integration of 16 linear first-order equations (since B_* is 4 × 4) with the single set of initial conditions specified by (7.9). Now let $\vec{w}^o$ be an arbitrary four-component vector. Consider $\vec{w}(\theta)$ defined by the equation

$$\vec{w}(\theta) = B_*(\theta)\vec{w}^o. \tag{7.10}$$

One easily checks that $\vec{w}(\theta)$ is a solution to the variational equations (7.6) and satisfies the arbitrarily prescribed initial condition

$$\vec{w}(0) = \vec{w}^o. \tag{7.11}$$

It is now easy to see that the linear part of $\mathcal{M}$ is available from the variational equations. Suppose the surface of section is located at $\theta = 0$. Let $\vec{u}_a(\theta)$ be the orbit with initial conditions $\vec{a}$. That is,

$$\vec{u}_a(0) = \vec{a}. \tag{7.12}$$

Then this trajectory must also satisfy the equation

$$\vec{u}_a(2\pi) = \mathcal{M}\vec{a} = \vec{b}. \tag{7.13}$$

Next, with the aid of the variational equations, the nearby trajectory $\vec{u}_{a+\varepsilon}(\theta)$ with initial conditions $\vec{a} + \vec{\varepsilon}$ is expressible in the form

$$\vec{u}_{a+\varepsilon}(\theta) = \vec{u}_a(\theta) + B_a(\theta)\vec{\varepsilon} + O(\vec{\varepsilon}^2). \tag{7.14}$$

[See Eqs. (7.5), (7.10), and (7.11).]

Now put $\theta = 2\pi$ in Eq. (7.14). The result is the relation

$$\mathcal{M}(\vec{a} + \vec{\varepsilon}) = \vec{u}_{a+\varepsilon}(2\pi) = \vec{u}_a(2\pi) + B_a(2\pi)\vec{\varepsilon} + O(\vec{\varepsilon}^2). \tag{7.15}$$

Comparison of (7.15) and (7.4), with the aid of (7.13), gives the result

$$M_a = B_a(2\pi). \tag{7.16}$$

The stage has now been set for the determination of the fixed points of $\mathcal{M}$. This will be done with the aid of another map $\mathcal{C}$, called a <u>contraction map</u>, defined in terms of $\mathcal{M}$. Let $\vec{e}$ be an <u>arbitrary</u> point in the vicinity of a fixed point $\vec{f}$. The contraction map $\mathcal{C}$ will be shown to have the remarkable property

$$\vec{f} = \lim_{n\to\infty} \mathcal{C}^n \vec{e}\ . \tag{7.17}$$

That is, a guess as to the whereabouts of a closed orbit is sufficient starting information to contract in on it exactly. In practice, the starting guess can be taken to be the initial conditions for the on-momentum design orbit.

The construction of the contraction mapping $\mathcal{C}$ is a generalization of Newton's method to the case of several variables.[22] The map $\mathcal{C}$ is defined by requiring that its action on the arbitrary point $\vec{a}$ be given by the rule

$$\mathcal{C}\vec{a} = \vec{a} - (I - M_a)^{-1} (\vec{a} - \mathcal{M}\vec{a}). \tag{7.18}$$

It is easily verified that the map $\mathcal{C}$ defined by (3.10) has the advertised property (7.17). First, suppose that $\vec{f}$ is a fixed point of $\mathcal{M}$. Then, it is easily verified that $\vec{f}$ is also a fixed point of $\mathcal{C}$,

$$\mathcal{C}\vec{f} = \vec{f}. \tag{7.19}$$

Next, suppose that $\vec{e}$ is some point in the vicinity of $\vec{f}$. Then $\vec{e}$ is of the form $\vec{f} + \vec{\varepsilon}$ where $\vec{\varepsilon}$ is some small vector. Inserting $\vec{e}$ in Eq. (7.19), one finds after a short calculation that

$$\begin{aligned} \mathcal{C}\vec{e} = \mathcal{C}(\vec{f} + \vec{\varepsilon}) &= \vec{f} + \vec{\varepsilon} - (I - M_{f+\varepsilon})^{-1} [(\vec{f} + \vec{\varepsilon}) - \mathcal{M}(\vec{f} + \vec{e})] \\ &= \vec{f} + \vec{e} - (I - M_{f+\varepsilon})^{-1} [(I - M_f)\vec{\varepsilon} + O(\vec{\varepsilon}^2)] \\ &= \vec{f} + O(\vec{\varepsilon}^2). \end{aligned} \tag{7.20}$$

Here use has been made of the relation

$$\mathcal{M}(\vec{f} + \vec{\varepsilon}) = \vec{f} + M_f\vec{\varepsilon} + O(\vec{\varepsilon}^2) \tag{7.21}$$

and the observation that

$$M_{f+\varepsilon} = M_f + O(\varepsilon). \tag{7.22}$$

Thus, according to (7.20), although the initial point $\vec{e}$ differs from the desired fixed point $\vec{f}$ by an amount $\vec{\varepsilon}$, the point $\mathcal{C}\,\vec{e}$ differs from the point $\vec{f}$ only by an amount of order $\vec{\varepsilon}^2$. Similarly, the point $\mathcal{C}^2\vec{e}$ differs from $\vec{f}$ only by an amount of order $(\vec{\varepsilon}^2)^2$, and $\mathcal{C}^n\vec{e}$ differs from $\vec{f}$ only by an amount of order $|\vec{\varepsilon}|^{2n}$. Consequently the convergence of the limit (7.17) to $\vec{f}$ is extremely fast. When applied to a typical problem in actual practice, one finds that the fixed point corresponding to an off-momentum closed orbit is given accurately to 1 part in 10^{10} by the time n = 4.[23]

Let p^o denote the design momentum, and p the momentum value of interest. Write

$$p = p^o(1 + \delta) \tag{7.23}$$

and give δ a small value. Also, let $\vec{f}(\delta)$ denote the fixed point (set of closed orbit initial conditions) corresponding to the momentum given by the relation (7.23). Examination of (7.18) and (7.20) shows that Newton's method succeeds as long as the matrix $I - M_{f(\delta)}$ has an inverse. Thus, as one gradually changes the momentum away from the design value, $\delta = 0$, the closed design orbit will continuously deform into the off-momentum closed orbit.

In order to complete the discussion, it is necessary to check whether the matrix $[I - M_{f(\delta)}]$ has an inverse. Evidently the inverse exists unless the related determinant satisfies the condition

$$\det[M_{f(\delta)} - I] = 0. \tag{7.24}$$

Equation (7.24) is equivalent to the condition that the Jacobian matrix $M_{f(\delta)}$ has eigenvalue +1.

Since $\mathcal{M}$ is a symplectic map, the Jacobian matrix $M_{f(\delta)}$ must be symplectic. However, according to section 2.3 and Fig. 2.2, all the eigenvalues of a symplectic matrix generally differ from one. Moreover, exercise 2.10 shows that if the eigenvalues of $M_{f(\delta)}$ differ from 1 for some value of δ, then the same will be true for nearby values of δ. Thus, if the eigenvalues of $M_{f(0)}$ for the design orbit are far away from +1, then the eigenvalues of $M_{f(\delta)}$ for the closed orbit associated with the initial condition $\vec{f}(\delta)$ will also differ from 1 for a range of δ values, and there will be one closed orbit for each value of δ. Finally, it will be shown in the next section that the eigenvalues of $M_{f(\delta)}$ for a closed orbit are related to the tunes for small betatron oscillations about this orbit. In particular, $M_{f(\delta)}$ can have eigenvalue +1 only if some tune has an integer value. Thus, the existence of closed orbits is assured for a range of momentum values provided the tunes of the design orbit are far from integer values.

A moment's reflection on the arguments just made for the general existence of an off-momentum closed orbit shows that similar arguments are applicable to other situations. The quantity δ simply appears as a parameter in the equations of motion, and what is being studied is the continued existence of a closed orbit as a parameter is continuously varied to a new value. Now, the difference between a perfect lattice and the perturbed lattice realized in actual construction may also be regarded as the result of a variation in certain parameters. Consequently, if the tunes for the closed design orbit in the perfect lattice have noninteger values, then the imperfect lattice will also have a closed orbit at the design momentum (and other nearby momenta as well) providing the perturbations in the lattice are not so large as to drive some tune to an integer value. In particular, small imperfections in a lattice, such as arise from magnet misalignment and misplacement, magnet under or over powering, magnetic fringe fields and general magnetic field inhomogenities, etc., do not destroy the existence of closed orbits but merely cause them to be slightly distorted.

Exercise 7.2: Verify Eq. (7.20).

7.2 Stability of closed orbits in the linear approximation

Let $\vec{f}(\delta)$ be the initial conditions in the Poincare surface of section for the closed orbit corresponding to the momentum deviation δ, and consider a nearby orbit having the same total momentum and the initial conditions $\vec{f} + \vec{\varepsilon}$ where $\vec{\varepsilon}$ is a small vector. Consider the action of $\mathcal{M}$ on the nearby orbit with initial conditions $\vec{f} + \vec{\varepsilon}$. Using (7.2), (7.3), and (7.4) one finds the result

$$\mathcal{M}(\vec{f} + \vec{\varepsilon}) = \vec{f} + M_f\vec{\varepsilon} + O(\vec{\varepsilon}^2). \tag{7.25}$$

This relation was already employed in verifying (7.20).

Similarly, the action of $\mathcal{M}^n$ is given by the relation

$$\mathcal{M}^n(\vec{f} + \vec{\varepsilon}) = \vec{f} + M_f^n\vec{\varepsilon} + O(\vec{\varepsilon}^2). \tag{7.26}$$

Thus, for nearby orbits, the long term behavior resulting from many lattice circuits is governed, in lowest approximation, by the matrix powers M_f^n.

But the behavior of large powers of a matrix is controlled, in turn, by the eigenvalue spectrum of the matrix. Moreover, as already seen earlier, M_f is a 4×4 symplectic matrix. Therefore, the various possibilities for the spectrum must be as illustrated in Fig. 2.2.

The behavior of M_f^n for large n is easily analyzed in each case. Suppose that some eigenvalue λ of M_f has a magnitude which exceeds one as in cases 1 through 4 of the generic configurations and cases 1, 2, and 6 of the degenerate configurations of Fig. 2.2. Then in these cases M_f^n grows exponentially without bound as n increases [essentially as $|\lambda|^n = \exp(n \log |\lambda|)$]. Correspondingly, the closed orbit associated with the fixed point $\vec{f}$ is unstable. That is, orbits with the same total momentum and initially nearby will, according to (7.26), deviate exponentially away from the closed orbit over the course of successive circuits around the lattice.

Next suppose that some eigenvalue λ of M_f has the value ± 1 while all others have absolute value one as in cases 3, 4, and 7 of the degenerate configurations. Then, as shown earlier in section 2.3, this eigenvalue must have even multiplicity. Consider, to be concrete, the case $\lambda = +1$, and suppose that M_f is brought to Jordan normal form by a similarity transformation S.[24] Then, in general the result will be a relation of the form

$$S^{-1} M_f S = \left(\begin{array}{cc|c} 1 & a & 0 \\ 0 & 1 & \\ \hline & 0 & \end{array}\right) \quad . \tag{7.27}$$

Here only the upper left-hand 2 × 2 block of (7.27) has been specified, and the quantity a denotes a number which is generally nonzero. See, for example, Eq. (6.53) for a 2 × 2 matrix of this type. Then when this 2 × 2 matrix is raised to the n'th power, as occurs in the calculation of M_f^n, one finds the result

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & na \\ 0 & 1 \end{pmatrix} \quad . \tag{7.28}$$

Thus, unless the quantity a vanishes, which is generally not the case, M_f^n grows linearly with n. Correspondingly, the closed orbit associated with the fixed point $\vec{f}$ is again unstable, although not as unstable as the exponential growth case of the previous paragraph. Note also that, under a small perturbation, an eigenvalue in this case can leave the unit circle through the point +1, and then the closed orbit becomes exponentially unstable.

Exercise 7.3: Carry out similar reasoning for the case $\lambda = -1$ to show that M_f^n again grows linearly with n, and hence the closed orbit associated with $\vec{f}$ is again linearly unstable. Also show that under a small perturbation an eigenvalue can leave the unit circle through the point -1, and the closed orbit then comes exponentially unstable.

Examination of Fig. 2.2 shows that there are still two possibilities which have not been discussed, namely case 5 of the generic configurations, and case 5 of the degenerate configurations. In the generic configuration of case 5, the matrix M_f can be diagonalized by a similarity transformation since all its eigenvalues are distinct. For this possibility, let λ_1 and λ_2 be two of the distinct eigenvalues which are not complex conjugates. Since they lie on the unit circle, they can be written in the form

$$\lambda_1 = e^{i\psi_1}, \ \lambda_2 = e^{i\psi_2} \tag{7.29}$$

where ψ_1 and ψ_2 are two real numbers. The remaining two distinct eigenvalues are then just the numbers $e^{-i\psi_1}$ and $e^{-i\psi_2}$. It follows that in this situation M_f^n merely oscillates with increasing n. That is, all matrix elements remain bounded. Correspondingly, the closed orbit associated with $\vec{f}$ must be <u>stable</u>. That is, to the accuracy of the linear approximation (7.26), nearby orbits exhibit bounded betatron oscillations about the closed orbit.

At this point it is possible to make another fundamental observation about imperfect lattices. Suppose a perfect lattice has been designed, all its closed orbits for a range of off-momentum values have been found, and all these closed orbits prove to be stable. What happens if this lattice is slightly perturbed? It has already been argued that the various closed orbits generally continue to exist and are only slightly deformed under perturbation. But do they remain stable? Fortunately, because not even the Swiss can build perfect lattices, the answer is yes thanks to the symplectic condition.

To understand this result, suppose the eigenvalues of M_f were originally all complex, of absolute value 1, and distinct. Then, according to exercise 2.10, this situation persists under a small perturbation. Indeed, for an eigenvalue to leave the unit circle, the eigenvalue spectrum must first pass through one of the degenerate configurations shown in part B of Fig. 2.2. Thus, from this perspective, the symplectic condition is really the key reason why it is possible to build workable accelerators and storage rings.

What about the last possibility, case 5 of the degenerate configurations? In this situation more needs to be known about the matrix M_f. If it can be diagonalized despite the fact that its eigenvalues are not distinct, then the same reasoning can be used as in the generic case, and the closed orbit is stable. However, suppose M_f cannot be diagonalized, but can only be brought to the Jordan form. Then, as in the other degenerate cases, the closed orbit is linearly unstable. To discover which of these possibilities holds, and the effect of small perturbations, requires considerably more discussion which is beyond the scope of these lectures.[25]

<u>Exercise 7.4</u>: Suppose that $\vec{f}$ is a fixed point of M and that M_f has a spectrum corresponding to one of the five <u>generic</u> configurations of Fig. 2.2. Show that in this case $\vec{f}$ must be an <u>isolated</u> fixed point. That is, there are no other fixed points near $\vec{f}$.

<u>Hint</u>: Assume that $\vec{g}$ is a nearby fixed point of M. Then $\vec{g} = \vec{f} + \vec{\varepsilon}$ where $\vec{\varepsilon}$ is small. Now show that $\vec{\varepsilon}$ must satisfy $M_f\vec{\varepsilon} = \vec{\varepsilon}$, and that this is impossible unless $\vec{\varepsilon} = 0$. For extra credit, show that if two or more fixed points are to coalesce as some parameter is varied, then the spectrum of M_f when they meet must be one of the degenerate configurations of Fig. 2.2 with an eigenvalue +1.

As intimated earlier, the eigenvalues λ_1 and λ_2 in the case of a stable closed orbit are related to the betatron oscillation tunes. Roughly speaking, the tunes of a closed stable orbit can be defined to be the number of vertical and horizontal oscillations made about the closed orbit by a nearby orbit during one lattice circuit. More precision in definition is possible for the case of the design orbit, if the vertical and horizontal degrees of freedom are uncoupled owing to the lattice having midplane symmetry, by the method of transfer matrices. In that case, tunes are related to the eigenvalues of the transfer matrix.

But a moment's reflection shows that the transfer matrix for the design orbit from the point θ=0 to the general point θ coincides with $B_{f(0)}(\theta)$. [Here $\vec{f}(0)$ denotes $\vec{f}(\delta)$ with δ=0.] Now consider the general off-momentum closed orbit. Assume that M is the Poincare map for a complete circuit about the lattice so that M_f and B_f for the general fixed point $\vec{f}(\delta)$ are related by the equation

$$M_f = B_f(2\pi). \tag{7.30}$$

It follows from a general Floquet analysis, in analogy with the method of transfer matrices, that the tunes T_1, T_2 of a general closed orbit (even in the absence of midplane symmetry for the lattice under consideration) can be defined to be the numbers

$$\begin{aligned} T_1 &= \pm(\psi_1/2\pi) + \text{some integer} \\ T_2 &= \pm(\psi_2/2\pi) + \text{some integer.} \end{aligned} \tag{7.31}$$

Here the quantities ψ_i are related to the eigenvalues of M_f by (7.29).

