SLAC-PUB-9939
June 10, 2003

Symplecticity in Beam Dynamics: An Introduction*

John R. Rees

Stanford Linear Accelerator Center, Stanford University, Stanford,CA 94309

Abstract

A particle in a particle accelerator can often be considered a Hamiltonian system, and when that is the case, its motion obeys the constraints of the Symplectic Condition. This tutorial monograph derives the condition from the requirement that a canonical transformation must yield a new Hamiltonian system from an old one. It then explains some of the consequences of symplecticity and discusses examples of its applications, touching on symplectic matrices, phase space and Liouville's Theorem, Lagrange and Poisson brackets, Lie algebra, Lie operators and Lie transformations, symplectic maps and symplectic integrators.

*Work supported by Department of Energy contract DF-AC03-76SF00515.

Contents

The motions of any Hamiltonian system obey a constraint called the symplectic condition or—more succinctly if less euphoniously—symplecticity. The same constraint applies to the motions of the system, of course, regardless of the mathematical formulation of the dynamical problem (Newtonian, Lagrangian *etc.*), but the Hamiltonian formulation leads more expeditiously to a general expression of the symplectic condition than other formulations do, so it offers a clearer and more straightforward pedagogical setting. Symplecticity has important consequences for beam physics. Probably the most commonly cited is the conservation of the density of system points in phase space—Liouville's Theorem—but that is only one of many consequences. This monograph is an exploration of symplecticity, its mathematical origins and some of its consequences.

A particle in an accelerator or a beam-transport system can frequently be considered a Hamiltonian system,*i.e.,* a system obeying Hamilton's equations of motion applied to the Hamiltonian for a charged particle in a given electromagnetic field. Of course, that is an approximation that ignores synchrotron radiation, particle-particle interactions and a number of other effects, and the approximation will not be permissible if, for example, the particle's radiation or the particle interactions are strong enough to affect the motion appreciably. Nonetheless, the approximation is often good enough, and it finds widespread use in beam physics.

In the following pages, we first develop Hamilton's equations in their symplectic (matrix) form. Then after introducing canonical transformations in terms of generating functions, we derive the symplectic condition, which constitutes a set of constraints on the Jacobian matrix of any canonical transformation. The symplectic condition is a necessary and sufficient condition for a transformation to be canonical; therefore it supplies a direct test of any transformation between related sets of dynamic variables to determine whether or not the transformation is canonical. By building up the motion of a Hamiltonian

system from a series of infinitesimal motions, we can show that the motions of any Hamiltonian system are themselves canonical transformations from the initial conditions to the final state of motion. (This fact underlies the Hamilton-Jacobi method of solution.) Consequently, the motions of Hamiltonian systems are canonical transformations, and the symplectic condition applies to their Jacobian matrices. Specifically, the motions of beam particles must be symplectic.

Viewed as a purely mathematical statement, the symplectic condition defines a set of symplectic matrices, and we discuss some of the mathematical properties of these matrices. Then returning to physics, we discuss the restrictions on physical dimensions in canonical transformations. We discuss phase space, emphasizing that it is a space in which the state of an entire Hamiltonian is represented by a single point, and we the use of the phase space for a single particle to characterize a beam as a swarm of points.

In beam dynamics we often expand the particle motions as Taylor series in the initial values of the coordinates and momenta, and frequently, we can attain sufficient accuracy even if we discard all but the linear terms and express the result as a "transport matrix." That matrix is the Jacobian matrix of the solution, and thus the transport matrix is symplectic.

Lie algebraic structures have proved to be powerful tools in developing methods of finding the motions of Hamiltonian systems, especially when the systems are highly non-linear. We introduce the elements of this approach in the last two sections.

We conclude with an appendix devoted to the derivation of the the Hamiltonian for non-interacting particles moving in a given electromagnetic guide-field.

The topics chosen are those that caught the writer's interest as he studied symplecticity to explain the subject to himself. The references reflect the main published sources he studied, which were greatly augmented by very pleasant and illuminating discussions with Alex Chao and Ron Ruth.

A list of the main notational styles used is included for the reader's convenience in the following table.

NOTATION

$\dot{a}$	An overhead dot denotes total differentiation with respect to time: da/dt.
a'	A prime denotes total differentiation with respect to s: da/ds.
$\mathbf{A}$	Bold-face Roman typeface denotes a vector.
$\hat{\mathbf{a}}$	An overhead caret denotes a unit vector.
A	Bold-face *san serif* typeface denotes a matrix.
$\widetilde{\mathsf{A}}$	A tilde denotes a transposed matrix: $\widetilde{A}_{ij} = A_{ji}$.
$\lvert\mathsf{A}\rvert$	The determinant of the elements of a matrix.
S	A constant rearranging matrix. See (4) and (5).
$\partial f/\partial\mathbf{x}$	Differentiation of a function by a column matrix. See (8).
$/x/$	The physical dimensions of x.
$\{u, v\}$	Lagrange brackets. See (112).
$[u, v]$	Poisson brackets. See (117).
$:f:$	Lie operator associated with the function f. See (134).

1 The Symplectic Form of Hamilton's Equations of Motion

Hamilton's equations of motion can be written in a compact matrix formulation, and recasting them in that way is a good first step to introduce the study of symplecticity.[1] We begin with the equations themselves, which are normally written

$$\dot{q}_k = \frac{\partial H}{\partial p_k} \ , \qquad \dot{p}_k = -\frac{\partial H}{\partial q_k} \ , \qquad k = 1, 2, \cdots, N \ , \tag{1}$$

where overhead dots denote total differentiation with respect to time. These equations carry the implicit understanding that *time* is the independent variable. It is possible under certain circumstances to interchange the role of independent variable between time and one of the coordinates or momenta so that the latter can be treated as the independent variable, and time can be treated as a coordinate. The essential requirement for the feasibility of this procedure is that the variable chosen to replace time as independent variable must vary monotonically and smoothly with time. In particle-beam dynamics this interchange is customarily made between time and the longitudinal coordinate of the particle, called s, which is promoted to the status of independent variable, time becoming the third coordinate. This change of variables, its justification, and the derivation of the resulting Hamiltonian and equations of motion are discussed fully in the appendix. The resulting equations of motion, generalized to N degrees of freedom, are

$$q_k' = \frac{\partial \mathcal{H}}{\partial p_k} \ , \qquad p_k' = -\frac{\partial \mathcal{H}}{\partial q_k} \ , \qquad k = 1, 2, \cdots, N \ , \tag{2}$$

where primes denote total differentiation with respect to s, and the cursive $\mathcal{H}$ denotes the Hamiltonian in which s, the longitudinal coordinate, is the independent variable. *We shall take (2) as our fundamental equations of motion, and we shall always treat s as the independent variable unless we state otherwise.*