Note that if +1 is an eigenvalue of M_f, then according to (7.29) and (7.31) at least one tune must be an integer. As indicated earlier, the proof of the existence and uniqueness of off-momentum closed orbits breaks down under these conditions. Moreover, according to the previous orbit stability discussion, if a closed orbit does exist when $\lambda = 1$, it is generally linearly unstable and subject to becoming exponentially unstable under small perturbation. Note also that when a tune is half integer, some eigenvalue of M_f has the value -1. Then the orbit is again generally linearly unstable and again subject to becoming exponentially unstable under a small perturbation. Consequently, integer and half integer tunes are generally to be avoided unless, perhaps, one is trying to exploit orbit instability combined with suitable nonlinear effects to achieve extraction.

Equation (7.31) states that tunes, in the general case, are only defined modulo an integer. This should not be surprising. It is not possible to detect how many whole oscillations are made simply by observing the orbit at one surface of section. Moreover, if the horizontal and vertical degrees of freedom are strongly coupled, it is not always possible to define what is meant by a whole oscillation even if the orbit is observed everywhere.

In the case that the lattice has N identical periods, the integer ambiguity can be resolved providing it is assumed, as is usually the case, that the number of betatron oscillations made in the passage through a single lattice period is less than 1. (This is equivalent to assuming that the phase advance per lattice period is less than 2π in the case that horizontal and vertical oscillations are uncoupled.) Suppose that we now take $\mathcal{M}$ to be the Poincare map for one lattice period so that M_f and B_f are related by the equation[26]

$$M_f = B_f(2\pi/N). \tag{7.32}$$

Again, as before, let λ_1 and λ_2 be two eigenvalues of M_f written in the form (7.29) with ψ_1, ψ_2 distinct and lying in the interval $(0,2\pi)$. In this case the numbers ψ_1, ψ_2 can be regarded as the phase advances per period, and the overall lattice tunes can be defined to be the numbers

$$\begin{aligned} T_1 &= N(\psi_1/2\pi) \\ T_2 &= N(\psi_2/2\pi). \end{aligned} \tag{7.33}$$

By contrast, the ψ values appearing in (7.31) are essentially phase advances, modulo signs and an unspecified number of 2π's, for an entire circuit of the whole lattice. Bearing this distinction in mind, it is easily verified that the definition (7.33) coincides with one of the set of values given by (7.31).

<u>Exercise 7.5</u>: In the case that the horizontal and vertical degrees of freedom of a lattice are uncoupled, the matrix M_f takes the reduced block form

$$M_f = \begin{pmatrix} \text{hor} & 0 \\ 0 & \text{ver} \end{pmatrix} \tag{7.34}$$

where the abbreviations "hor" and "ver" refer to 2×2 matrices which separately describe the horizontal and vertical motions. In this case, show that the phase advances ψ are given by the simple formulas

$$\psi_1 = \cos^{-1} [\,(1/2)\ \mathrm{tr}\ (\text{hor})\,] \tag{7.35a}$$

$$\psi_2 = \cos^{-1} [\,(1/2)\ \mathrm{tr}\ (\text{ver})\,]. \tag{7.35b}$$

In the more general case in which the horizontal and vertical degrees of freedom are coupled, it might be thought that a general diagonalization procedure would be required to find the eigenvalues of M_f. However show, using the results of exercise 2.11, that the two phase advances, now denoted by $\psi_\pm$, can be found directly from the relations

$$\psi_\pm = \cos^{-1} [-b \pm (b^2 - c)^{1/2}\,]\,. \tag{7.36}$$

In the discussion so far, it has been assumed that the total momentum p remains constant along an orbit. It is also possible to treat more general situations. Consider, for example, the case of a proton storage ring with one or several radio frequency bunching elements. Then in this case there is one design orbit with constant total momentum, and other orbits which exhibit synchrotron oscillations. To describe these various orbits, it is necessary to introduce the additional variables t and p_t as well as the previously used two coordinates and two momenta transverse to the beam. Indeed, it is useful to employ, as in Section 6, the time and conjugate momentum deviation variables T and P_T. Then the Poincare surface of section becomes a full six dimensional phase space, and the Poincare mapping $\mathcal{M}$ maps this phase space into itself.

As before, a fixed point $\vec{f}$ of $\mathcal{M}$ corresponds to a closed orbit. In particular, for a perfect machine there is a fixed point corresponding to the synchronous on-momentum design orbit, and the T and P_T components of $\vec{f}$ satisfy the conditions $T = 0$ and $P_T = 0$.

Now consider M_f, the linear part of $\mathcal{M}$ at $\vec{f}$. The matrix M_f will be 6×6 and symplectic. Suppose all its eigenvalues are distinct and lie on the unit circle. This is the 6×6 analog of case 5 of the generic configurations. (See Fig. 2.2). Then in this case the design orbit will be stable, and all other nearby orbits, including off-momentum and off-synchronous orbits, will make synchro-betatron oscillations (possibly coupled) about the design orbit. Of course, now there will be three tunes, and one of them will be quite small because synchrotron oscillation frequencies are usually much lower than betatron oscillation frequencies. Finally one can show, by arguments similar to those made before, that the fixed point $\vec{f}$ of $\mathcal{M}$ is isolated, and continues to exist and to remain stable under small perturbations of the lattice including variations in the strength and frequency of the bunching fields. These are all again benefits of the symplectic condition.

The preceeding discussion has shown that a fixed point is stable if the eigenvalues of M_f are distinct and lie on the unit circle. Otherwise, the fixed point is generally unstable. Still more can be said about the fixed points of two dimensional symplectic mappings. This restricted case is of interest in its own right, and is directly applicable to accelerators in the approximation that the various degrees of freedom are uncoupled. Moreover, the restricted case suggests results that may have useful analogs in the general case of four and six dimensional symplectic mappings.[27] The series of exercises below develops what is known about the two dimensional case, and explains the significance of the names hyperbolic, elliptic, etc. that occur in Fig. (2.1).

<u>Exercise 7.6</u>: Suppose $\mathcal{M}$ is a symplectic map in two dimensions with a fixed point f. Without loss of generality, the fixed point may be taken to be located at the origin. Let M denote the linear part of $\mathcal{M}$ at the fixed point. Suppose further that the spectrum of M corresponds to case 3 of Fig. (2.1). Show that M then sends a certain set of nested ellipses around the origin into themselves. Its only action is to advance or "rotate" points along the ellipses. (The rate of advance is related to the tune of the associated closed orbit.) For this reason, if the spectrum of M is as in case 3, the fixed point is said to be <u>elliptic</u>. It follows from (7.26) that, in the linear approximation, points initially very near the origin remain near the origin under repeated application of $\mathcal{M}$. (Each point must remain on its particular ellipse.) For this reason, an elliptic fixed point is said to be <u>stable</u>.

<u>Solution</u>: Suppose the eigenvalues of M are $\exp(\pm iw)$. Then $\mathrm{tr}\, M = 2 \cos w$. Introduce Twiss parameters α, β, γ by writing without loss of generality

$$M = \begin{pmatrix} \cos w + \alpha \sin w & \beta \sin w \\ -\gamma \sin w & \cos w - \alpha \sin w \end{pmatrix}. \qquad (7.37)$$

The symplectic condition that det M=1 requires the relation

$$\beta\gamma = 1 + \alpha^2. \tag{7.38}$$

Next observe that M can be expressed in the form

$$M = I \cos w + K \sin w \tag{7.39}$$

where K is the matrix

$$K = \begin{pmatrix} \alpha & \beta \\ -\gamma & -\alpha \end{pmatrix} \quad . \tag{7.40}$$

Moreover, thanks to (7.38), K has the property

$$K^2 = -I. \tag{7.41}$$

Consequently, M can also be written in the exponential form

$$M = \exp(wK). \tag{7.42}$$

Finally, K can be written in the form

$$K = J\ S \tag{7.43}$$

where S is the symmetric matrix

$$S = \begin{pmatrix} \gamma & \alpha \\ \alpha & \beta \end{pmatrix} \quad . \tag{7.44}$$

Now introduce a polynomial g_2 by the definition

$$\begin{aligned} g_2(z) &= -(1/2) \sum_{ij} S_{ij}\ z_i\ z_j \\ &= -(1/2)(\gamma q^2 + 2\alpha qp + \beta p^2). \end{aligned} \tag{7.45}$$

It follows from section 5.1 that M satisfies the relation

$$\exp(w{:}g_2{:})z = Mz. \tag{7.46}$$

Therefore g_2 is an invariant function. That is, g_2 has the property

$$g_2(Mz) = g_2(z). \tag{7.47}$$

Here, as usual, z stands for the complete collection of phase-space variables. See Eqs. (5.5) through (5.7) for a review of why g_2 should have the invariance property (7.47). Indeed, apart from the factor (-1/2), g_2 is the Courant-Snyder invariant.[28] Note that the discrimanant of the quadratic form (7.45), namely the quantity $(2\alpha)^2 - 4\beta\gamma$, has the negative value -4 thanks to Eq. (7.38). Consequently, the level lines of g_2 are indeed nested ellipses.

<u>Exercise 7.7</u>: Repeat exercise 7.6 under the supposition that the spectrum of M corresponds to case 1 of Fig. (2.1). Show that M then sends a certain set of hyperbolae into themselves. For this reason, if the spectrum of M is as in case 1, the fixed point is said to be hyperbolic. Show also that if a point is initially very near the fixed point, then in the linear approximation it will in general eventually be moved away from the fixed point under repeated action of M. Therefore, a hyperbolic fixed point is said to be <u>unstable</u>.

<u>Hint</u>: Suppose the eigenvalues of M are written in the form $\exp(\pm w)$. Then tr M = 2 cosh w. In this case, introduce parameters α, β, γ by writing

$$M = \begin{pmatrix} \cosh w + \alpha \sinh w & \beta \sinh w \\ -\gamma \sinh w & \cosh w - \alpha \sinh w \end{pmatrix} . \tag{7.48}$$

Show that in this case

$$\beta\gamma = \alpha^2 - 1. \tag{7.49}$$

Also show that M can be expressed in the form

$$M = I \cosh w + K \sinh w \tag{7.50}$$

with K as before. Next show that K now satisfies the relation

$$K^2 = +I. \tag{7.51}$$

Now show that the rest of the exercise goes through as before except that the discrimant of g_2 now has the positive value +4, and therefore level lines of g_2 are hyperbolae. Note that the level line $g_2 = 0$ corresponds to the fixed point itself and the asymptotes of the hyperbolae. Check that the asymptotes are given by the equations

$$q = -p\beta/(\alpha \pm 1). \tag{7.52}$$

Also verify that the eigenvectors of M lie along the asymptotes. Consider the asymptote corresponding to the eigenvector with eigenvalue greater than 1. Show that the action of M is to move points on this asymptote away from the origin, and to move points on the other asymptote toward the origin. For this reason, these two asymptotes are called the unstable and stable manifolds respectively. Show that the action of M on any given hyperbola is to move points along the hyperbola. See Fig. (7.1) which illustrates the action of M on a general point in the hyperbolic case.

Exercise 7.8: Repeat exercise 7.6 under the supposition that the spectrum of M corresponds to case 4 of Fig. (2.1). Show that M then sends a family of straight lines into themselves. In this case the fixed point is said to be parabolic, although the invariant curves are straight lines and not parabolas. Note that the parabolic case is a transitional case between the elliptic and hyperbolic cases. It is called parabolic out of deference to the convention that the word "parabolic" is used in other contexts to describe a circumstance intermediate to an elliptic or hyperbolic case. Show that the invariant line through the origin consists entirely of fixed points. (A parabolic fixed point is not isolated in the linear approximation. See exercise 7.4.) Also show that the action of M on each line is that of a linear displacement. For this reason, a parabolic fixed point is said to be linearly unstable.

Hint: Since M has +1 as an eigenvalue with multiplicity two, it must satisfy tr M = 2. Suppose M is parameterized by writing it in the form

$$M = \begin{pmatrix} 1 + \alpha & \beta \\ -\gamma & 1 - \alpha \end{pmatrix}. \tag{7.53}$$

Then the condition det M = 1 entails the restriction

$$\alpha^2 = \beta\gamma. \tag{7.54}$$

Next define a matrix K as in (7.46). Show that K satisfies the relation

$$K^2 = 0, \tag{7.55}$$

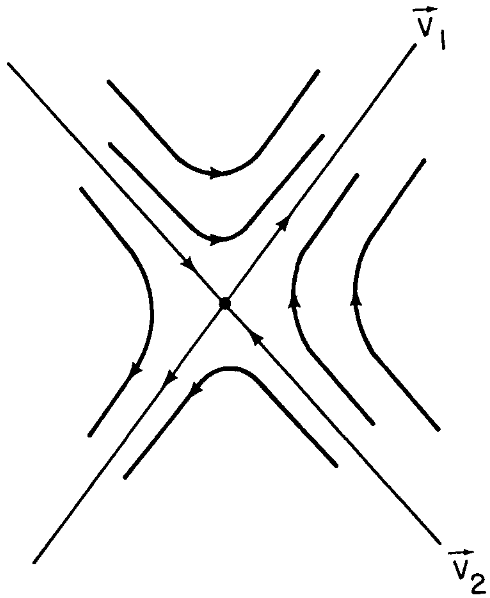


Figure (7.1): The action of M on points near a hyperbolic fixed point. Suppose that $\vec{v}_1$ and $\vec{v}_2$ are eigenvectors of M having eigenvalues greater and less than one respectively. Then points on $\vec{v}_1$ are moved outward, those on $\vec{v}_2$ are moved inward, and others are moved on hyperbolas.

and that M can be written in the form

$$M = \exp(K) . \tag{7.56}$$

Now proceed as before. Then show that (7.54) implies that g_2 can be written as a perfect square in the form

$$g_2(z) = -(1/2)(\gamma^{1/2}q + \beta^{1/2}p)^2 . \tag{7.57}$$

Let $\vec{u}$ be the vector with entries

$$\vec{u} = \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} .$$

Show that $\vec{u}$ has eigenvalue +1. Next show that the line

$$g_2(z) = g_2(0) = 0 \tag{7.58}$$

is along $\vec{u}$. Therefore, all points on this line will remain fixed under the action of M. Let $\vec{v}$ be the vector with entries

$$\vec{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} .$$

Verify that $\vec{v}$ is orthogonal to $\vec{u}$. Suppose $\vec{w}$ is a general point in phase space. Since $\vec{u}$ and $\vec{v}$ are orthogonal, $\vec{w}$ can be expanded in terms of $\vec{u}$ and $\vec{v}$ to give a relation of the form

$$\vec{w} = a\vec{u} + b\vec{v}$$

where <u>a</u> and <u>b</u> are expansion coefficients. Verify that the action of M on the general point $\vec{w}$ is given by the relation

$$M\cdot\vec{w} = [\, a + b(\beta+\gamma)]\ \vec{u} + b\vec{v} = \vec{w} + b(\beta+\gamma)\ \vec{u}.$$

Thus, the action of M on a general point, in the parabolic case, is as illustrated in Fig. (7.2).

<u>Exercise 7.9</u>: Suppose the spectrum of M corresponds to the inversion hyperbolic or inversion parabolic cases of Fig. (3.1). Show that M can then be written as a product in the form

$$M = (-I)\ N \tag{7.59}$$

where N is symplectic and either hyperbolic or parabolic, respectively. Thus, M is a product of a hyperbolic or parabolic symplectic matrix followed by <u>inversion</u> through the origin. Let g_2 be the invariant function for N as constructed in exercises 7.7 or 7.8. For a fixed value, g_2 generally has two branches since it is an even function. The action of N is to

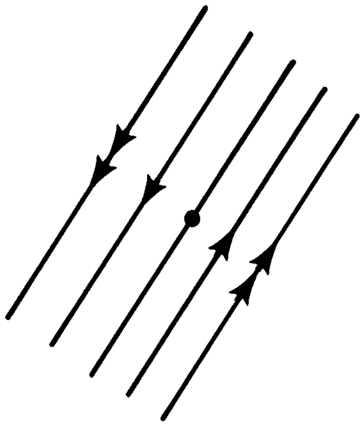


Figure (7.2): The action of M on points near a parabolic fixed point. The action is such as to produce a "sheared" flow along straight lines. The displacement along a line is proportional to the distance of that line from the fixed point.

move points along on a given branch, and the action of (-I) is to jump between corresponding branches. Consequently, the action of M^n for increasing n is to send a point jumping back and forth between two branches. Observe that M^2 is either hyperbolic or parabolic, respectively. Therefore, after two jumps, a point is always back on the same branch, and the net effect is motion along that branch just as in the hyperbolic or parabolic cases.

7.3 Stability of closed orbits including nonlinear effects

Suppose $\mathcal{M}$ is a symplectic map with fixed point $\vec{f}$. Without loss of generality, one can select a new canonical coordinate system in such a way that $\vec{f}$ is located at the origin. With this choice of coordinates, $\mathcal{M}$ becomes a symplectic map which sends the origin into itself.