To cast these equations in matrix form, we first define x, a column matrix of the canonical variables, indicated by a bold *sans serif* typeface, with $(\mathsf{x})_i = q_{(i+1)/2}$ for odd i, $(\mathsf{x})_i = p_{i/2}$ for even i, and with i ranging from 1 to $2N$. It looks like this:

$$\mathsf{x} = \begin{pmatrix} q_1 \\ p_1 \\ q_2 \\ \vdots \\ q_N \\ p_N \end{pmatrix} . \tag{3}$$

The arrangement of this matrix has been chosen to accord with typical usage in accelerator beam dynamics—more on this subject at the end of the section.

Hamilton's equations of motion equate q' to derivatives of $\mathcal{H}$ with respect to p, and they equate p' to *negative* derivatives with respect to q. In order to

express the equations in terms of $\mathbf{x}$, a "rearranging matrix" is needed. To help construct the rearranging matrix, we introduce an antisymmetric two-by-two constant matrix, $\mathbf{s}$.

$$\mathbf{s} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{4}$$

The rearranging matrix, which we shall denote by a capital $\mathbf{S}$, can be constructed of $\mathbf{s}$ and the null matrix by placing s-matrix partitions along the diagonal up to the desired dimension and null matrix partitions everywhere else to form a $2N$ by $2N$ matrix.

$$\mathbf{S} = \begin{pmatrix} \mathbf{s} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{s} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{s} \end{pmatrix} \tag{5}$$

It is a very simple, sparse, square, even-dimensioned, antisymmetric matrix with only one entry in each row and one entry in each column. S-matrices have the following obvious properties:

$$\mathbf{S}^{-1} = \widetilde{\mathbf{S}} = -\mathbf{S}, \qquad \mathbf{S}^2 = -\mathbf{I}\ , \qquad |\mathbf{S}| = 1, \tag{6}$$

where the tilde denotes the transposed matrix, $\mathbf{I}$ is the identity matrix of appropriate order and $|\mathbf{A}|$ signifies the determinant of the matrix $\mathbf{A}$.

Using the matrix $\mathbf{S}$, we can now write Hamilton's equations of motion in compact form in terms of the column matrix $\mathbf{x}$.

$$\mathbf{x}' = \mathbf{S}\ \frac{\partial \mathcal{H}}{\partial \mathbf{x}}\ , \tag{7}$$

where we define the derivative of a scalar function, f, with respect to the column matrix $\mathbf{x}$ as another column matrix according to

$$\left(\frac{\partial f}{\partial \mathbf{x}}\right)_i \equiv \frac{\partial f}{\partial x_i}\ . \tag{8}$$

Equation 7 is called the *symplectic form of Hamilton's canonical equations of motion*. The term, "symplectic," comes from the Greek $\sigma\upsilon\mu\pi\lambda\epsilon\kappa\tau\iota\kappa os$ which, according to the Oxford English Dictionary, means twining, plaited together or copulative.

It is important to reëmphasize that the form of the matrix $\mathbf{S}$ was dictated entirely by the arbitrary organization of the column matrix $\mathbf{x}$. Our choice—alternating coordinates and momenta—although usually used in beam dynamics, differs from the usual practice in classical mechanics textbooks of grouping the coordinates together at the top and the momenta at the bottom. In the latter case, the matrix $\mathbf{S}$, is replaced by another, often called $\mathbf{J}$, which has a different form:

$$\mathbf{J} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix}, \tag{9}$$

where $\mathbf{I}$ is the unit matrix. Of course, these are merely notational differences; the dynamics are the same.

2 Canonical Transformations

In Hamiltonian mechanics the state of a system with N degrees of freedom is specified in terms of $2N$ variables—the canonically conjugate coordinates and momenta—and its evolution is described by the variation with s of those variables. One may choose the coordinates and momenta to use for a particular system to suit the symmetries of the system or according to habit. As simple examples, one might think of cartesian or spherical coordinates, each with appropriate conjugate momenta, to describe the motion of a particle. A canonical transformation is a set of equations that relates one of these sets to another, say a "new" set of $2N$ canonical variables, the Q's and P's, and an "old" set, the q's and p's. It may optionally involve also the independent variable s. Supposing we had found the motion of the system in terms of the old variables $q_1(s), q_2(s), \cdots, p_{(N-1)}(s), p_N(s)$, we could use the transformation equations to describe the motion in terms of the new variables.

The transformations may be written[1]

$$Q_k = Q_k(q, p, s) \ , \qquad P_k = P_k(q, p, s) \ , \qquad k = 1, 2, \cdots, N \ , \tag{10}$$

$2N$ equations in all. Here and subsequently the argument list notation (q, p, s) symbolizes the full list $(q_1, p_1, \cdots, q_N, p_N, s)$. In the special case that the transformation (10) does not depend explicitly on the independent variable, it is designated a *restricted canonical transformation*. The old set of variables obeys (2), and since the new set is canonical, there must exist a corresponding new Hamiltonian, call it $K(Q, P, s)$, such that

$$Q'_k = \frac{\partial K}{\partial P_k} \ , \qquad P'_k = -\frac{\partial K}{\partial Q_k} \ , \qquad k = 1, 2, \cdots, N \ , \tag{11}$$

Furthermore the function K must somehow be produced from the old Hamiltonian by the process of the transformation itself by a scheme that is independent of the system under consideration. In other words the process of canonical transformation includes more than transforming the canonical variables; it includes transforming the Hamiltonian.

A common method of making a canonical transformation is through the agency of a *generating function*, an arbitrary but well-behaved function of the generalized coordinates, the canonical momenta, and the independent variable.[1] The coordinates and the momenta of the generating function may not come from the same side of the transformation, *i.e.*, some must come from the old variables and some from the new, but the selection must be such as to specify the state of the system completely. Four combinations are used in four kinds of generating functions numbered one through four:

$$F_1(q, Q, s) \ , \quad F_2(q, P, s) \ , \quad F_3(p, Q, s) \ , \quad F_4(p, P, s) \ . \tag{12}$$

These arguments assume that the transformation can be uniquely inverted, but transformations with that property are the only ones that are useful for our purposes. The corresponding canonical transformations are given by the following

four sets of equations.

$$p_k = +\frac{\partial F_1}{\partial q_k} \ , \qquad P_k = -\frac{\partial F_1}{\partial Q_k} \ , \qquad K = \mathcal{H} + \frac{\partial F_1}{\partial s} \ , \tag{13}$$

$$p_k = +\frac{\partial F_2}{\partial q_k} \ , \qquad Q_k = +\frac{\partial F_2}{\partial P_k} \ , \qquad K = \mathcal{H} + \frac{\partial F_2}{\partial s} \ , \tag{14}$$

$$q_k = -\frac{\partial F_3}{\partial p_k} \ , \qquad P_k = -\frac{\partial F_3}{\partial Q_k} \ , \qquad K = \mathcal{H} + \frac{\partial F_3}{\partial s} \ , \tag{15}$$

$$q_k = -\frac{\partial F_4}{\partial p_k} \ , \qquad Q_k = +\frac{\partial F_4}{\partial P_k} \ , \qquad K = \mathcal{H} + \frac{\partial F_4}{\partial s} \ . \tag{16}$$

These recipes produce strictly canonical transformations as they are defined in (50). There is a type of transformation, called a scale transformation, that satisfies the requirements that the transformation equations take the form (10) and that a new Hamiltonian is obeyed by the system, but scale transformations are not usually admitted as canonical transformations. More is said about scale transformations in Section 3 and in Reference [1]. The formulae above cannot produce scale transformations.

A few examples will be useful in fixing the idea and the application of the generating-function method in mind.

Example 1—The beam-dynamical Hamiltonian $\mathcal{H}$ given in equation (228) in the appendix is expressed in terms of the canonical variables x, P_x, y, P_y, t and $-H$, where t is the time and H is the total energy of the particle. A more useful set of variables for beam dynamics replaces time and energy with their increments from the time and energy of the reference particle. As an example of the generating-function procedure we shall transform to that set. The new time coordinate is $\tau = t - t_0(s)$, where $t_0(s)$ is the schedule of the progress of the reference particle along the reference trajectory, and the conjugate momentum (an energy) is $-\epsilon = -H + H_0$ where $H_0(s)$ is the energy of the reference particle. Both $t_0(s)$ and $H_0(s)$ are given *a priori*. The time-difference variable measures the lead or lag in the arrival at s of the particle in question relative to that of the reference particle. To summarize, the old variables q, p are $t, -H$ and the new variables Q, P are $\tau, -\epsilon$. Using a generating function of the third type,

$$F_3 = (H - H_0)\,(t_0(s) + \tau) \ , \tag{17}$$

the transformation equations are

$$t = -\frac{\partial F_3}{\partial(-H)} = t_0 + \tau, \tag{18}$$

$$-\epsilon = -\frac{\partial F_3}{\partial \tau} = -H + H_0 \ . \tag{19}$$

The new Hamiltonian becomes

$$K(x, P_x, y, P_y, \tau, -\epsilon) = \mathcal{H} + \frac{\partial F_3}{\partial s} = -(1 + hx)\Psi + \frac{\epsilon}{v_0} \ , \tag{20}$$

where

$$\Psi = \sqrt{\left(\frac{H_0 + \epsilon - e\Phi}{c}\right)^2 - (P_x - eA_x)^2 - (P_y - eA_y)^2 - m^2c^2} + eA_s \ . \quad (21)$$

In the result, all occurrences of t are replaced by $t_0 + \tau$ and all occurrences of H by $H_0 + \epsilon$. The variables of both $\mathcal{H}$ and K are canonical, *i.e.*, they form sets of canonically conjugate pairs. That qualifies either set to form the dimensions of a single-particle phase space in terms of which some important general principles, such as Liouville's Theorem, can be asserted.

Example 2—Another instructive example of the generating-function method of making canonical transformations is one that transforms the relativistic Hamiltonian for a charged particle from cartesian coordinates (in which it is usually expressed) to the curvilinear coordinates of beam dynamics.[2] See the appendix in which the coordinates are defined and in which we derive the same beam dynamical Hamiltonian by a different procedure which is more perspicuous but more laborious.

We begin with the familiar form of the Hamiltonian in cartesian coordinates with time as the independent variable,

$$H(\mathbf{r}, \mathbf{p}, t) = c\sqrt{(\mathbf{p} - e\mathbf{A})^2 + m^2c^2} + e\Phi \ , \quad (22)$$

The old coordinates are the cartesian coordinates and the corresponding conjugate momenta are the cartesian components of the kinetic momentum plus the corresponding components of the vector potential (times the charge). The new variables are x, y, s as defined at the beginning of the appendix and their conjugate momenta. We choose again a generating function of the third type:

$$F_3 = -\mathbf{p} \cdot [\mathbf{r}_0(s) + \hat{\mathbf{x}}(s)x + \hat{\mathbf{y}}(s)y] \ , \quad (23)$$

where $\mathbf{p}$ is the conjugate momentum vector in the cartesian system and the other quantities are defined at the beginning of the appendix where the curvilinear coordinate system is discussed. The transformations, (15), give the results

$$\mathbf{r}(s) = \mathbf{r}_0(s) + \hat{\mathbf{x}}(s)x + \hat{\mathbf{y}}(s)y \ , \quad (24)$$

the desired relation between old and new coordinates, and

$$P_x = \mathbf{p} \cdot \hat{\mathbf{x}} \ , \quad P_y = \mathbf{p} \cdot \hat{\mathbf{y}} \ , \quad P_s = \mathbf{p} \cdot \hat{\mathbf{s}}(1 + hx) \ , \quad (25)$$

for the new momenta. Since the generating function has no explicit time dependence, the transformations are of the restricted variety, and $K = H$ with the new variables substituted according to (25). The transformed Hamiltonian is

$$K = c\sqrt{(P_x - eA_x)^2 + (P_y - eA_y)^2 + \left(\frac{P_s}{1 + hx} - eA_s\right)^2 + m^2c^2} \quad + \quad e\Phi \quad (26)$$

the same as (222) of the appendix.