Since $\mathcal{M}$ is a symplectic map which sends the origin into itself, it follows from the factorization theorem 5.1 that there exist homogeneous polynomials g_2, g_3, g_4, etc. such that $\mathcal{M}$ can be written in the form

$$\mathcal{M} = \exp(:g_2:)\ \exp(:g_3:)\ \exp(:g_4:) \cdots . \qquad (7.60)$$

(Here, in order to avoid confusion with the symbol f employed to refer to the fixed point $\vec{f}$, the letters g_n rather than f_n will be used to denote polynomials.)

From this perspective, the linear stability analysis made for closed orbits in the previous section is equivalent to a study of the "Gaussian" piece of $\mathcal{M}$, namely $\exp(:g_2:)$. Indeed, if the variables in the surface of section on which $\mathcal{M}$ acts are denoted by the letters z_1 z_{2n}, then one has the relation

$$\exp(:g_2:)\ z_i = \sum_j (M_f)_{ij}\ z_j . \qquad (7.61)$$

The linear stability analysis neglected the effect of the factors $\exp(:g_3:)\ \exp(:g_4:)\cdots$. It showed that closed orbits are stable, in lowest approximation, if the eigenvalues of M_f are distinct and all lie on the unit circle. The purpose of this section is to examine the effect of the higher order terms.

As stated earlier, the long-term behavior of orbits for a large number of turns is equivalent to a knowledge of $\mathcal{M}^n$ for large n. Suppose that $\mathcal{M}$ could be written as the exponential of a single Lie operator :h: in the form

$$\mathcal{M} = \exp(:h:). \qquad (7.62)$$

Then the evaluation of $\mathcal{M}^n$ would be simple, for one would simply have the result

$$\mathcal{M}^n = \exp(n:h:). \qquad (7.63)$$

But, thanks to remarkable pedagogical foresight, formulas for combining exponents into a single exponent were developed in section 5.4. To see how they may be applied, consider as a simple example the case where all terms in (7.60) beyond $\exp(:g_3:)$ are neglected. That is, suppose $\mathcal{M}$ is approximated by the truncated symplectic map

$$\mathcal{M} = \exp(:g_2:)\, \exp(:g_3:). \tag{7.64}$$

Then $:g_2:$ and $:g_3:$ can be combined using Eqs. (5.53), (5.81), and (5.83). One finds the result

$$\begin{aligned} h &= g_2 + :g_2:[1 - \exp(-:g_2:)]^{-1}\, g_3 \\ &\quad + \text{terms of degree 4 and higher.} \end{aligned} \tag{7.65}$$

Note that Eqs. (5.81) and (5.83) omit terms of degree t^2 and higher. In the present context, this omission is equivalent to neglecting all commutators in an expression of the form (5.53) which contain two or more g_3's and any number of g_2's. However, according to exercise 5.6, these neglected terms are of degree 4 and higher as indicated in (7.65). It should also be remarked that these omitted terms can in principle be computed if needed. In particular, the term of degree t^2 is known.[29] Thus, h can be computed through degree 4 with existing tools. If such a calculation were to be made, then for consistency the contribution of degree four from the g_4 term in (7.60), which is easily computed, should be included as well.

In order to continue with the calculation for h as given by (7.65), it is necessary to specify more about g_2 or, equivalently, M_f. Consider for simplicity the case of a coasting proton beam. That is, all accelerating sections and bunchers have been turned off. Suppose further that all particles in the beam have the design momentum. Then attention can be restricted to the two transverse positions and momenta. Moreover, if the Poincare surface is indeed located in some straight section as has been assumed, then these transverse coordinates can be conveniently taken to be the Cartesian quantities y, z, p_y, p_z at some fixed location. Finally, assume that the horizontal and vertical betatron oscillations are uncoupled in the linear approximation, and that the design orbit is stable with tunes $T_y = (w_y)/(2\pi)$ and $T_z = (w_z)/(2\pi)$.[30] Then, without loss of generality, the polynomial g_2 can be taken to be the expression[31]

$$g_2 = -w_y(p_y^2 + y^2)/2 - w_z(p_z^2 + z^2)/2. \tag{7.66}$$

It is easily verified that $\exp(:g_2:)$ has the desired action

$$\exp(:g_2:)\begin{bmatrix} y \\ p_y \\ z \\ p_z \end{bmatrix} = \begin{bmatrix} \cos(w_y) & \sin(w_y) & 0 & 0 \\ -\sin(w_y) & \cos(w_y) & 0 & 0 \\ 0 & 0 & \cos(w_z) & \sin(w_z) \\ 0 & 0 & -\sin(w_z) & \cos(w_z) \end{bmatrix} \begin{bmatrix} y \\ p_y \\ z \\ p_z \end{bmatrix}. \tag{7.67}$$

<u>Exercise 7.10</u>: Verify (7.67) and convince yourself that the quantities $T_y = (w_y)/(2\pi)$ and $T_z = (w_z)/(2\pi)$ are tunes.

As indicated in (5.82), the evaluation of the operator expression in (7.65) requires the summation of an infinite series. In order to carry out the summation, it is convenient to expand the quantity g_3 in terms of eigenfunctions of the operator $:g_2:$. At first sight, it may not be obvious that $:g_2:$ should have eigenfunctions. However, suppose that g_n is a homogeneous polynomial of degree n. Then, according to exercise 5.6, the quantity $:g_2:g_n$ is also a homogeneous polynomial of degree n. Therefore, $:g_2:$ may be viewed as a <u>linear operator</u> which maps the vector space of homogeneous polynomials of degree n into itself. It follows from simple matrix theory that $:g_2:$ must have at least one eigenvector.

Indeed, it is easy by explicit construction to see that $:g_2:$ has a complete set of eigenfunctions. For ease of exposition, consider first the simpler case of one degree of freedom with canonical coordinates q,p. Also, for the moment, let g_2 denote the polynomial

$$g_2 = -w(p^2 + q^2)/2. \tag{7.68}$$

Now replace the variables q,p by the action-angle variables a,ϕ defined by the relation

$$q = (2a)^{1/2} \sin \phi \tag{7.69a}$$

$$p = (2a)^{1/2} \cos \phi. \tag{7.69b}$$

Then it is easily verified that

$$2a = q^2 + p^2 = -2g_2/w . \tag{7.70}$$

It follows that

$$:g_2:a = 0. \tag{7.71}$$

Consider the quantity $(p + iq)^m$. One finds the relation

$$(p + iq)^m = [(2a)^{1/2} (\cos\phi + i \sin\phi)]^m$$
$$= (2a)^{m/2} e^{im\phi} \quad . \tag{7.72}$$

Also compute the quantity $[\exp(\tau{:}g_2{:})](p + iq)$ where τ is some parameter. The result is easily found to be the relation

$$[\exp(\tau{:}g_2{:})](p + iq) = e^{i\tau w} (p + iq). \tag{7.73}$$

Next, using the isomorphism property (3.35), one finds the general relation

$$[\exp(\tau{:}g_2{:})](p + iq)^m = \{[\exp(\tau{:}g_2{:})](p + iq)\}^m$$
$$= e^{im\tau w}(p + iq)^m. \tag{7.74}$$

Now differentiate both sides of (7.74) with respect to the parameter τ, and then set $\tau = 0$. The result is the relation

$${:}g_2{:}(p + iq)^m = imw(p + iq)^m$$

or

$${:}g_2{:}[(2a)^{m/2}e^{im\phi}] = imw(2a)^{m/2}e^{im\phi} \quad . \tag{7.75}$$

Finally, in view of (3.34) and (7.71), the factor $(2a)^{m/2}$ may be removed from both sides of (7.75) to give the result

$${:}g_2{:}e^{im\phi} = imwe^{im\phi} \quad . \tag{7.76}$$

It follows that the eigenfunctions of ${:}g_2{:}$ are the functions $e^{im\phi}$ with the eigenvalues imw.

Evidently, an equivalent result holds in the two-dimensional case. Introduce two pairs of action-angle variables by the relations

$$y = (2a_y)^{1/2} \sin\phi_y \tag{7.77a}$$

$$p_y = (2a_y)^{1/2} \cos\phi_y \tag{7.77b}$$

$$z = (2a_z)^{1/2} \sin\phi_z \tag{7.77c}$$

$$p_z = (2a_z)^{1/2} \cos\phi_z. \tag{7.77d}$$

Then it follows that the eigenfunctions of the full g_2, now again given by (7.66), are the functions $e^{im\phi_y} e^{in\phi_z}$ with the eigenvalue relations

$$:g_2: e^{im\phi_y} e^{in\phi_z} = i(mw_y + nw_z) e^{im\phi_y} e^{in\phi_z} \quad . \tag{7.78}$$

<u>Exercise 7.11</u>: Verify Eqs. (7.73), (7.75), (7.76), and (7.78).

With this background, we are ready to proceed. From (7.77) it follows that g_3 has an expansion of the form

$$g_3 = \sum_{m,n} c_{mn}(a_y, a_z)\, e^{im\phi_y} e^{in\phi_z} \tag{7.79}$$

where, as indicated, the coefficients c_{mn} depend on the variables a_y, a_z. Using this expansion, it is easy to evaluate the operator expression in (7.65). One finds the result

$$:g_2:[1 - \exp(-:g_2:)]^{-1}\, g_3 = \sum_{m,n} c_{mn}\, e^{im\phi_y} e^{in\phi_z} \times$$

$$i(mw_y + nw_z)/[1 - \exp(-imw_y - inw_z)]. \tag{7.80}$$

Consequently, after a little trigonometric tidying up and neglecting the terms of degree 4 and higher, the expression (7.65) can be written in the explicit form

$$h = g_2 + \sum_{m,n} c_{mn}\, e^{im(\phi_y + w_y/2)}\, e^{in(\phi_z + w_z/2)} \times$$

$$(mw_y/2 + nw_z/2)/[\sin(mw_y/2 + nw_z/2)]. \tag{7.81}$$

<u>Exercise 7.12</u>: Verify Eqs. (7.80) and (7.81).

Examination of (7.81) shows that the denominator in the summation vanishes if the argument of the sine function vanishes or equals an integral multiple of π. That is, there are possible difficulties in computing h when there is a relation involving integers of the form

$$mw_y/2 + nw_z/2 = \ell\pi$$

or

$$m(w_y)/(2\pi) + n(w_z)/(2\pi) = \ell \quad . \tag{7.82}$$

Here the quantities ℓ, m, and n can in principle take all possible positive and negative integer values including zero. Equation (7.82) is the usual condition for nonlinear structure resonances.[32] It indicates that, in principle, difficulties can occur whenever a tune has a rational value, or two tunes are relatively rational. In actual practice, the summation range of m and n in (7.81) is limited by the degree of g. For example, in (7.79) and hence in (7.81), the summations over m and n are restricted to lie between +3 and -3. Similarly, for a g_4, they would be restricted to lie between +4 and -4. Thus, if one assumes that the higher degree g's are much smaller or less important than the lower order g's, then the lower order resonances dominate.

The complete details of nonlinear structure resonances have yet to be worked out from a Lie algebraic perspective. However, some work has been done on the simpler one-dimensional case in which g_2 is given by (7.68) and g_3 has the form

$$g_3 = sq^3/3. \tag{7.83}$$

This may be viewed as the case of a perfect ring with a short strong sextupole insertion, provided attention is restricted to horizontal betatron oscillations and all orbits are assumed to lie in the midplane.[33]

Indeed, Eq. (6.47) takes the form (7.83) in the midplane z = 0 when due account is taken of the change of notation. Here s is a measure of the sextupole strength integrated over its length.

<u>Exercise 7.13</u>: Suppose $\mathcal{M}$ is given by (7.64) with g_2 and g_3 given by (7.68) and (7.83) respectively. Find the action of $\mathcal{M}$ on the general point (q,p).

<u>Answer</u>:

$$\begin{aligned} \bar{q} = \mathcal{M}q &= q\cos w + p\sin w \\ \bar{p} = \mathcal{M}p &= -q\sin w + p\cos w + s(q\cos w + p\sin w)^2. \end{aligned} \tag{7.84}$$

When (7.69) is inserted into (7.83), one finds the expansion

$$\begin{aligned} g_3 &= (s/3)(2a)^{3/2}(\sin\phi)^3 = \sum_m c_m(a)e^{im\phi} \\ &= (s/3)(2a)^{3/2}(e^{3i\phi} - 3e^{i\phi} + 3e^{-i\phi} - e^{-3i\phi})/(2i)^3. \end{aligned} \tag{7.85}$$

Thus, as expected, only a few of the coefficients c_m are nonzero. In analogy to (7.81), one now has the relation

$$h = g_2 + \sum_m c_m e^{im(\phi+w/2)}(mw/2)/[\sin(mw/2)]. \qquad (7.86)$$

Finally, when the explicit coefficients as given by (7.85) are inserted into (7.86) and use is made of (7.70), one finds the result

$$h = -wa - (ws/8)\,(2a)^{3/2}\{[\sin(3\phi+3w/2)]/[\sin(3w/2)]$$

$$- [\sin(\phi+w/2)]/[\sin(w/2)]\}. \qquad (7.87)$$

<u>Exercise 7.14</u>: Verify Eqs. (7.85), (7.86), and (7.87).

Inspection of (7.87) shows that resonances occur whenever the quantities (3w/2) or (w/2) are multiples of π. That is, the resonance conditions are

$$T = w/(2\pi) = \ell/3 \text{ or } T = w/(2\pi) = \ell. \qquad (7.88)$$

Consequently, for a short sextupole insertion, there are resonances when the tune T is an integer multiple of (1/3), and at an integer tune values.

The behavior of the symplectic map $\mathcal{M}$ at and near these resonant values will be examined shortly. Suppose for the moment that w is far from these resonant values. Then the expression for h is well behaved. Moreover, since $\mathcal{M}$ can be written in the form (7.62), it follows that h is an <u>invariant function</u> for the mapping $\mathcal{M}$. That is, denoting the general action of $\mathcal{M}$ as in (5.2), one has the relation

$$h(\mathcal{M}z) = h(z). \qquad (7.89)$$

Here, as usual, z stands for the complete collection of phase-space variables. See Eqs. (5.5) through (5.7) for a review of why h should have the invariance property (7.89).

Equation (7.89) says that the value of h should be the same at all those points in the surface of section which are the image of a given point under successive applications of the map $\mathcal{M}$. Or, put another way, the mapping $\mathcal{M}$ must send each level line of the function h into itself. This condition may place great restrictions on the action of $\mathcal{M}$. For example, if h has a level line which is a simple closed curve, then $\mathcal{M}$ can never map the interior of this curve to the exterior because a continuous mapping, which $\mathcal{M}$ is, must preserve topological properties. Thus, if there is indeed an invariant function with a closed curve level line, then orbits having initial conditions in the surface of section which are initially inside this curve must always intersect the surface of section at points inside the curve for all future times. Therefore such orbits are <u>stable</u> because they are bounded for all time.

Indeed, even more can be said. Comparison of (7.62) and (5.34) shows that the function (-h) can be viewed as playing the role of an "effective Hamiltonian" for the Poincare map M. That is, suppose (-h) is viewed as a Hamiltonian defined on the surface of section. Then following the Hamiltonian flow specified by (-h) for n units of time from some initial point is equivalent to computing the action of M^n on the initial point.

As an example of how this works out in practice, look at Fig. 7.3. The points on the inner curve are the images of the point

$$(q_o,p_o) = (0.5,0.0) \tag{7.90}$$

under successive applications of M.

That is, they are the points

$$(q_n,p_n) = M^n(q_o,p_o) \tag{7.91}$$

where M is the symplectic map given explicitly by (7.84). The tune angle is set at the nonresonant value w=5 degrees, which corresponds to the tune T having the value T = .0139, and the sextupole has the value s = .05 for its strength.

Similarly, the points on the outer curve are the images of an initial point on the q axis with a larger radius.

For comparison, the circles on the inner curve are centered on the points obtained by following (by numerical integration) the Hamiltonian flow specified by (-h) for successive units of time starting with the initial conditions (7.90). Here h is the quantity given by (7.87). The actual points themselves are not shown, but only the open circles surrounding them. Similarly, the circles on the outer curve are centered about points, again not shown, which are the result of following the flow specified by (-h) starting with, as initial conditions, the point on the q axis having a larger radius.

Evidently, the circles on the inner curve appear to surround the points (7.91). Similarly, the circles on the outer curve appear to surround the points on the outer curve. The close agreement of the points and circles illustrates the close agreement between M^n and exp(n:h:) for the case of a tune far away from resonant values.

In the absence of the sextupole term (i.e. when s=0), the mapping M would map the nested circles $q^2 + p^2 = 2a =$ constant into themselves. This fact can easily be seen from (7.84). That is, the quantity g_2 given by (7.70) is an invariant of M in lowest approximation. According to exercise (7.6), this quantity is the Courant-Snyder invariant for betatron oscillations.

Apparently, in the nonresonant case, the only effect of a sextupole of moderate strength is to produce an "egg shaped" distortion of the invariant curves. Moreover, the shape of the distorted curves is well described by the level lines h = constant. Consequently, h provides a generalization of the Courant-Snyder invariant to the case of nonlinear motion.