Both of the preceding examples have a noteworthy characteristic: The generating functions (17) and (23) are real functions of real variables and therefore generate transformations from real variables to real variables. No such restriction is placed on generating functions by the theory. Indeed, complex generating functions and the resulting complex transformations are useful in some contexts, but when they are used, supplementary rules must be carried along to translate the results into real, physical variables. The point to remember is that the theory of canonical transformations does not prohibit or eliminate complex results; that is left to the practitioner.

The transformation between old and new variables in *Example 1* is intrinsically unique, because it is linear. A description of the system in terms of the old variables $(t, -H)$ defines a single, unique description in the new variables $(\tau, -\epsilon)$ and *vice versa*. In other words, these transformations are one-to-one mappings. The generating-function method does not guarantee this property. While the generating function must possess continuous second derivatives to be useful, it is otherwise utterly arbitrary.[1] Consequently it is entirely possible to define a generating function with benign behavior in every respect except that of securing uniqueness of the mathematical transformations it generates. Indeed generating functions of higher degree ordinarily do not automatically generate one-to-one mappings.

If the transformation $Q_k = Q_k(q, p, s),\ P_k = P_k(q, p, s)$ is non-linear, *i.e.,* is composed of non-linear functions of (q, p), then a single point in (q, p)-space may correspond mathematically to multiple points in (Q, P)-space, and *vice versa*. Such cases often arise in non-linear beam dynamics, a field of vigorous activity, and recipes must be used in these cases to select the inverse transformation that leads back to the original point. *Example 3* explicates these methods.

Example 3—As a simple yet practical example consider the generating function $F_2(q, P) = P(q + aq^2)$ which generates the following "forward" transformation from the q-p phase space to the Q-P phase space.

$$Q = q + aq^2, \qquad\qquad P = p/(1 + 2aq). \tag{27}$$

Clearly these expressions map one point in q-p phase space to a single point in Q-P phase space. But what of the inverted relations? The coordinate q is a solution of the quadratic equation $aq^2 + q - Q = 0$, which has two distinct roots that generate two distinct "backward" transformations which we can label with a and b.

$$q_a = (-1 + \sqrt{1 + 4aQ})/2a, \qquad p_a = P\sqrt{1 + 4aQ}, \tag{28}$$

$$q_b = (-1 - \sqrt{1 + 4aQ})/2a, \qquad p_b = -P\sqrt{1 + 4aQ}. \tag{29}$$

Let us consider the forward transformation of the point $q = 0,\ p = 0$. It transforms to the point $Q = 0,\ P = 0$ in the Q-P phase space, and only the "a" transformation carries this point back to the point $q = 0,\ p = 0$ as

required. Thus we must eliminate the "b" transformation and choose the "a" transformation as the proper backward transformation.

Canonical transformations possess the *group* property; they form a group. A group is a set of abstract elements with a law of combination to form a *product* such that (a) products obey the associative law, (b) every product of two elements and the square of each element is a member of the set, (c) the set contains the unit element which, combined in a product with any other element in either order, yields the other element itself, and (d) every element has an inverse, also a member of the set, so that the product of the element with its inverse yields the unit element. For canonical transformations, each transformation $(q, p) \longrightarrow (Q, P)$ as a whole is an element, and the product operation is called *composition*. It is the successive application of two transformations, $(q, p) \longrightarrow (Q, P)$ followed by $(Q, P) \longrightarrow (\bar{Q}, \bar{P})$ which can be abbreviated $(q, p) \longrightarrow (Q, P) \longrightarrow (\bar{Q}, \bar{P})$. Since the variables $(\bar{Q}, \bar{P})$ obey a Hamiltonian, the transformation $(q, p) \longrightarrow (\bar{Q}, \bar{P})$ is canonical and thus a member of the set. A tripartite product does not depend on the order in which the transformations are carried out, so the associative law applies. The identity transformation exists trivially. And finally, the inverse of the transformation $(q, p) \longrightarrow (Q, P)$ is simply the inversion of it $(Q, P) \longrightarrow (q, p)$. We say that canonical transformations form a group under composition, a group of infinite extent. Anticipating the next section, we may note that to every canonical transformation there belongs a square Jacobian matrix; we shall find that those Jacobian matrices also form an infinite group.

3 The Symplectic Condition

Canonical transformations are most often introduced in the context of generating function theory as we have done in the preceding section. Starting from a description of the motion in terms of one set of canonical variables, one writes down a generating function and then carries out the manipulations prescribed in the theory, whereby one arrives at a new set of canonical variables and a new Hamiltonian. The transformation is guaranteed to be canonical. The generating function method is a constructive method. On the other hand, suppose we are confronted *a priori* with an arbitrary transformation which may or may not be canonical. How can we find out whether it is or is not? Using generating function theory we could only depend on the *deus ex machina* while guessing at generating functions and trying them. If we found one that generated the transformation in question, we would have proved that the transformation is canonical, but if we failed to find one we could not thereby prove that the appropriate generating function does not exist—only that we had not found it. In other words, to test an arbitrary transformation we need an analytic method rather than a constructive one. The symplectic condition answers this need.

Stated in words, the symplectic condition requires that the Jacobian matrix of any canonical transformation be a symplectic matrix, which in turn means

that the Jacobean matrix must satisfy a certain equation involving only the Jacobian matrix and the rearranging matrix, $\mathbf{S}$, introduced at the beginning of the monograph. The Jacobian matrix is a matrix of derivatives of the new coordinates relative to the old ones, similar in form to the well-known Jacobian determinant. It is not unreasonable that it should be constrained by a condition that selects only canonical transformations from all possible transformations.