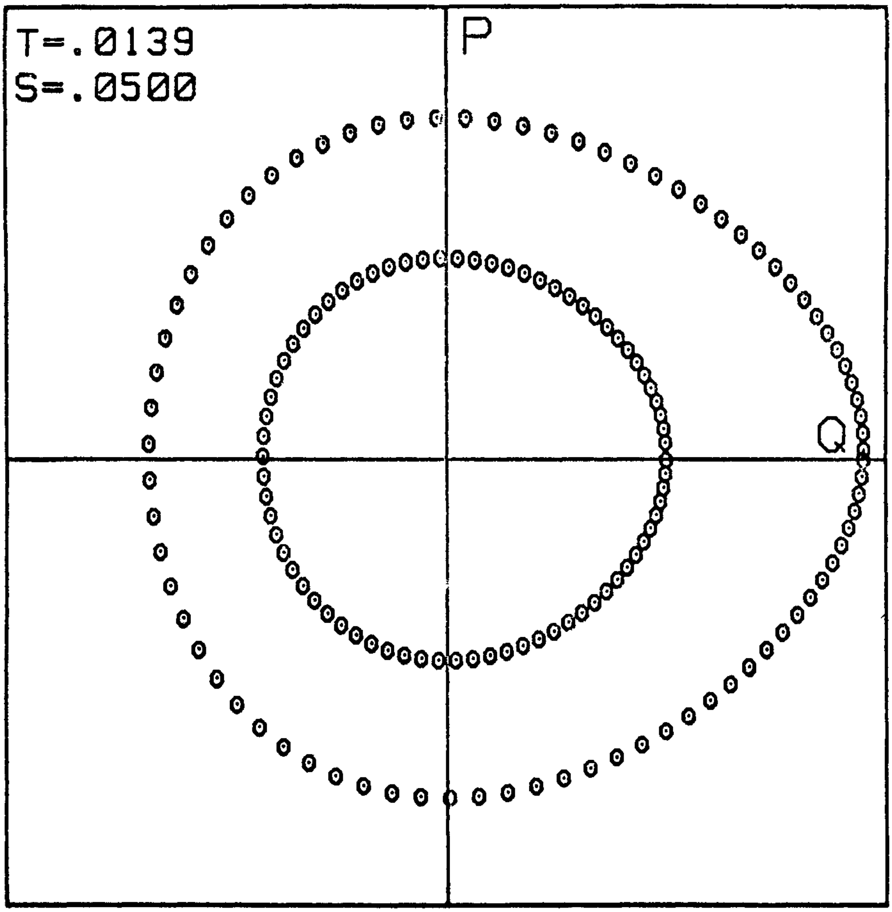


Figure (7.3): Typical behavior when the tune is far from resonance. Orbits are stable, and iterates of appear to agree perfectly with trajectories derived from h. The tune is T = .0139, and the sextupole strength is s=.05. In the viewing area the coordinates q and p range from -1 to 1.

The fact that the invariant curves are now "egg shaped" rather than circular means that the amplitude of betatron oscillation is not constant, but instead varies between certain minimum and maximum values. The extent of this variation is sometimes referred to as the <u>beating range</u>.[34] Also, it can be shown that the rate of progress of successive points along a given curve, which is a measure of the tune of an orbit, depends on which curve is considered. This circumstance may be described roughly by saying that, in the nonlinear case, the tune of an orbit depends on the betatron amplitude. Evidently, complete information about both these nonlinear phenomena is given by h.

Figure 7.4 shows another example. In this case, the sextupole strength is as before, but the tune angle has the almost-zero value w=1 degree. Again, there is good agreement between points, which are the result of computing the action of $\mathcal{M}^n$ on various initial points, and circles which are the result of following the Hamiltonian flow generated by (-h). Only a slight departure of the circles from the points is visible in the far upper right-hand corner of the figure. Presumably this departure arises from the omitted terms in (7.65). Note that g_3 as given by (7.83) is proportional to the quantity s which is small, and that multiple commutators involving several factors of g_3 will contain higher powers of s. It follows that the omitted terms are not only of degree 4 and higher as indicated, but also in this case are proportional to s^2 and higher powers of s. Therefore, one should expect very good agreement near the origin where the higher degree terms are small, and moderately good agreement away from the origin where the contributions of the higher degree terms are still diminished by the higher powers of s.

Observe that in this case the map has two fixed points and therefore there are two closed orbits. One of the fixed points is at the origin, as expected, and the second is located somewhere near the q axis as indicated in Fig. 7.4.

To the extent that (-h) acts as an effective Hamiltonian for $\mathcal{M}$, then fixed points of $\mathcal{M}$ should correspond to equilibrium (critical) points of h. That is, fixed points should be solutions to the equations

$$\partial h/\partial q = 0 \tag{7.92a}$$

$$\partial h/\partial p = 0. \tag{7.92b}$$

Evidently, Eqs. (7.92) are satisfied at the origin. Indeed, at the origin h has an expansion of the form

$$h = -w(p^2 + q^2)/2 + \text{terms of degree 3 and higher}\,. \tag{7.93}$$

Consequently, the quadratic terms dominate at the origin. It follows that the linear part of the map at the origin is given by the matrix

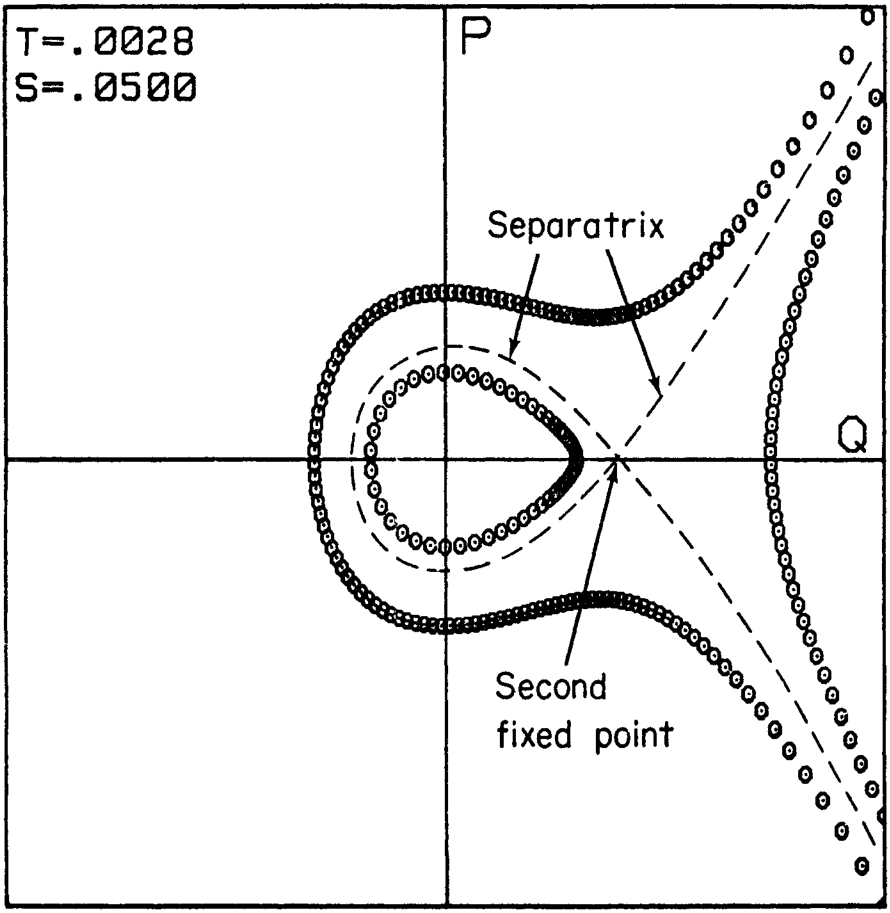


Figure (7.4): Behavior near an integer tune showing stable and unstable regions and their separatrix. Note that the regions of stability and instability are well predicted by h. In the viewing area the coordinates q and p range from -1 to 1.

$$M = \begin{pmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{pmatrix} \quad . \tag{7.94}$$

This result is also obvious from Eq. (7.84). Note that the eigenvalues of M are $\exp(\pm iw)$. Therefore, according to exercise 7.6, the fixed point at the origin is elliptic. Consequently, the fixed point at the origin is stable in the sense that points initially very near the origin remain near the origin (in the linear approximation) under repeated applications of $\mathcal{M}$.

Equations (7.92) are also satisfied at a second point. This point is most easily found using action-angle variables. Since the transformation (7.49) is canonical, equations (7.92) can be written in the equivalent form

$$\partial h/\partial\phi = 0 \tag{7.95a}$$

$$\partial h/\partial a = 0. \tag{7.95b}$$

Upon carrying out the indicated differentiations and solving Eqs. (7.95) using (7.87), one finds the results

$$\phi = \pi/2 - w/2 \tag{7.96a}$$

$$(2a)^{1/2} = [(8)/(3s)][\sin(3w/2)\sin(w/2)]/[\sin(3w/2) + \sin(w/2)]. \tag{7.96b}$$

These results can then be substituted in Eqs. (7.69) to find, in the approximation that the terms of degree 4 and higher in (7.65) can be neglected, that the fixed point should have coordinates q,p satisfying the relations

$$q = 2s^{-1} (\sin w)(1 + \cos w)^{-1} (1 + 2 \cos w)/3 \tag{7.97a}$$

$$p/q = (\sin w)/(1 + \cos w). \tag{7.97b}$$

Because the mapping under study is rather simple, its fixed points can also be found directly and exactly. Suppose one requires that $\bar{q} = q$ and $\bar{p} = p$ in Eqs. (7.84). One finds, in addition to the fixed point at the origin, the result

$$q = 2s^{-1}(\sin w)(1 + \cos w)^{-1} \tag{7.98a}$$

$$p/q = (\sin w)/(1 + \cos w). \tag{7.98b}$$

Evidently, Eqs. (7.97b) and (7.98b) are identical, and hence Eq. (7.96a), although derived from an expression for h which neglects terms of degree 4 and higher, is actually exact. The exactness of this result is related to the observation that, as a result of time reversal invariance, the function h must be symmetric to all orders about the line $\phi = \pi/2 - w/2$.

Comparison of Eqs. (7.97a) and (7.98a) shows that they differ by the factor $(1 + 2\cos w)/3 = 1 - w^2/3 + \cdots$. Thus, the approximate and exact expressions for q agree in the small w limit, and differ only by terms of order w^2. This is what should be expected. Note that, according to (7.98a), when w is of order 1, then the quantity (qs) is also of order 1. But, in this case, the neglected terms of higher degree in the expansion for h can no longer be ignored.

At this point it is interesting to recall that section 7.1 and exercise 7.4 showed the existence and local uniqueness of closed orbits as long as tunes are away from integer values. Inspection of Eqs. (7.97) or (7.98) shows that the fixed point away from the origin coalesces with the origin as $w \to 0$. That is, this simple example illustrates that two distinct closed orbits can merge as a tune goes through an integer value.

Exercise 7.15: Verify Eqs. (7.96), (7.97), and (7.98).

The stability of the fixed point away from the origin can be examined by expanding h about the fixed point. Since the fixed point is an equilibrium point, the first nonvanishing derivatives are given by the expressions

$$\begin{aligned}\partial^2 h/\partial\phi^2 &= (ws/8)(2a)^{3/2}\{[9\sin(3\phi + 3w/2)]/[\sin(3w/2)] \\ &\quad - [\sin(\phi + w/2)]/[\sin(w/2)]\} \\ &= -(ws/8)(2a)^{3/2}\{9[\sin(3w/s)]^{-1} + [\sin(w/2)]^{-1}\},\end{aligned} \tag{7.99a}$$

$$\partial^2 h/\partial\phi\partial a = 0\,, \tag{7.99b}$$

$$\begin{aligned}\partial^2 h/\partial a^2 &= -(3/4)(ws/8)(2)^{3/2}\, a^{-1/2}\{[\sin(3\phi + 3w/2]/[\sin(3w/2)] \\ &\quad - \sin(\phi + w/2)]/[\sin(w/2)]\} \\ &= (3/4)(ws/8)(2)^{3/2}\, a^{-1/2}\{[\sin(3w/2)]^{-1} + [\sin(w/2)]^{-1}\}.\end{aligned} \tag{7.99c}$$

Here use has been made of (7.96a). Observe that the two terms (7.99a) and (7.99c) are of opposite signs. Consequently, in the neighborhood of the equilibrium point, h has an expansion which looks like that of an harmonic oscillator with a negative spring constant. Therefore, the equilibrium is unstable, and correspondingly the fixed point is hyperbolic.

Let z^0 be the coordinates of the hyperbolic fixed point as given by Eq. (7.97) or (7.98). Near z^0, $\mathcal{M}$ can be approximated by its linear part. Therefore, there is a local hyperbolic structure as described in exercise 7.7. In particular, there are the asymptotes which were referred to as the stable and unstable manifolds. The concept of stable and unstable manifolds can be generalized to the full map. Let z denote a general point. Then, the stable manifold, denoted by W_s, is defined to be the set of all points z which are sent into z^0 by $\mathcal{M}^n$ in the limit of large n. In set theoretic notation, one has the definition

$$W_s = \{z \mid \lim_{n\to\infty} \mathcal{M}^n z = z^0\}. \qquad (7.100)$$

Evidently, the stable manifold is the analog of the asymptote corresponding to that eigenvector of the linear part having eigenvalue less than one. Similarly, the unstable manifold is the analog of the asymptote corresponding to the eigenvector having eigenvalue greater than one. Note that points on the unstable manifold, W_u, should move away from z^0 under the action of $\mathcal{M}$, and therefore they should approach z^0 under the action of $\mathcal{M}^{-1}$. Consequently, the unstable manifold is cleverly defined as the set of all points which are sent into z^0 by $\mathcal{M}^{-n}$ in the large n limit,

$$W_u = \{z \mid \lim_{n\to\infty} \mathcal{M}^{-n} z = z^0\}. \qquad (7.101)$$

To the extent that h is an invariant function, the stable and unstable manifolds are given by the level lines of h that pass through z^0. That is, the stable and unstable manifolds are solutions of the equation

$$h(z) = h(z^0). \qquad (7.102)$$

These curves are indicated schematically in Fig. 7.4. Note that two branches of the stable and unstable manifold join, and two branches go off to infinity. Note also that the behavior of points under the action of $\mathcal{M}$ is different on different sides of the stable and unstable manifolds. For this reason, the stable and unstable manifolds are also sometimes referred to together as the separatrix. That is, the stable and unstable manifolds separate phase space into regions of qualitatively different behavior. In particular, points inside that portion of the separatix surrounding the origin remain inside, and points outside are eventually sent to infinity. In the content of a circulating charged particle beam, points inside the separatrix loop correspond to trajectories which remain within the machine, and points outside correspond to trajectories which eventually escape.

It can be shown that for most symplectic maps the stable and unstable manifolds of a hyperbolic fixed point do not join smoothly as Fig. 7.4 would suggest, but rather they intersect at some point at some finite angle. This phenomena is called separatrix splitting. The point of intersection is called a homoclinic point, and the angle is referred to as a homoclinic angle.[35]

When separatrix splitting occurs, the symplectic nature of the map causes the stable and unstable manifolds to oscillate about each other. Correspondingly, the concept of a separatrix loses its meaning. Separatrix splitting occurs for the simple map we have been studying.[33,36] Figure 7.5 shows the behavior of the stable and unstable manifolds for the map (7.84) for a tune angle of $w \simeq 70$ degrees. For convenience of plotting, the origin has been shifted to the hyperbolic fixed point, and the axes, now called x and y, have been lined up along the asymptotes. The existence of a homoclinic point is evident. Figures 7.6 and 7.7 show how the stable and unstable manifolds oscillate about each other as they return to the neighborhood of the hyperbolic fixed point.

In many cases, the angle of intersection between the stable and unstable manifolds at the homoclinic point is too small to be readily visible. For example, when the tune angle has the value of 5 degrees corresponding to Fig. (7.3), the homoclinic angle is less than 10^{-13} degrees. However, the existence of homoclinic angles, no matter how small, rules out the true existence of invariant functions.[36,37] Consequently, the series (7.65) is generally divergent when all terms are included. Nevertheless, the truncated series is still useful, as has been seen, provided the homoclinic angle is small or one does not inquire about properties of the map that are too detailed.

A second and related consequence of the intersection of stable and unstable manifolds is the appearance of chaotic behavior. Figure 7.8 displays the result of iterating the map (7.84) for a variety of initial conditions when the tune angle has the value $w \simeq 76$ degrees. The scale is such as to show only the region in the neighborhood of the elliptic fixed point at the origin. The hyperbolic fixed point given by (7.98) is outside the viewing area of the drawing. Inspection shows that there appear to be five other elliptic fixed points and five other hyperbolic fixed points. These points are actually fixed points of M^5, the fifth power of the map. Figure 7.9 shows a magnification of the behavior near one of these five hyperbolic fixed points.[38] Many of the points in the figure appear to be distributed in a random or chaotic fashion. This behavior arises from the fact that the stable manifolds from one hyperbolic fixed point intersect (at a finite angle) the unstable manifolds from another fixed point. Such points of intersection of manifolds from different fixed points are called heteroclinic points. As is the case with homoclinic points, the manifolds then go into wild oscillation about each other thus producing the apparently chaotic behavior.

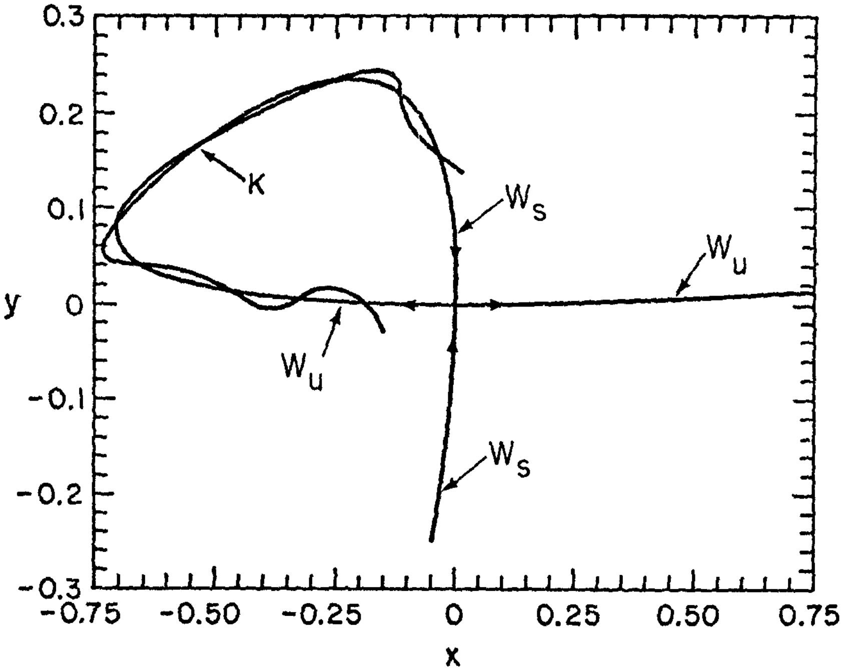


Figure (7.5): The intersection of stable and unstable manifolds emanating from a hyperbolic fixed point resulting in a homoclinic point K.

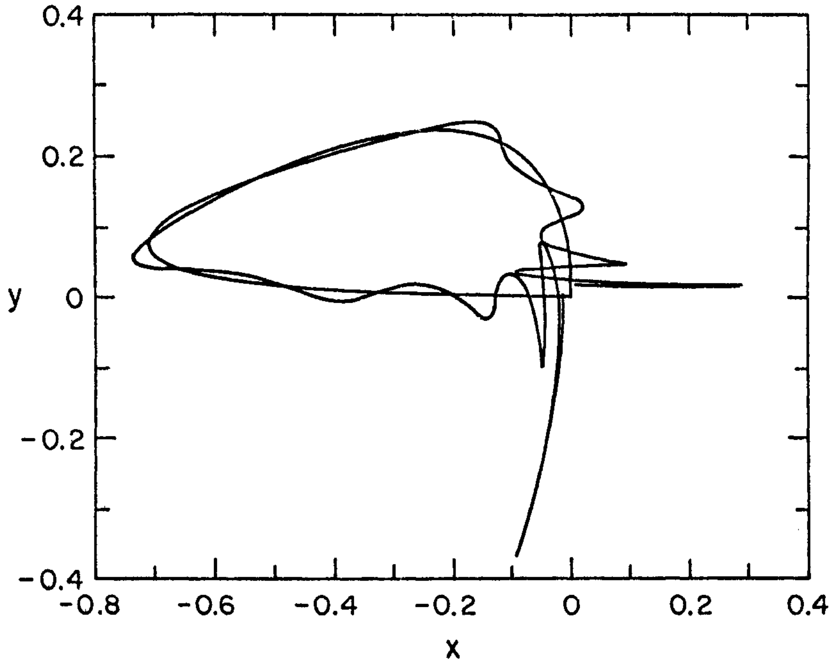


Figure (7.6): Successive homoclinic intersections and oscillations of the stable and unstable manifolds. The other halves of W_u and W_s, those pieces that go off to infinity, are not shown.

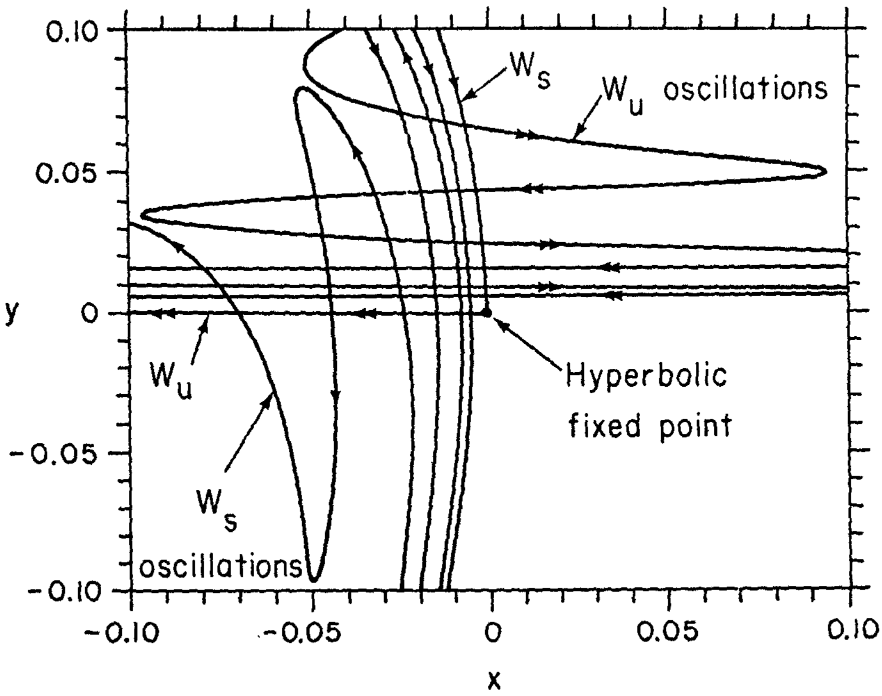


Figure (7.7): A continuation of Fig. (7.6) near the origin showing the formation of a grid of intersecting lines. The spacing of the grid becomes finer and finer as it approaches the hyperbolic fixed point. Each grid intersection is a homoclinic point. The result of all these intersections is an ever denser cloud of homoclinic points which has the hyperbolic fixed point as a limit point.

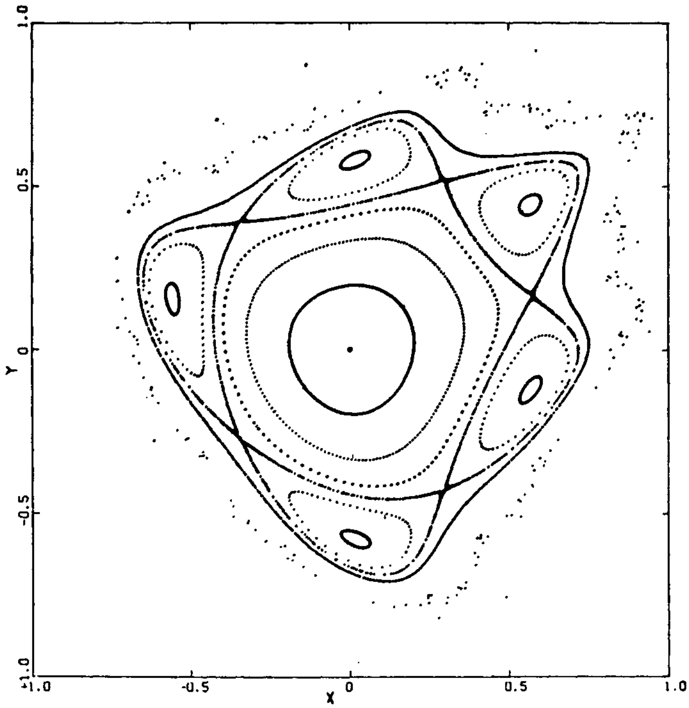


Figure (7.8): Behavior of M near a tune of 1/5. There are five elliptic and five hyperbolic fixed points of M^5. The variables are selected so that the hyperbolic fixed point given by equation (7.98) has coordinates $x \simeq 1.6$, $y \simeq 1.2$.

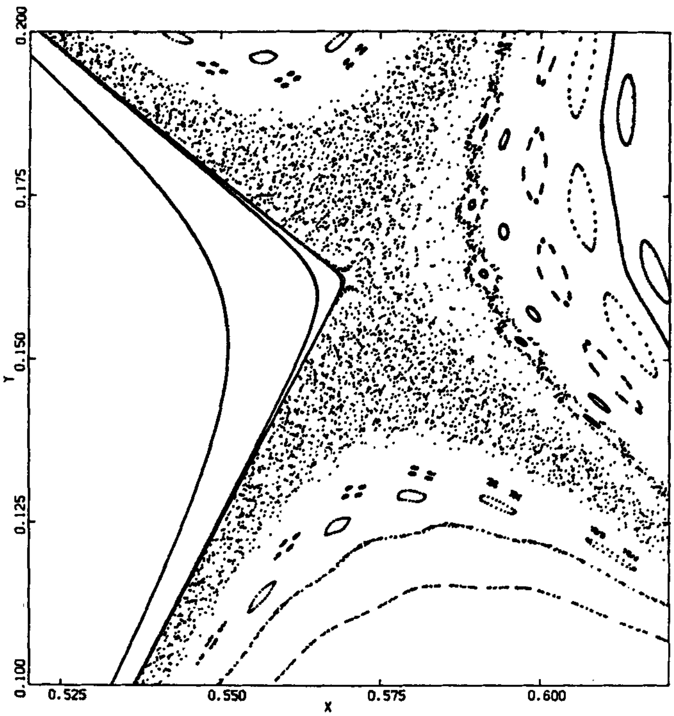


Figure (7.9): A magnification of Fig. (7.8) near one of the hyperbolic fixed points. The chaotic behavior arises from the heteroclinic intersections of the stable and unstable manifolds emanating from the hyperbolic fixed points. Note also the presence of many small islands corresponding to very high order resonances. Indeed, it can be shown that in the case of homoclinic or heteroclinic seperatrix splitting for a hyperbolic fixed point, there must be fixed points of high powers of M in every neighborhood of the "parent" hyperbolic point.

Evidently, even the simplest of nonlinear symplectic maps can exhibit an extremely complicated behavior.

Our discussion of the example of a perfect ring with a short sextupole insertion has so far been restricted to the nonresonant case. Suppose, now, that the tune is at or near one of the resonant values 1/3 or 2/3. In these cases the expansion (7.87) is no longer valid, and the symplectic map $\mathcal{M}$ cannot be written in the exponential form (7.62).

However, we will see that although $\mathcal{M}$ cannot be written in exponential form at a third integer resonance, the <u>cube</u> of the map, $\mathcal{M}^3$, can still be written in exponential form. For brevity, attention will be restricted to tunes at and near the resonant value 1/3. The case of the resonant value 2/3 can be handled analogously.

Suppose the tune is written in the form

$$T = w/(2\pi) = (1/3) + \delta \tag{7.103}$$

where the quantity δ measures departure from exact resonance. Now watch closely a bit of algebraic sleight of hand! If $\mathcal{M}$ is written in the form (7.64), then $\mathcal{M}^3$ is obviously the product

$$\begin{aligned} \mathcal{M}^3 = {} & \exp(:g_2:)\exp(:g_3:) \times \\ & \exp(:g_2:)\exp(:g_3:) \times \\ & \exp(:g_2:)\exp(:g_3:). \end{aligned} \tag{7.104}$$

However, this product can be rewritten in the form

$$\begin{aligned} \mathcal{M}^3 = {} & \exp(3:g_2:)\exp(-2:g_2:)\exp(:g_3:)\exp(2:g_2:) \times \\ & \exp(-:g_2:)\exp(:g_3:)\exp(:g_2:) \times \\ & \exp(:g_3:). \end{aligned} \tag{7.105}$$

Several simplications are now possible. The term $\exp(3:g_2:)$ can be rewritten with the aid of (7.103) and (7.68). One finds the result

$$\exp(3:g_2:) = \exp(-\pi:p^2 + q^2:)\exp(:f_2:) \tag{7.106}$$

where f_2 denotes the function

$$f_2 = -3\delta\pi(p^2 + q^2). \tag{7.107}$$

However, the first Lie transformation in (7.106) corresponds to rotation by 2π in the q,p phase space, and is therefore just the identity operator. See exercise 3.23. Therefore, one also has the relation

$$\exp(3:g_2:) = \exp(:f_2:). \qquad (7.108)$$

The remaining terms in (7.105) can be simplified using the calculus developed in section 5.4. Employing Eq. (5.52), one finds the results

$$\begin{aligned} &\exp(-2:g_2:)\exp(:g_3:)\exp(2:g_2:) \\ &= \exp(:\exp(-2:g_2:)g_3:), \end{aligned} \qquad (7.109a)$$

and

$$\begin{aligned} &\exp(-:g_2:)\exp(:g_3:)\exp(:g_2:) \\ &= \exp(:\exp(-:g_2:)g_3:). \end{aligned} \qquad (7.109b)$$

It follows from exercise 5.6 that the quantities $\exp(-2:g_2:)g_3$ and $\exp(-:g_2:)g_3$ are both homogeneous polynomials of degree three. Therefore, using the Campbell-Baker-Hausdorff formula to combine exponents and neglecting polynomials of degree four and higher, which is in keeping with the spirit of our calculations, one finds that the remaining terms in (7.105) can be rewritten in the form

$$\begin{aligned} &\exp(-2:g_2:)\exp(:g_3:)\exp(2:g_2:) \times \\ &\exp(-:g_2:)\exp(:g_3:)\exp(:g_2:) \times \\ &\exp(:g_3:) = \\ &\exp(:[\exp(-2:g_2:) + \exp(-:g_2:) + 1]g_3:). \end{aligned} \qquad (7.110)$$

Combining (7.108) and (7.110) gives the result

$$M^3 = \exp(:f_2:)\exp(:[\exp(-2:g_2:) + \exp(-:g_2:) + 1]g_3:). \qquad (7.111)$$

Suppose the Campbell-Baker-Hausdorff formula is now used to combine the two exponents in (7.111) to obtain a single exponent. For later convenience, call this exponent $3h_r$. Then one finds the result

$$M^3 = \exp(3:h_r:) \qquad (7.112)$$

with h_r given by the relation

$$3h_r = f_2 + :f_2:[1 - \exp(-:f_2:)]^{-1} \times$$
$$[\exp(-2:g_2:) + \exp(-:g_2:) + 1]\, g_3. \qquad (7.113)$$

See Eq. (5.81). However, from the algebraic identity

$$a^2 + a + 1 = (1 - a^3)/(1 - a), \qquad (7.114)$$

there follows the operator identity

$$\exp(-2:g_2:) + \exp(-:g_2:) + 1$$
$$= [1 - \exp(-3:g_2:)]/[1 - \exp(-:g_2:)]. \qquad (7.115)$$

Also, from (7.108), it follows that

$$1 - \exp(-3:g_2:) = 1 - \exp(-:f_2:). \qquad (7.116)$$

Therefore, using (7.115) and (7.116), the expression (7.113) for h_r can also be written in the form

$$3h_r = f_2 + :f_2:[1 - \exp(-:g_2:)]^{-1} g_3. \qquad (7.117)$$

Inspection of Eqs. (7.65) and (7.117) shows that the expressions for h and h_r are rather similar. Consequently, for g_3 given by (7.83), we can use our previous calculations to find the specific form of h_r. One finds the result

$$h_r = -2\pi\delta a - (2\pi\delta s/8)(2a)^{3/2} \times$$
$$\{[\sin(3\phi+3w/2)]/[\sin(3w/2)] - [\sin(\phi+w/2)]/[\sin(w/2)]\}.$$
$$(7.118)$$

Suppose that the tune $(w/2\pi)$ approaches the resonant value (1/3). Then one has the pleasure of evaluating the limiting ratio

$$\delta/[\sin(3w/2)] = \delta/\sin(\pi+3\pi\delta) = -1/(3\pi) + O(\delta). \qquad (7.119)$$

Evidently the function h_r, unlike h, is well behaved both at and near resonant tunes.

<u>Exercise 7.16</u>: Verify Eqs. (7.105), (7.106), (7.113), (7.117), and (7.118).

How well does all this work in practice? Figure 7.10 shows a phase-space plot in the Poincare surface of section for the near-resonant case w = 119 degrees. The points on the various curves are the images, under successive applications of M^3, of a given initial point on each curve. For comparison, the circles are centered on points (again not shown) obtained by taking the given initial point as an initial condition, and following the Hamiltonian flow generated by $-h_r$ for 3n units of time. Evidently, there are two kinds of curves. Those that are far from the origin lead to infinity, and correspond to orbits on which particles are eventually lost. Those that are sufficiently close to the origin (corresponding to sufficiently small betatron oscillation amplitudes) encircle the origin, and hence describe stable orbits. In the limit of exact resonance this stable region shrinks to zero, and all of phase space becomes unstable.

Observe also that there is good agreement between the points and the circles. That is, Eq. (7.112) is well satisfied with h_r given by (7.118). Consequently, in the resonant or near resonant case, h_r can be used to predict the stable and unstable regions of phase space, and the behavior of orbits in each region. In particular, h_r can be used to predict regions of stability and instability, resonance widths, and the betatron amplitude growth rates of unstable orbits.