Again letting the old variables be designated q and p and the new ones Q and P, a general transformation between them would be expressed

$$Q_i = Q_i(q, p, s) \ , \tag{30}$$

$$P_i = P_i(q, p, s) \ , \tag{31}$$

or as the inverted relationship

$$q_i = q_i(Q, P, s) \ , \tag{32}$$

$$p_i = p_i(Q, P, s) \ , \tag{33}$$

where the coordinate and momentum symbols in the argument lists are meant to imply the whole sets. To repeat, the motion of the system in terms of the old coordinates and momenta obeys Hamilton's equations of motion,

$$q_i' = \frac{\partial \mathcal{H}}{\partial p_i} \quad , \quad p_i' = -\frac{\partial \mathcal{H}}{\partial q_i} \ . \tag{34}$$

If the transformation is canonical, there exists some function, $K(Q, P, s)$ such that the equations of motion in the new coordinates are

$$Q_i' = \frac{\partial K}{\partial P_i} \quad , \quad P_i' = -\frac{\partial K}{\partial Q_i} \ , \tag{35}$$

and such that K is arrived at from the old Hamiltonian $\mathcal{H}$ through the procedure of the transformation.

Returning to matrix notation and letting

$$\mathbf{X} = \begin{pmatrix} Q_1 \\ P_1 \\ Q_2 \\ P_2 \\ \vdots \end{pmatrix} \ , \tag{36}$$

the reciprocal transformations (30), (31) and (32), (33) take the forms

$$\mathbf{X} = \mathbf{X}(\mathbf{x}, s) \qquad \text{and} \qquad \mathbf{x} = \mathbf{x}(\mathbf{X}, s) \ , \tag{37}$$

simply symbolizing the fact that the elements of $\mathbf{X}$ are functions of the elements of $\mathbf{x}$ *etc.* If the transformation is canonical, in addition to (7) we have also

$$\mathbf{X}' = \mathbf{S} \frac{\partial K}{\partial \mathbf{X}} \ , \tag{38}$$

and we may take the existence of a function $K(\mathbf{X}, s)$ such that these are the equations of motion as the touchstone of a canonical transformation. The fact that $\partial^2 K/\partial x_i \partial x_j = \partial^2 K/\partial x_j \partial x_i$ must hold for any i and j turns out to impose strong constraints on the Jacobian matrix of the transformation.

The total derivative of $\mathbf{X}$ with respect to s is given by

$$\mathbf{X}' = \frac{\partial \mathbf{X}}{\partial \mathbf{x}}\mathbf{x}' + \frac{\partial \mathbf{X}}{\partial s} = \mathbf{M}\mathbf{S}\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \frac{\partial \mathbf{X}}{\partial s} \ , \tag{39}$$

where we have introduced $\mathbf{M}$, the *Jacobian matrix of the first of the transformations* (37) with respect to the components of $\mathbf{x}$:

$$M_{ij} = \left(\frac{\partial \mathbf{X}}{\partial \mathbf{x}}\right)_{ij} \equiv \frac{\partial X_i}{\partial x_j} \ . \tag{40}$$

Combining (38) and (39) we obtain

$$\frac{\partial K}{\partial \mathbf{X}} = \widetilde{\mathbf{S}}\left(\mathbf{M}\mathbf{S}\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \frac{\partial \mathbf{X}}{\partial s}\right) \ . \tag{41}$$

Then using the Jacobian relation

$$\frac{\partial K}{\partial \mathbf{x}} = \widetilde{\mathbf{M}}\frac{\partial K}{\partial \mathbf{X}} \ , \tag{42}$$

we arrive at the column-matrix equation

$$\frac{\partial K}{\partial \mathbf{x}} = \widetilde{\mathbf{M}}\mathbf{S}\mathbf{M}\widetilde{\mathbf{S}}\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \widetilde{\mathbf{M}}\widetilde{\mathbf{S}}\frac{\partial \mathbf{X}}{\partial s} \ . \tag{43}$$

This matrix equation represents a set of $2N$ equations of the form

$$\frac{\partial K}{\partial x_j} = a_j + \sum_{k=1}^{2N} b_{jk}\frac{\partial \mathcal{H}}{\partial x_k} \ , \tag{44}$$

where $\mathbf{a}$ and $\mathbf{b}$ are given by

$$\mathbf{a} = \widetilde{\mathbf{M}}\widetilde{\mathbf{S}}\left(\frac{\partial \mathbf{X}}{\partial s}\right) \qquad \text{and} \qquad \mathbf{b} = \widetilde{\mathbf{M}}\mathbf{S}\mathbf{M}\widetilde{\mathbf{S}} \ , \tag{45}$$

a column matrix and a $2N$-by-$2N$ matrix respectively. These matrices depend only on the transformation, not on the Hamiltonian. In other words, they are independent of the system under consideration. Equations (44) are simultaneously satisfied if the transformation is canonical.

In order to be a Hamiltonian, K must be a simple scalar function for which $\partial^2 K/\partial x_i \partial x_j = \partial^2 K/\partial x_j \partial x_i$ for any i and j.[3] From this condition we determine that

$$\left(\frac{\partial a_j}{\partial x_i} - \frac{\partial a_i}{\partial x_j}\right) + \sum_{k=1}^{2N}\left(\frac{\partial b_{jk}}{\partial x_i} - \frac{\partial b_{ik}}{\partial x_j}\right)\frac{\partial \mathcal{H}}{\partial x_k} + c_{ij} = 0 \tag{46}$$

for any function $\mathcal{H}$ whatever, where

$$c_{ij} = \sum_{k=1}^{2N} \left(b_{jk} \frac{\partial^2 \mathcal{H}}{\partial x_i \partial x_k} - b_{ik} \frac{\partial^2 \mathcal{H}}{\partial x_j \partial x_k} \right) \ , \tag{47}$$

The condition means that, *for any i and j*:

1. The first parenthetical expression in (46) must vanish.

2. The sum in the second term of (46) must vanish, but since the partial derivatives are independent, their coefficients (the parentheses) must individually vanish.
$$\frac{\partial b_{jk}}{\partial x_i} = \frac{\partial b_{ik}}{\partial x_j}$$

3. The sum c_{ij} must vanish. In order to deduce the implications of that requirement, let us first consider the case $i \neq j$. In compiling the sum there will be a term
$$\left(b_{jj} \frac{\partial^2 \mathcal{H}}{\partial x_i \partial x_j} - b_{ij} \frac{\partial^2 \mathcal{H}}{\partial x_j^2} \right)$$
when $k = j$ and a term
$$\left(b_{ji} \frac{\partial^2 \mathcal{H}}{\partial x_i^2} - b_{ii} \frac{\partial^2 \mathcal{H}}{\partial x_j \partial x_i} \right)$$
when $k = i$. These are the only terms in the sum in which the mixed derivative $\partial^2 \mathcal{H} / \partial x_i \partial x_j$ appears, and the coefficient of that derivative must vanish. These terms are also the only ones in which the second derivatives with respect to the particular variables x_i and x_j appear, and their coefficients must also vanish. Therefore
$$b_{ij} = 0 \ , \qquad b_{ii} = b_{jj} \ , \qquad \text{for all} \ \ i \neq j \tag{48}$$
For completeness we note that all the c_{ii} vanish identically.