The examples just described have all been limited to the one-dimensional case. It has been seen that the functions h and h_r can be used to compute all relevant linear and nonlinear features of orbits (apart from the consequences of possible homoclinic and heteroclinic behavior which are assumed to be visible only on a fine scale) in both the nonresonant and resonant cases. Preliminary calculations indicate that similar results can be obtained in the full two-dimensional case. In this case, the function h is given by (7.81), and there is a related formula for h_r to be used under resonance and near resonance conditions. As before, h and h_r will be invariant functions, and their negatives will also act as effective Hamiltonians. However, since the Poincare surface of section is now a four-dimensional phase space, a knowledge of the "level surfaces" of h or h_r may not be sufficient to give complete qualitative information about orbits. Indeed, one is now in principle faced with the full general problem of motion in four-dimensional phase space, and such motion can be extremely complicated. For this reason, it may be useful to look for additional invariant functions, perhaps by the method of normal forms.[29,39] These additional invariant functions would be formal integrals of motion for the Hamiltonians $-h$ or $-h_r$.

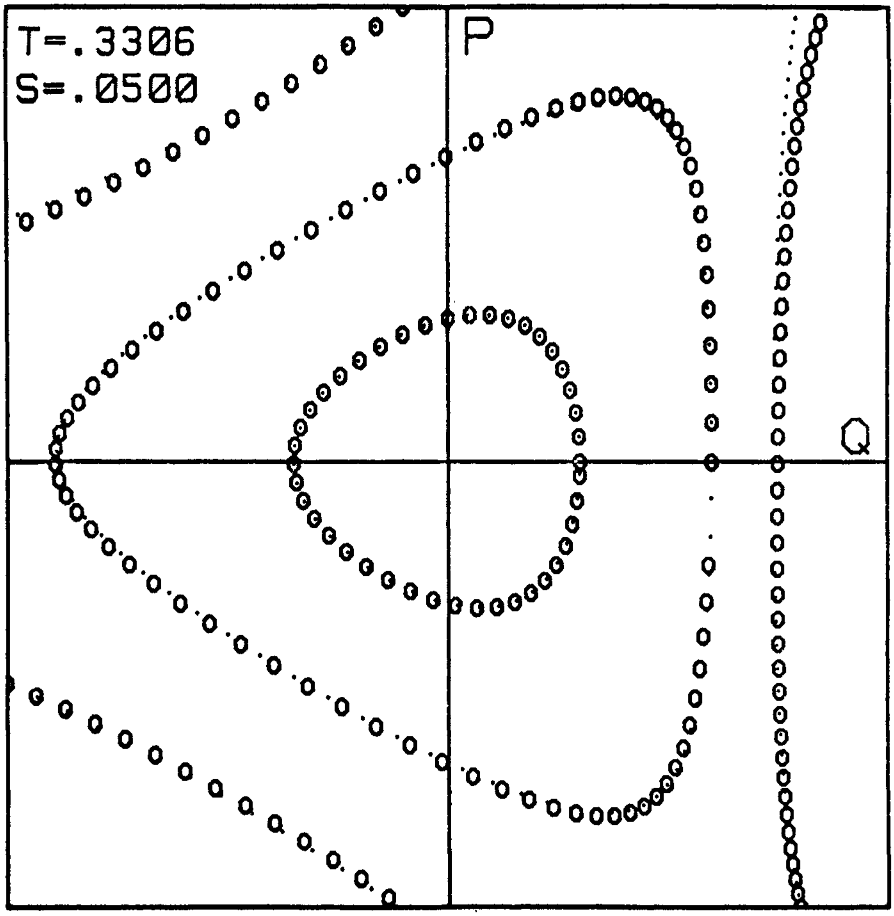


Figure (7.10): The behavior of M^3 near a third-integer tune. Observe that the regions of stability and instability are well predicted by h_r. In the viewing area the coordinates q and p range from -1 to 1.

7.4 The beam-beam interaction

Consider two charged particle beams, cirulating in the same or different rings, which intersect in a certain collision region. Suppose further that attention is restricted to the "weak-strong" limit or approximation. In this case, one beam, the strong beam, is taken to be fixed and unaffected by the second beam. The other beam, the weak beam, is treated as a collection of particles that are affected by their passage through the strong beam, but not by each other.

In the weak-strong limit, the net motion of a particle in the weak beam can be viewed as the continual repetition of two sequential motions: passage through the storage ring followed by passage through the strong beam. See Fig. 7.11. The equations of motion for each of these two passages (through the ring and through the strong beam) are derivable from Hamiltonians, and therefore each passage is described by a symplectic map.

Suppose the passages through the ring and the strong beam are described by the symplectic maps $\mathcal{M}_R$ and $\mathcal{M}_B$ respectively. By design, the passage through the storage ring is well described by a linear map. Thus, in lowest approximation, $\mathcal{M}_R$ can be written in the form

$$\mathcal{M}_R = \exp(:g_2:). \tag{7.120}$$

Higher order nonlinear corrections, as described in the previous sections, can be included if desired; but they will be omitted for the time being for simplicity.

What can be said about the map $\mathcal{M}_B$ that describes passage through the strong beam? According to the factorization theorem, $\mathcal{M}_B$ can be written in the factored product form (5.15). However, in this case the factored product representation, if truncated after a small number of terms, is not a particularly good representation of $\mathcal{M}_B$. This is because by design the charge and current density of the strong beam vary rapidly over a small spatial region, and hence $\mathcal{M}_B$ is very nonlinear. This apparent difficulty can be overcome by observing that, for $\mathcal{M}_B$, each exponent occurring in the factored product representation must be rather small. (This is because the beam-beam interaction is quite weak in the sense that not very much happens in a single passage through the strong beam.) Consequently, the different exponents can in principle be combined into one grand exponent using the Campbell-Baker-Hausdorff formula. Thus there is some function f_b, expected to be small but quite nonlinear, such that $\mathcal{M}_B$ can be written in the form

$$\mathcal{M}_B = \exp(:f_b:). \tag{7.121}$$

Upon combining the results of the previous two paragraphs, it follows that the grand symplectic map $\mathcal{M}$ that describes the result of passage through the ring and then the strong beam can be written in the form

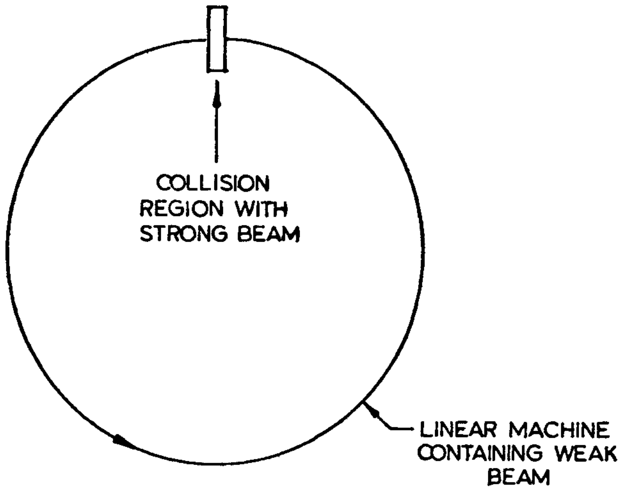


Figure (7.11): Schematic representation of particle motion in a storage ring and a colliding beam region.

$$\mathcal{M} = \exp(:g_2:)\exp(:f_b:). \tag{7.122}$$

In order to compute the long-time behavior of particles in the weak beam, the quantity of interest is again $\mathcal{M}^n$ for large n. As before, we try to find a function h such that $\mathcal{M}$ can be written in the form (7.62).

From this point on, the discussion of the beam-beam interaction in the weak-strong limit parallels the discussion of the stability of closed orbits in the presence of nonlinearities. In particular, formulas (7.65), (7.79), and (7.81) may be taken over directly simply by replacing g_3 by the function f_b. Similarly, there are related expressions for h_r in resonant and near resonant cases. Only two important differences must be kept in mind. First, the remainder terms in (7.65) are no longer of degree 4 and higher, but rather are of second order and higher in the beam-beam interaction strength.[40] Second, because of the possibility of nonlinearities of arbitrarily high degree in f_b, the summation ranges in (7.79) and (7.81) are generally infinite. Thus, in principle, nonlinear resonances of arbitrarily high order can occur.

As is the case with nonlinear structure resonances, the complete details of the effect of the beam-beam interaction in the full two-dimensional problem have not yet been worked out from a Lie algebraic perspective. Some work has been done on a one-dimensional model, and a discussion for a simple example will be given to close this section.[41]

Suppose the strong beam is assumed to be an unbunched ribbon in the horizontal plane. That is, the horizontal dimension of the strong beam is much larger than the vertical dimension. Assume also that the vertical charge distribution of the strong beam is well described by a Gaussian shape. Finally, assume that the weak beam lies in the same horizontal plane and crosses the strong beam at a fixed angle. Then the primary effect of the strong beam on particles in the weak beam is to cause vertical deflections.[42] Consequently, attention will be restricted to the vertical degree of freedom of the weak beam.

To find the map $\mathcal{M}_B$ exactly, it is necessary to integrate the nonlinear equations of motion for a particle passing through the strong beam. However, a good approximation to this map is given by assuming that the particle suffers a vertical momentum change depending only upon its initial vertical position, and that the vertical position itself remains unaffected. Thus, there is some "deflection function" u(q) in terms of which $\mathcal{M}_B$ has the effect

$$\begin{aligned} \mathcal{M}_B q &= q \\ \mathcal{M}_B p &= p + u(q). \end{aligned} \tag{7.123}$$

This impulse approximation becomes exact in the limit that the interaction region becomes a point and/or the transit time through the region approaches zero. In any case, the mapping (7.123) is symplectic, and therefore its use will produce no qualitative error.

The function u is proportional to the electrostatic force exerted by the strong beam. In the Gaussian model employed and with a suitable choice of coordinates, u is given by the relation[43]

$$u(q) = 4\pi D/\sqrt{3} \int_0^{q\sqrt{3}} dt\, e^{-t^2}. \qquad (7.124)$$

Here D is the beam-beam strength parameter that typically has values ranging from 10^{-3} to 10^{-2}. It is normalized in such a way that the beam-beam interaction depresses the tune for infinitesimal betatron oscillations by an amount D when D is small. Now let f_b be the function defined by

$$f_b(q) = \int_0^q u(q')dq'. \qquad (7.125)$$

Then it is readily verified that $\mathcal{M}_B$ as given by the Lie transformation (7.121) produces the desired action (7.123).

<u>Exercise 7.17</u>: Verify that (7.121) with f_b given by (7.125) satisfies (7.123).

Explicit calculation shows that f_b has an expansion of the form

$$f_b = \sum_{-\infty}^{\infty} c_n(a) \exp(2in\phi) \qquad (7.126)$$

where the coefficients $c_n(a)$ are given by the formulas[41]

$$c_o(a) = 2\pi Da \sum_{m=0}^{\infty} (-3a/2)^m (2m)!/\{(m!)[(m+1)!]^2\} \qquad (7.127a)$$

and

$$c_n(a) = -(4\pi/3)D(3a/2)^n \times$$
$$\sum_{m=0}^{\infty}(-3a/2)^m(2n+2m-2)!/[m!(n+m-1)!(m+2n)!] \qquad (7.127b)$$

when $n > 0$.

When use is made of this expansion, and g_2 is taken to be given by (7.68) as before, then h is found to be given by the relation

$$h = -wa + c_o(a) + 2\sum_1^{\infty} c_n(a)[nw/\sin(nw)]\cos[2n(\phi+w/2)]$$
$$+ \text{ terms of order } D^2 \text{ and higher.} \qquad (7.128)$$

Inspection of (7.128) indicates that the beam-beam interaction should produce nonlinear resonances whenever

$$nw = \ell\pi \text{ or } T = w/(2\pi) = \ell/(2n). \qquad (7.129)$$

Thus, there should be resonances at half-integer, quarter-integer, sixth-integer tunes, etc. The strength of these resonances should be proportional to $nc_n(a)$ for n=1,2,3, etc. It can be shown that the quantity (nc_n) falls off faster than exponentially in n with increasing n.[41] Moreover, according to (7.127b), c_n for large n falls off rapidly as $a \to 0$. Consequently, the sizes of various resonance features in phase space should decrease rapidly with the order of the resonance. In addition, these features should decrease in size according to their proximity to the origin in phase space.

How does this work out in practice? Figures (7.12) through (7.15) show phase-space plots generated numerically by applying successive iterates of $\mathcal{M}$ to various initial conditions, and using various tune values. The phase-space coordinates range over (-2,2), and the scale is chosen so that the strong beam lies within $-1 \leq q \leq 1$. The tunes are near the resonant values 1/2, 1/4, 1/6, and 1/8 respectively, and the beam-beam interaction strength is 10^{-2}. Observe that the expected resonances are all present, and that the size of resonance features, e.g. island dimensions, do indeed decrease with increasing order of the resonance.

Figure (7.16) shows a tenth-order resonance obtained by running near a tune of 1/10. It was not shown as part of the previous sequence because the island structure becomes too small to see when (by adjusting the tune) it is located closer to the origin. This example illustrates that the sizes of resonant features do indeed decrease with proximity to the origin, and in fact the higher the order of the resonance, the more rapid is the decrease.

Figure (7.17) shows the case of running with a nonresonant tune of 77/100. On the scale shown, and for the number of iterations made, there seems to be no evidence that any points will escape from the beam. Just as in the treatment given in the previous section of the effect of nonlinearities, the quantity h should again act as a generalization of the Courant-Snyder invariant in the non-resonant case. Figure (7.18) illustrates this proposition. It shows plots of the two quantities (-wa) and h as a function of $[\phi/(2\pi)]$ for each of the curves in Fig. (7.17). It is evident that the quantity (-wa), which is proportional to the ordinary Courant-Snyder invariant, can, in some cases, show substantial variations. By contrast, the quantity h is more nearly constant in all cases.

In the situation of tunes near and at resonant values, it is again useful to work with powers of $\mathcal{M}$. Suppose the tune is written in the form

$$T = w/(2\pi) = k/m + \delta \qquad (7.130)$$

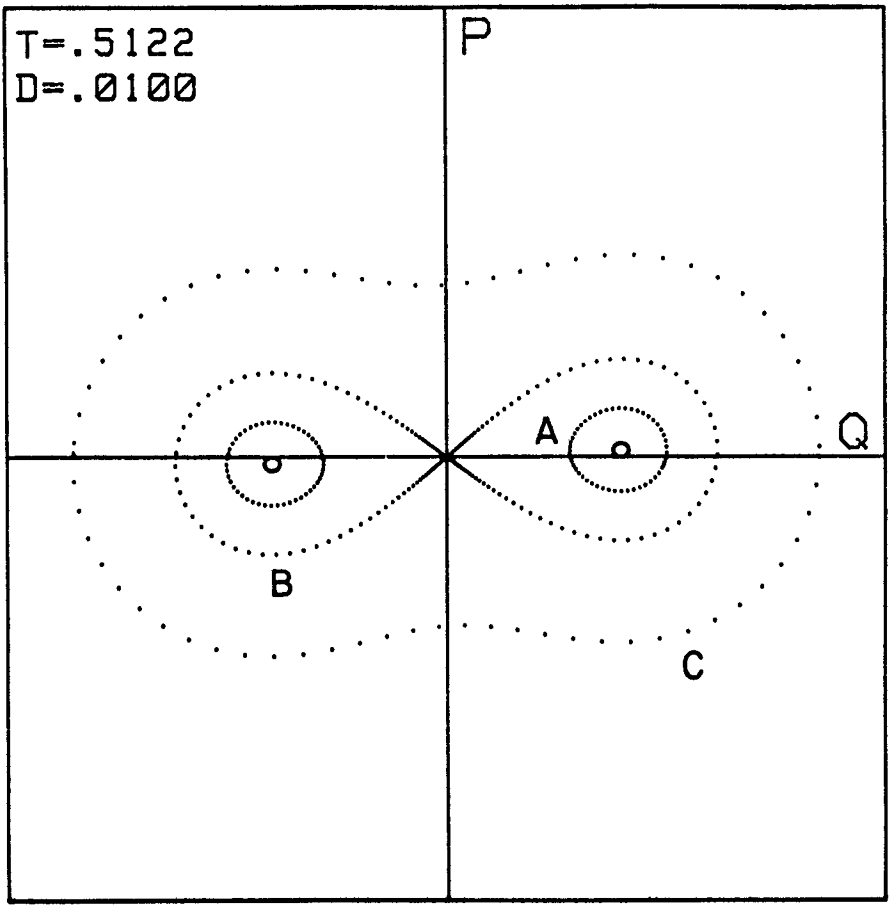


Figure (7.12): Phase-space plot generated by successive iterations of M for various initial conditions. The tune is near one half, and the beam-beam interaction strength is 10^{-2}. The coordinates extend from −2 to 2, and are normalized in such a way that the strong beam has fallen off by $1/e^3$ at $q = \pm 1$.