It follows that $\mathbf{b}$ is diagonal and all the diagonal elements are equal; in other words, $\mathbf{b}$ is a multiple of the unit matrix: $\mathbf{b} = \lambda(\mathbf{x}, s)\ \mathbf{I}$. Then from Point (2) above we deduce that the function is independent of the coordinates and momenta and is therefore a function of s alone. The deduction goes as follows. The condition requires that each diagonal element b_{kk} depend only upon on the variable bearing the same index x_k; but the diagonal elements must all be the same; therefore they may depend on none of the canonical variables, only on the independent variable: $\lambda = \lambda(s)$. But it is possible to show from the definitions of $\mathbf{a}$ and $\mathbf{b}$, (45), that $\partial b_{ij}/\partial s = \partial a_j / \partial x_i - \partial a_i / \partial x_j$. Therefore $\partial b_{ij}/\partial s = 0$ and λ is merely a constant. We conclude that $\widetilde{\mathbf{M}}\mathbf{S}\mathbf{M}\widetilde{\mathbf{S}} = \lambda \mathbf{I}$ which leads to the most general statement of the result:

$$\widetilde{\mathbf{M}}\mathbf{S}\mathbf{M} = \lambda \mathbf{S} \ . \tag{49}$$

We may call this equality (with Goldstein[1]) the *extended symplectic condition.*

The terms "canonical transformation" and "symplectic condition" as we have used them—and indeed as most authors use them—allow only transformations for which $\lambda = 1$, so the equation,

$$\widetilde{\mathbf{M}}\mathbf{S}\mathbf{M} = \mathbf{S} \ , \tag{50}$$

expresses the *symplectic condition.* It is a constraint upon the Jacobian matrix of a canonical transformation. In fact, it is a necessary and sufficient condition for all canonical transformations.[1] If a transformation satisfies the symplectic condition, it is canonical, and if a transformation is canonical, it satisfies the symplectic condition. The symplectic condition provides us with a direct test of canonicality of a transformation.

Along with the symplectic condition we get the following conditions on the new Hamiltonian: applying $\mathbf{b} = \mathbf{I}$ to (43), we get a set of simultaneous partial differential equations.

$$\frac{\partial K}{\partial \mathbf{x}} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}} - \widetilde{\mathbf{M}}\mathbf{S}\frac{\partial \mathbf{X}}{\partial s} \ , \tag{51}$$

relating the new Hamiltonian K to the old Hamiltonian $\mathcal{H}$ and the given canonical transformation $\mathbf{X}(\mathbf{x}, s)$. If the transformation (37) has no explicit s-dependence, the second term on the right vanishes, and we can satisfy the conditions by choosing $K = \mathcal{H}$, just as we do in the generating-function method when the generating function has no explicit s-dependence. Otherwise, we have to augment $\mathcal{H}$ by the quantity $-\widetilde{\mathbf{M}}\mathbf{S}(\partial\mathbf{X}/\partial s)$ to produce K. If we define the augmentation as Δ, so that $K = \mathcal{H} + \Delta$, then

$$\Delta = -\widetilde{\mathbf{M}}\mathbf{S}(\partial\mathbf{X}/\partial s), \tag{52}$$

and we emphasize the important fact that the augmentation of the Hamiltonian depends only on the canonical transformation itself and not on the Hamiltonian.

For the sake of reminding ourselves of the meaning of the Jacobian matrix, it is illuminating to write it out in terms of the canonical variables themselves:

$$\mathbf{M} = \begin{pmatrix} \partial Q_1/\partial q_1 & \partial Q_1/\partial p_1 & \partial Q_1/\partial q_2 & \cdots \\ \partial P_1/\partial q_1 & \partial P_1/\partial p_1 & \partial P_1/\partial q_2 & \cdots \\ \partial Q_2/\partial q_1 & \partial Q_2/\partial p_1 & \partial Q_2/\partial q_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} . \tag{53}$$

Symplecticity constrains these derivatives, which are, in general, functions of the canonical coordinates x_i and of s. Therefore the symplectic condition must be read like this: $\widetilde{\mathbf{M}}(\mathbf{x}, s)\mathbf{S}\mathbf{M}(\mathbf{x}, s) = \mathbf{S}$. It requires each of the elements of the product matrix on the left to be a constant for *all values* of the x_i's and s. In the special case that the transformation $\mathbf{X}(\mathbf{x}, s)$ is linear in the x_i's, the Jacobian matrix of the transformation is independent of the x_i's, but it still may depend on s.

A useful type of transformation called the *scale transformation* can always be used to transform a system satisfying the extended symplectic condition (49) into one satisfying (50):

$$Q_i = \mu q_i, \quad P_i = \nu p_i, \quad K = \mu\nu\mathcal{H} \ , \tag{54}$$

where μ and ν are constants, *i.e.*, independent of the canonical variables and of the independent variable, and $\mu\nu = \lambda$. Indeed we may consider that we ruled out these transformations in setting λ equal to one in order to reach (50).