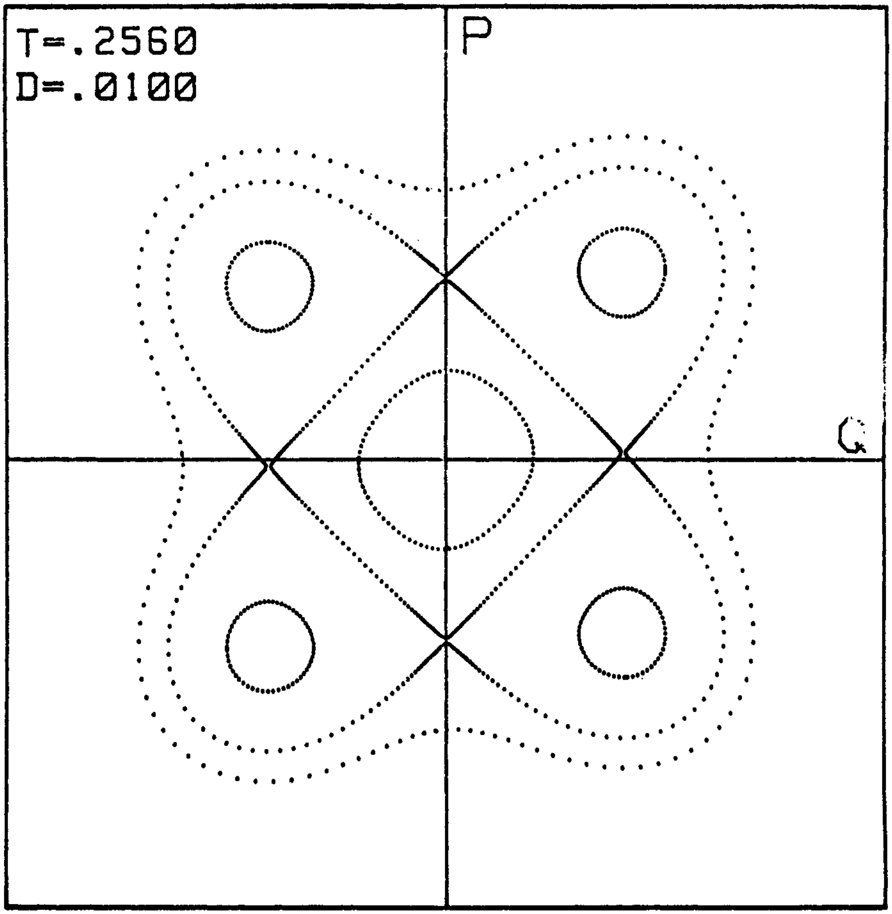


Figure (7.13): Phase-space plot when the tune is near one fourth.

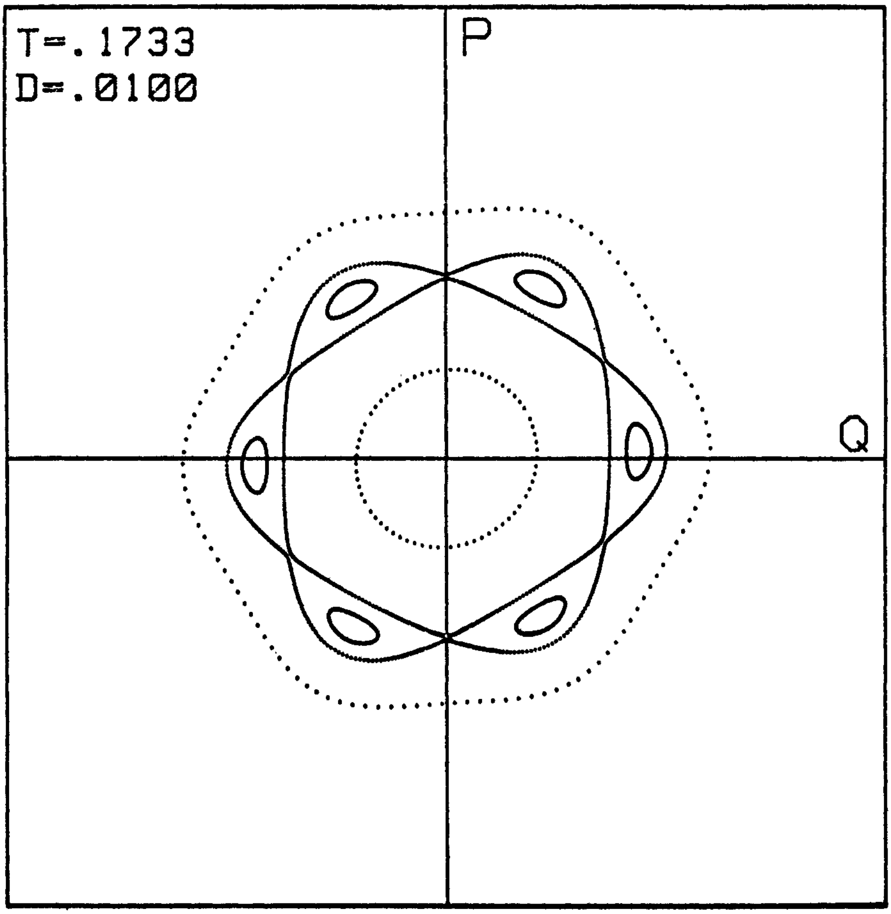


Figure (7.14): Phase-space plot when the tune is near one sixth.

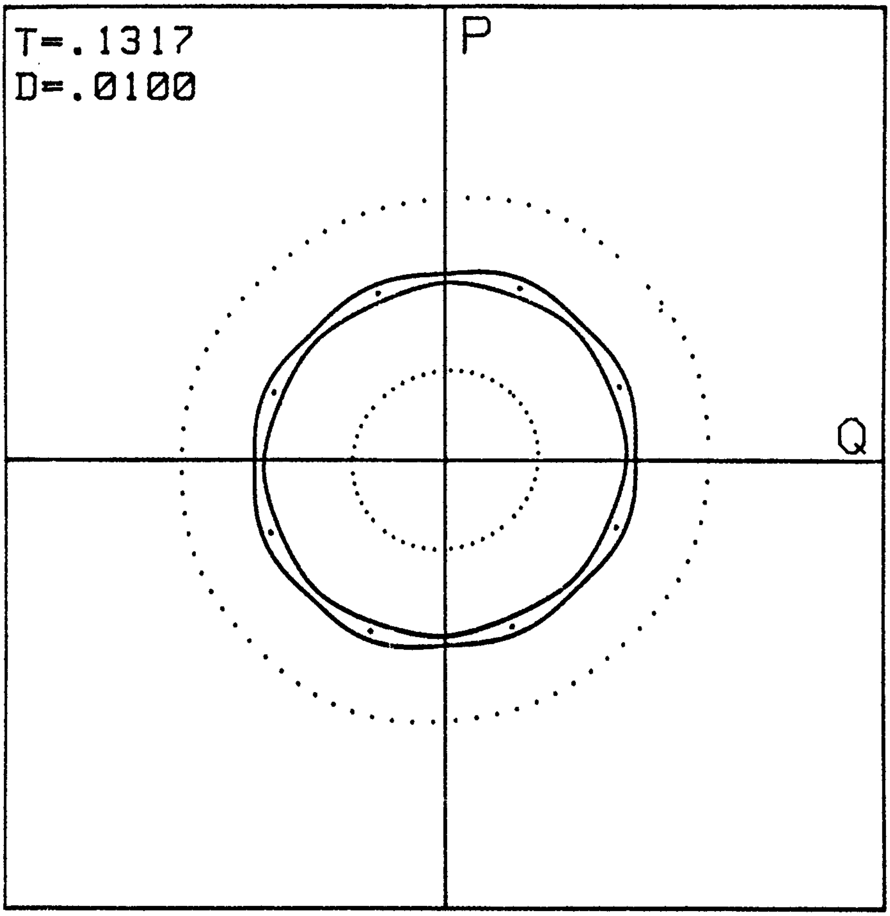


Figure (7.15): Phase-space plot when the tune is near one eigth.

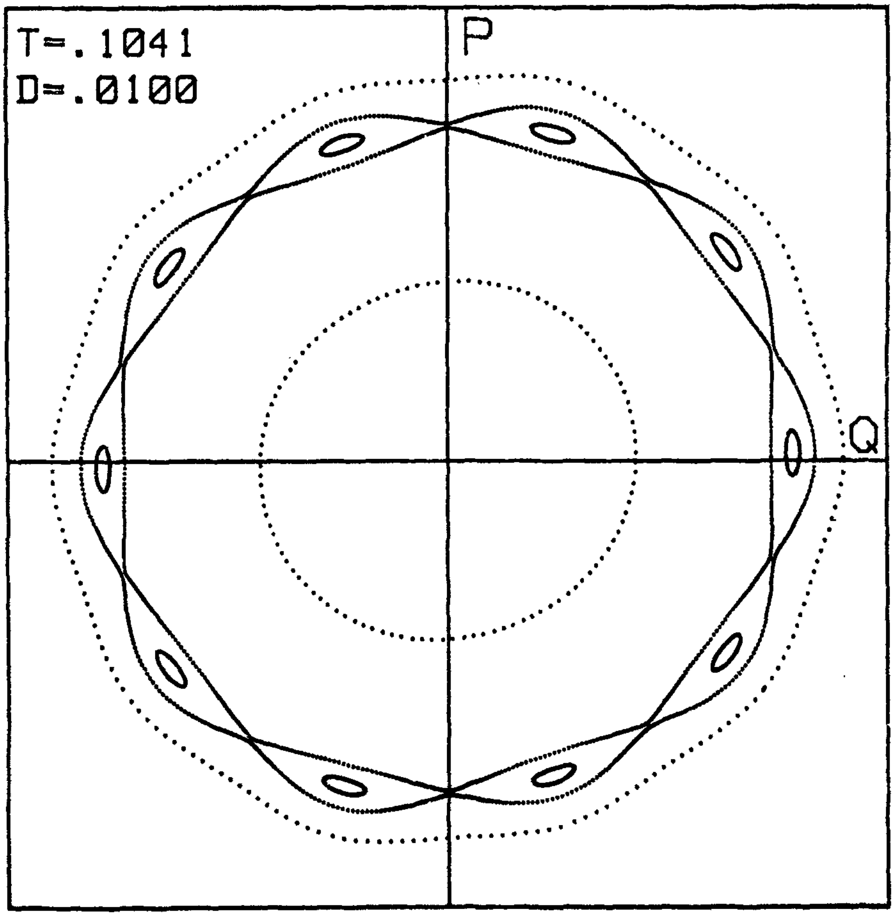


Figure (7.16): Phase-space plot near a tune of one tenth.

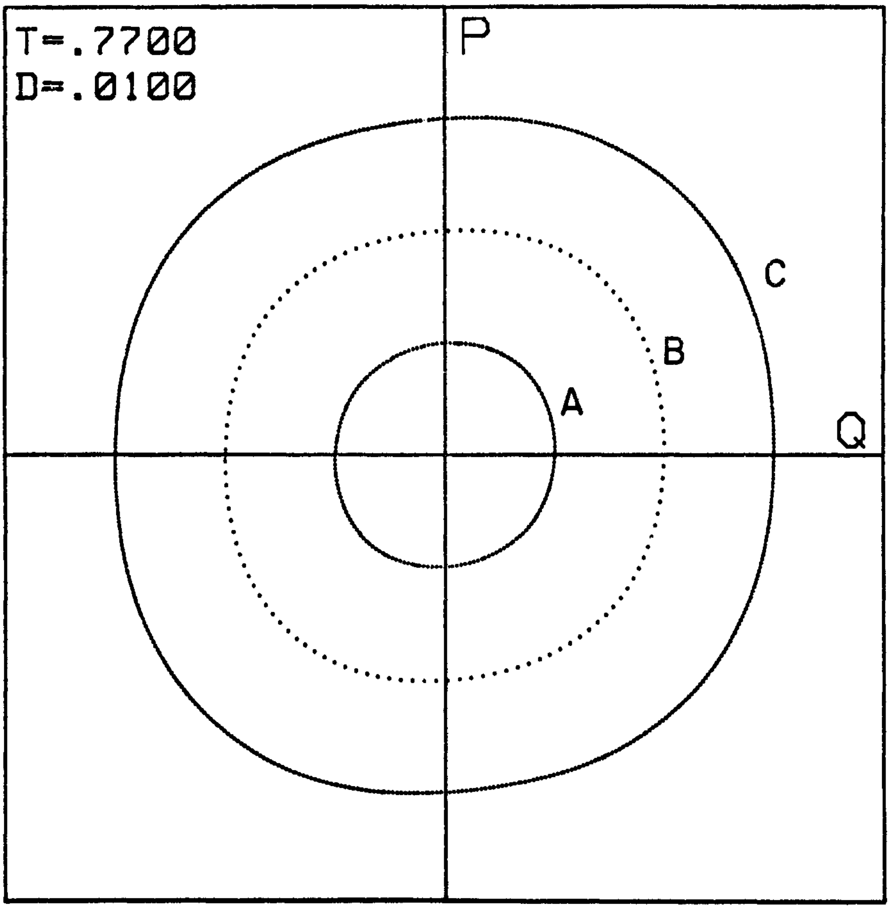


Figure (7.17): Phase-space plot for a nonresonant tune.

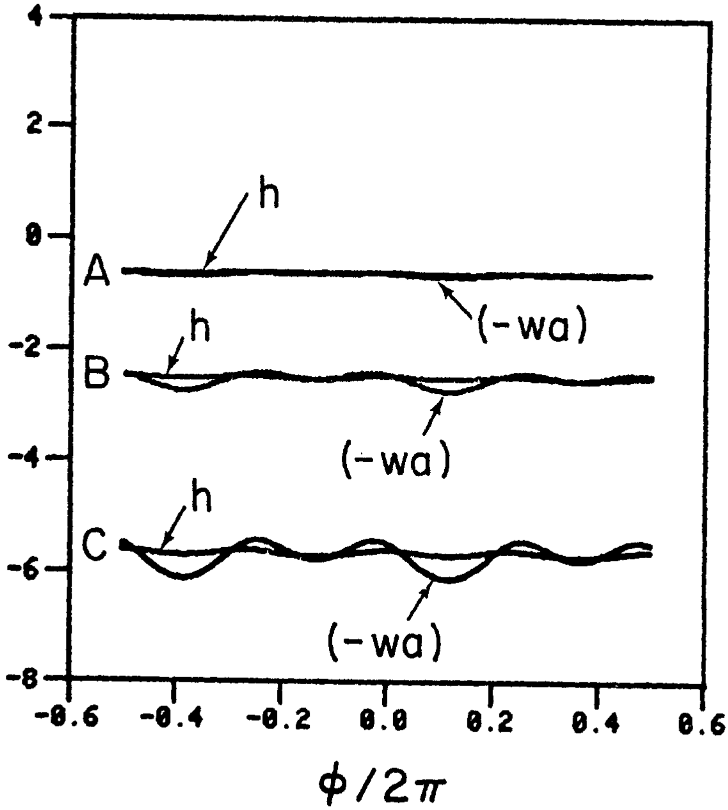


Figure (7.18): Plots of the quantities (-wa) and h for each of the three cases of Fig. (7.17). Observe that h is more nearly constant than the ordinary Courant-Snyder invariant (-wa). Observe also that the larger variations in the ordinary Courant-Snyder invariant occur in the cases with the larger betatron amplitudes, i.e., the cases where nonlinearities are more important.

where the quantity δ measures departures from the exact resonance values k/m. Now consider the m'th power of the map $\mathcal{M}$. By proceeding in a manner analogous to that used for the special case m=3 in the last section, one finds that the m'th power of the map $\mathcal{M}$ can be written in the form

$$\mathcal{M}^m = \exp(m{:}h_r{:}) \tag{7.131}$$

with h_r given by the formula

$$h_r = -\delta a + c_o(a) + 2 \sum_1^\infty c^n(a)[n\delta/\sin(nw)] \cos[2n(\phi+w/2)]$$

$$+ \text{ terms of order } D^2 \text{ and higher.} \tag{7.132}$$

Again, h_r is well defined both near and at resonant tune values.

<u>Exercise 7.18</u>: Provide the derivation of Eq. (7.132).

The quantity h_r should be an invariant function in the resonant case. As an illustration of how this works out in the particular example of a near half-integer tune, Fig. (7.19) shows values of h_r plotted as functions of q for each of the three labeled curves of Fig. (7.12). Observe that the quantity h_r remains remarkably constant. Thus, the curves of Fig. (7.12) are very nearly level lines of h_r, and the major features of Fig. (7.12) can be predicted using $(-h_r)$ as an effective Hamiltonian.

We have again seen that the functions h and h_r can be used to compute all relevant nonlinear features of orbits in both the nonresonant and resonant regimes for the one-dimensional case. Similar results should be within reach for the two-dimensional case. This general problem should be examined after the related but simpler two-dimensional problem of the previous section has been thoroughly studied.

We close this section with an illustration of separatrix splitting in the case of the beam-beam interaction. Study of the fixed point at the origin of Fig. (7.12) shows that it is inversion hyperbolic. Consequently, the origin is a hyperbolic fixed point of $\mathcal{M}^2$. Next observe that the stable and unstable manifolds of $\mathcal{M}^2$ emanating from the origin appear to join smoothly to form the figure eight curve labeled B. On general grounds, according to the discussion of the previous section, one would expect these manifolds to intersect and then go into oscillation about each other. Apparently, the angle of homoclinic intersection in Fig. (7.12), if it exists at all, must be very small.

By contrast, Fig. (7.20) shows the case of Fig. (7.12) extended to larger values of the beam-beam interaction with the value of the tune suitably adjusted to maintain a near resonance condition. All

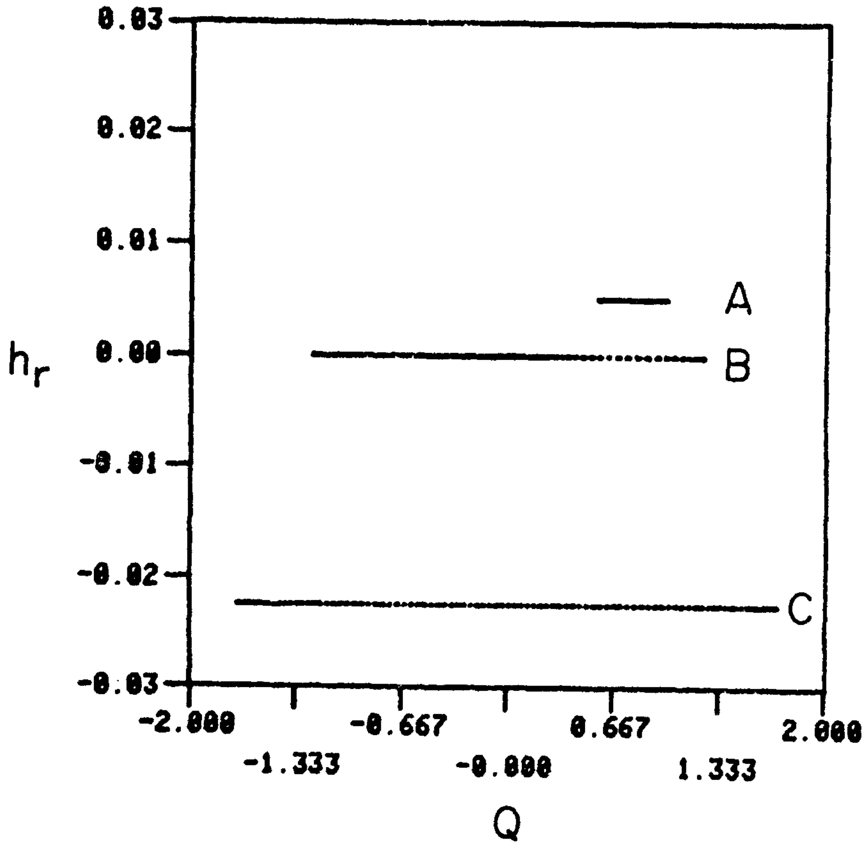


Figure (7.19): Values of h_r, the generalized Courant-Snyder invariant in the resonant case, plotted for each of the three cases of Fig. (7.12). Note that h_r is remarkably constant despite the large distortions from circular behavior evident in Fig. (7.12).

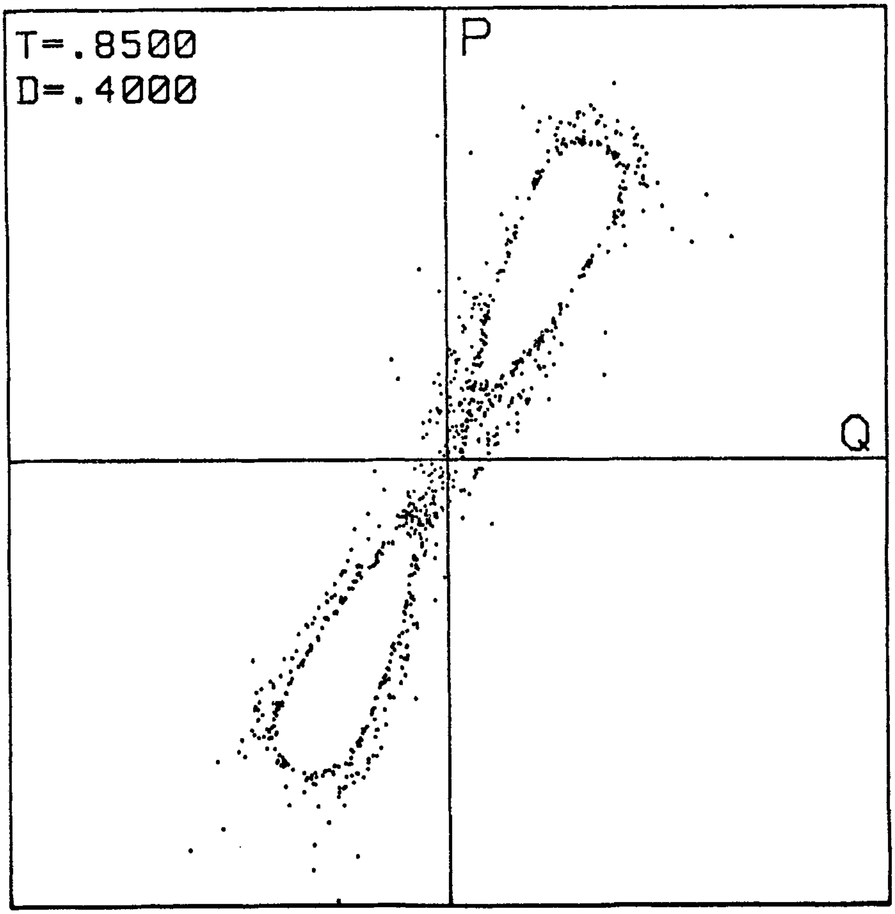


Figure (7.20): A phase-space plot showing chaotic behavior for a large value of the beam-beam interaction strength. All the points displayed are the images of a single point under repeated application of M.

the points shown are the image of a single point under the action of successive powers of . Chaotic behavior is clearly now in evidence. This behavior arises because the angle of homoclinic intersection, as illustrated schematically in Fig. (7.21), is now quite large.

ACKNOWLEDGMENTS

Much of this work was carried out with the collaboration of Mr. David R. Douglas. The author is also grateful to the Los Alamos National Laboratory for the fine hospitality provided during a sabbatical year in which many of the ideas for this work were conceived.

This work was supported in part by U.S. Department of Energy contract number DE-AS05-80ER10666.A000.

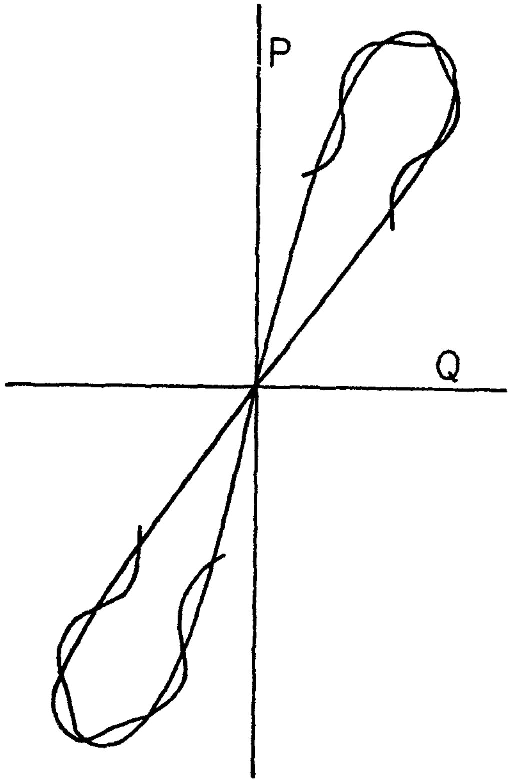


Figure (7.21): Schematic presentation of the intersections of the stable and unstable manifolds in the case of Fig. (7.20) leading to chaotic behavior.

REFERENCES

[1] H. Goldstein, Classical Mechanics 2nd ed. (Addison-Wesley, Reading, Massachusetts, 1980).

[2] J. Moser, M.V. Berry, J. Ford, Y.M. Treve et. al., Topics in Nonlinear Dynamics, AIP Conference Proceedings #46, Siebe Jorna, Ed. (1978).

[3] V. I. Arnold, Mathematical Methods of Classical Mechanics (Springer-Verlag, 1978).

[4] M. Hammermesh, Group Theory and its Applications to Physical Problems (Addison-Wesley, Reading, Massachusetts, 1962), p. 406.

[5] A.J. Dragt and J.M. Finn, J. Math. Phys. 17, 2215 (1976).

[6] N. Jacobson, Lectures in Abstract Algebra, Vol. II - Linear Algebra (D. Van Nostrand, Princeton, 1953), p. 188.

[7] A.J. Dragt and J.M. Finn. J. Geophys. Res. 81, 2327 (1976).

[8] A.J. Dragt and J.M. Finn, J. Math. Phys. 20, 2649 (1979).

[9] J. Moser, Stable and Random Motion in Dynamical Systems (Princeton University Press, 1973).

[10] In this discussion, the time t simply plays the role of a parameter. It is included in the notation to indicate that the transformation may depend on the time. That is, the transformation may be different at different times.

[11] C.L. Siegel and J.K. Moser, Lectures on Celestial Mechanics (Springer, New York, 1971).

[12] A.J. Dragt and J.M. Finn, J. Math. Phys. 17, 2215 (1976).

[13] The concept of an "adjoint" operator in the Lie algebraic context should not be confused with the concept of Hermitian adjoint often employed in the theory of linear operators on Hilbert spaces.

[14] In the general case, one may require a product of two exponentials. That is, Eq. (5.63) has the general form

$$\exp(:h_2^c:)\,\exp(:h_2^a:) = \exp(:f_2^c:) \times$$
$$\exp(:f_2^a:)\,\exp(:g_2^c:)\,\exp(:g_2^a:). \tag{5.92}$$

Similarly, Eq. (5.60) then takes the more general form

$$M^g = \exp(JS^a)\,\exp(JS^c). \tag{5.93}$$

With this understanding, Eq. (5.64) and all subsequent equations hold in the general case. It is also worth remarking that in many cases of practical interest a single exponential, which may not have either the "a" or "c" symmetry, is sufficient. See Ref. 12.

[15] D.R. Douglas and A.J. Dragt, IEEE Trans. Nucl. Sci. NS-28, 2522 (1981).

[16] H. Goldstein, Classical Mechanics, 2nd ed. (Addison-Wesley, 1980).

[17] H. Bruck, Circular Particle Accelerators, Los Alamos National Laboratory Translation of Accelerateurs Circulaires de Particules, Presses Universitaires de France (1966).

[18] B. Friedman, Principles and Techniques of Applied Mathematics, (John Wiley, New York, 1956), p. 92.

[19]A.J. Dragt, Exact Numerical Calculation of Chromaticity in Small Rings, Univ. of Maryland Physics Publication No. 81-115 (1981).
[20]A.J. Dragt, A Lie Algebraic Theory of Geometrical Optics and Optical Aberrations, J. Opt. Soc. Am. 72, 372 (1982).
[21]K.L. Brown, SLAC-75, Rev. 3 (1975).
[22]P. Henrici, Elements of Numerical Analysis, John Wiley, New York (1964).
[23]A. J. Dragt, Exact Numerical Calculation of Chromaticity in Small Rings, to appear in Particle Accelerators (1982).
[24]B. Friedman, Principles and Techniques of Applied Mathematics, John Wiley, New York (1956).
[25]For a discussion of some of the relevant mathematical background required for a proper treatment of this subject, see V. I. Arnold, Mathematical Methods of Classical Mechanics, page 228, Springer-Verlag, New York (1978); I. M. Gelfand, and V. B. Lidskii, Amer. Math. Soc. Translations, Series 2, Volume 8, page 143 (1958); V. A. Yakubovic, Amer. Math. Soc. Translations, Series 2, Volume 10, page 125 (1958).
[26]In this case, it is necessary to set up two surfaces of section. One is located at the beginning of the lattice period, and the other at its end. The Poincare map M now sends points in the initial surface of section at the beginning of the lattice period to image points in the final surface of section.
[27]A preliminary study of the four-dimensional case has been made by D. A. Edwards, and L. C. Teng, IEEE Trans. on Nucl. Sci. Vol. 20, No. 3, 885 (1973). See also A. J. Dragt, Symplectic Floquet Theory, manuscript in preparation.
[28]E. D. Courant, and H. S. Snyder, Annals of Physics 3, 1 (1958); H. Bruck, Circular Particle Accelerators, Chapters 4-8, Los Alamos National Laboratory technical report LA-TR-72-10 Rev., an English translation of Accelerateurs Circulaires de Particules, Presses Universitaires de France (1966).
[29]A. J. Dragt, and J. M. Finn, J. Math. Phys. 17, 2215 (1976).
[30]This assumption is made only for simplicity. In point of fact, the betatron oscillations can always be uncoupled by a suitable choice of coordinates.
[31]C. Bovet, R. Gouiran, I. Gumowski, and K. H. Reich, A Selection of Formulae and Data Useful for the Design of A. G. Synchrotrons, CERN/MPS-SI/Int. DL/70/4, page 17 (25 April 1970); J. C. Herrera, M. Month, and R. F. Peierls, Nonlinear Dynamics and the Beam-Beam Interaction, page 202, M. Month and J. C. Herrera, Edits., AIP Conf. Proc. No. 57, Amer. Inst. Phys., N.Y. (1979).
[32]G. Guignard, A General Treatment of Resonances in Accelerators, Report CERN 78-11 (1978).
[33]A. J. Dragt, IEEE Trans. Nuc. Sci. Vol. NS-26, No. 3, page 3601 (1979).
[34]The term beating range is also used to refer to a phenomenon which occurs in the more general problem of two degrees of freedom (with a four-dimensional phase space). In this case a resonance can cause a bounded periodic exchange of amplitude between the betatron oscillations in the two degrees of freedom.

[35]V. I. Arnold, Mathematical Methods of Classical Mechanics, page 394, Springer-Verlag, New York (1978); J. Moser, Stable and Random Motions in Dynamical Systems, Princeton University Press, Princeton, N.J. (1973).

[36]A. J. Dragt, and J. M. Finn, J. Geophys. Res. 81, No. 13, page 2327 (1976).

[37]In the absence of the existence of invariant functions, it is generally not possible within our present knowledge to make positive statements about long-term stability. If the use of the first few terms of h predicts instability, as in the unstable region of Fig. 7.4, then we may be sure that orbits in this region are in fact unstable. However, if stability is predicted by the first few terms of h, then we are in fact assured of stability only for moderately long periods of time. That is, given enough time, there is the possibility that homoclinic behavior will bring a point from what was judged to be the stable region into the unstable region. Thus, when nonlinear effects are taken into account, we are in the uncomfortable position of being generally unable to make any rigorous statements about very long term behavior. From this perspective, the fact that storage rings work so well must be viewed, at the very least, as a minor miracle.

Under certain conditions it can be shown in the one degree of freedom case (two-dimensional phase space) that there are still isolated closed invariant curves despite the nonexistence of invariant functions. The existence of these curves was first shown by Komolgorov, Arnold, and Moser, and consequently they are often referred to as KAM curves. See, for example, J. Moser, Stable and Random Motions in Dynamical Systems, Princeton University Press, Princeton, N.J. (1973). The existence of closed invariant curves would guarantee the stability of all orbits corresponding to initial conditions inside these curves. There is some optimism that such curves exist for regions of phase space of physical interest. However, the present mathematical proof of their existence holds only for regions of phase space much too small to be of physical significance.

[38]The map (7.84) was first studied by Henon in a different context. Figures 7.8 and 7.9 are taken from his work. See, M. Henon, Quarterly of Applied Mathematics 27, 291 (1969). For remarkably beautiful color pictures of phase-space plots of this map, see the work of M. Feigenbaum in Los Alamos Science, Volume 1, Number 1 (1980), published by the Los Alamos National Laboratory, Los Alamos, New Mexico.

[39]A. J. Dragt, and J. M. Finn, J. Math. Phys. 20, p. 2649 (1979).

[40]The remainder terms are all commutators which contain two or more f_b's and any number of g_2's. Therefore, these terms are of second order and higher in the beam-beam interaction strength. The terms of order two can be computed if needed. See Ref. 29.

[41]A. J. Dragt, Nonlinear Dynamics and the Beam-Beam Interaction, p. 143, M. Month and J. Herrera, Edits., Am. Inst. of Phys. Proceedings

No. 57 (1979); A. J. Dragt, and O. Jakubowicz, Analysis of the Beam-Beam Interaction Using Transfer Maps, Proceedings of the SLAC Symposium on the Beam-Beam Interaction, SLAC-Pub-2624, Stanford University, Stanford, California (1980).

2. J. C. Herrera, Nonlinear Dynamics and the Beam-Beam Interaction, p. 29, M. Month and J. C. Herrera, Edits., Am. Inst. Phys. Proceedings No. 57 (1979).

3. J. C. Herrera, M. Month, and R. F. Peierls, Nonlinear Dynamics and the Beam-Beam Interaction, p. 202, M. Month and J. C. Herrera, Edits., Amer. Inst. Phys. Proc. No. 57 (1979).