Example 4—In Example 1, Sec. 2, we developed the Hamiltonian for a beam particle in terms of the canonical variable set, $x,\ P_x,\ y,\ P_y,\ \tau$ and $-\epsilon$. Another widely used set of variables is one called the "Transport variables," so-called because this set is used in a computer program for beam-transport design called `Transport`.[4] Here is a case in which we have, on one hand, the set of variables above, that we know to be canonical and, on the other hand, a set given to us solely in terms of the transformation between the two sets. Are the Transport variables a canonical set? Answering this question is an excellent task for the symplectic condition. The Transport variables are

- $\mathtt{x}$: The horizontal coordinate, same as x.
- $\mathtt{x}'$: The horizontal slope of the particle's trajectory.
- $\mathtt{y}$: The vertical coordinate, same as y.
- $\mathtt{y}'$: The vertical slope of the particle's trajectory.
- ℓ : The longitudinal separation from the reference particle.
- δ : The fractional momentum deviation from the reference particle.

With the help of the symplectic condition we can test whether the transformation to these variables is canonical, and therefore whether these variables form a canonical set. The transverse Transport variables, which we have symbolized by a `typewriter font` to distinguish them from the standard variables, are related to the canonical variables above by the following relations.

$$\mathtt{x} = x \tag{55}$$

$$\mathtt{x}' = \frac{\partial K}{\partial P_x} \tag{56}$$

$$\mathtt{y} = y \tag{57}$$

$$\mathtt{y}' = \frac{\partial K}{\partial P_y} \tag{58}$$

where K is the Hamiltonian, (20). The transverse momenta are supplanted by the transverse slopes $\mathtt{x}'$ and $\mathtt{y}'$. Quite naturally the question arises, do these variables form canonical pairs? Is $\mathtt{x}'$ canonically conjugate to $\mathtt{x}$, and is $\mathtt{y}'$ canonically conjugate to $\mathtt{y}$? Since these variables are used primarily in linearized treatments of beam dynamics, let us expand K to second order, and let us omit accelerating sections from consideration and confine ourselves to beam transport magnets that can be characterized by purely longitudinal vector potentials so that $A_x = A_y = 0$. Also let us symbolize the second-order Taylor expansion (in x and y) of $A_s(x, y, s)$ by $S(x, y, s)$. The Hamiltonian then takes the form,

$$K_{2nd} = -p_0 + \frac{P_x^2}{2p_0} + \frac{P_y^2}{2p_0} + \frac{m^2\epsilon^2}{2p_0^3} - \frac{H_0 h x \epsilon}{p_0 c^2} + eS(x, y, s)\ . \tag{59}$$

From this Hamiltonian we find the new canonical momenta.

$$\mathtt{x}' = \frac{P_x}{p_0} \ , \tag{60}$$

$$\mathtt{y}' = \frac{P_y}{p_0} \ . \tag{61}$$

The transformations of the longitudinal variables are given by

$$\ell = -v\tau = -c\tau\sqrt{1 - \left(\frac{mc^2}{H_0 + \epsilon}\right)^2}, \tag{62}$$

$$\delta = \frac{p - p_0}{p_0} = \frac{1}{p_0}\sqrt{\left(\frac{H_0 + \epsilon}{c}\right)^2 - m^2c^2} - 1 \ . \tag{63}$$

Summarizing the transformation: the transverse coordinates do not change; the new transverse momenta are mere constant scalings of the old momenta; and the new longitudinal variables depend only on the old longitudinal variables. The new variable ℓ is negatively proportional to the old variable τ because τ is negative when the particle in question leads the reference particle, *i.e.,* arrives earlier. For this transformation, the Jacobian matrix is very sparse. We can partition it into two-by-two submatrices, and only the submatrices on the diagonal are non-zero; the Jacobian matrix is a block-diagonal matrix.

$$\mathsf{M} = \begin{pmatrix} \mathsf{M}_x & \mathsf{O} & \mathsf{O} \\ \mathsf{O} & \mathsf{M}_y & \mathsf{O} \\ \mathsf{O} & \mathsf{O} & \mathsf{M}_s \end{pmatrix} \ , \tag{64}$$

where the submatrices are

$$\mathsf{M}_x = \mathsf{M}_y = \begin{pmatrix} 1 & 0 \\ 0 & 1/p_0 \end{pmatrix} \ , \quad \mathsf{M}_s = \begin{pmatrix} -v & (m^2c^6\tau)/v(H_0 + \epsilon)^3 \\ 0 & -1/(vp_0) \end{pmatrix} \tag{65}$$

For a matrix like (64), the symplectic condition reduces to the requirement that the determinants of the submatrices each be unity. In fact, the three determinants are equal, but their common value is not unity, it is $1/p_0$. The transformation, therefore, is not strictly canonical.

It can, however, be considered an extended canonical transformation. For extended transformations, the determinants of the diagonal submatrices are merely required to be constants, but all three determinants must be the same constant. That criterion is satisfied: the determinants are all $1/p_0$. Then, since we are considering only beam transport in magnets, and p_0, the momentum of the reference particle, is constant in magnets, we may in fact regard the transformation to Transport variables to be an extended canonical transformation. Indeed we might have noted at the outset that the transformation from P_x to $\mathtt{x}'$ (for example) changes the dimensions of the momentum from kinetic momentum to unity, so the appearance of a constant with the dimensions of momentum

was inevitable. To summarize: the set $\{\mathtt{x}, \mathtt{x}', \mathtt{y}, \mathtt{y}', \ell, \delta\}$ is "extended canonical" in transport systems in which the momentum of the particle is constant.

Of course, we have reached these results on the basis of the linearized Hamiltonian (59), having dodged the exact Hamiltonian (20) at the outset. On the basis of the exact Hamiltonian, the relationships between the set $\{\mathtt{x}, \mathtt{x}', \mathtt{y}, \mathtt{y}', \ell, \delta\}$ and the "old" canonical variable set, $\{x,\ P_x,\ y,\ P_t,\ \tau, -\epsilon\}$ are not so simple as those we found with the approximate Hamiltonian, and the transformation is not canonical or even extended canonical.

Also in this example, by taking p_0 to be constant, we have excluded accelerating systems like rf cavities and traveling-wave accelerators, in which the momentum of the reference particle is not constant. The transformation to Transport variables in those regions is not canonical, even on the basis of the linearized Hamiltonian. One consequence of that fact is that Liouville's Theorem (which is discussed in Section 8) does not hold for the Transport variables in such regions, and the phenomenon generally called "adiabatic damping" takes place. In adiabatic damping, acceleration of the reference particle is accompanied by an increase in the density of particle points in the transverse pseudo-phase planes $(\mathtt{x}, \mathtt{x}')$ and $(\mathtt{y}, \mathtt{y}')$.

Example 5—In the preceding example, the transformations of the variables involved only the old and the new canonical variables; they did not depend explicitly on the independent variable s. In that case the new Hamiltonian was simply the old Hamiltonian re-expressed in terms of new variables instead of the old ones, and we had no need to use the formula (51), which comes into play only when the transformations depend explicitly on the independent variable. Now let us construct a simple example in which the transformations do depend on the independent variable, and for the sake of familiarity, let us temporarily take the independent variable to be time. Equation (51) will be valid for this case with the simple substitution of t for s.

We don't need a very elaborate system to illustrate the method. Let us consider motion in one degree of freedom along the x-axis. Let x be the coordinate of a particle of mass m in a stationary (inertial) coordinate system S_1, and let ξ be the coordinate in a relatively moving system S_2 which coincides with S_1 at $t = 0$ but which is accelerating to the right at rate a so that the coordinate of a point x in S_1 is related to the coordinate of the same point in S_2 by the equation,

$$x = \xi + at^2/2. \tag{66}$$

The non-relativistic Hamiltonian of a free particle in the S_1 system is

$$H = \frac{p_x^2}{2m}, \tag{67}$$

where $p_x = m\dot{x}$. Now suppose we are given *ex cathedra* the following transformation.

$$\xi = x - at^2/2, \qquad p_\xi = p_x. \tag{68}$$

The questions are: is this transformation canonical, and if it is, what is the new Hamiltonian K for a free particle?

First, we must form the Jacobian matrix, $\mathbf{M} = (\partial\mathbf{X}/\partial\mathbf{x})$, of the transformation.

$$\mathbf{x} = \begin{pmatrix} x \\ p_x \end{pmatrix} \qquad \text{and} \qquad \mathbf{X} = \begin{pmatrix} \xi \\ p_\xi \end{pmatrix} = \begin{pmatrix} x - at^2/2 \\ p_x \end{pmatrix}.$$

The Jacobian turns out to be the unit matrix, $\mathbf{M} = \mathbf{I}$, so the transformation is symplectic, and therefore canonical. That being the case, we turn to the prescription (51) for the new Hamiltonian,

$$\frac{\partial K}{\partial \mathbf{x}} = \frac{\partial H}{\partial \mathbf{x}} - \widetilde{\mathbf{M}}\mathbf{S}\frac{\partial \mathbf{X}}{\partial t}, \tag{69}$$

in which we have replaced s with t. If now we let $K = H + F$, the partial differential equations we have to solve are simplified to

$$\frac{\partial F}{\partial \mathbf{x}} = -\widetilde{\mathbf{M}}\mathbf{S}\frac{\partial \mathbf{X}}{\partial t} = -\mathbf{S}\frac{\partial \mathbf{X}}{\partial t}. \tag{70}$$

Differentiating the transformation equations with respect to time, we find

$$\frac{\partial \mathbf{X}}{\partial t} = \begin{pmatrix} -at \\ 0 \end{pmatrix}, \tag{71}$$

and then combining the last two equations we arrive at the partial differential equation for F.

$$\frac{\partial F}{\partial p_x} = -at, \tag{72}$$

with the solution,

$$F = -atp_x. \tag{73}$$

Then adding F to H and using the transformation to express the result in terms of the new variables, we arrive at the new Hamiltonian,

$$K = \frac{p_\xi^2}{2m} \; - \; atp_\xi. \tag{74}$$

The equations of motion in the accelerated frame S_2 are

$$\dot{\xi} = \frac{p_\xi}{2m} - at \qquad \text{and} \qquad \dot{p}_\xi = 0. \tag{75}$$

The canonical momentum in the accelerated system is constant at its initial value, and the particle is uniformly accelerated at rate $-a$. The solution for the coordinate is $\xi = (p_{\xi 0}/m)t - at^2/2$. Viewed in the inertial system S_1, $x = (p_{x0}/m)t$, which is the correct motion for a free particle.

Before leaving the subject of the symplectic condition, one further important fact needs to be discussed, and we state the fact before proving it: *the solutions*

of Hamilton's equations of motion—the motions of Hamiltonian systems—are themselves canonical transformations and therefore obey the symplectic condition.

For any system, the Hamiltonian is a function of the canonical coordinates and momenta and the independent variable (and no other variables), so application of the equations of motion leads to a system of $2N$ differential equations of the form

$$\varphi_i' = f_i(\varphi_1, \varphi_2, \ldots, \varphi_{2N}, s) \ , \qquad i = 1, 2, \ldots, 2N \tag{76}$$

with initial conditions $\varphi_i(0) = y_{0i}$ which specify a single point in phase space. If the functions f_i are well-behaved, we know from the theory of differential equations that this system has a unique solution set,

$$\varphi_i(s; y_{01}, y_{02}, \cdots, y_{02N}) \ , \qquad i = 1, 2, \ldots, 2N \ , \tag{77}$$

determined by the initial conditions.[6] In the $2N$-dimensional space of the variables, the solution describes a curve parametrized by s that passes through the point specified by the initial conditions, and each different initial point defines a different curve.

Casting the statements of the preceding paragraph in terms of x, our vector of canonical variables defined in (3), the solution (77) takes the form $\mathsf{x}(s) = \mathsf{x}(s; \mathsf{x}_0)$. Clearly we may regard any such functional relation as an s-dependent transformation: x_0 is transformed to $\mathsf{x}(s)$. But is it a canonical transformation? The answer is, yes, and to prove it we shall construct an infinitesimal transformation. Consider the motion of a system from s_0 to $s_0 + \delta s$. In the infinitesimal interval δs the coordinates and momenta change by the increments $q'\delta s$ and $p'\delta s$, so if Q and P denote the "new" values (at $s_0 + \delta s$), we can consider the motion a transformation and write it $Q_k = q_k + (\partial\mathcal{H}/\partial p_k)\delta s$ and $P_k = p_k - (\partial\mathcal{H}/\partial q_k)\delta s$ or in matrix form

$$\mathsf{X} = \mathsf{x} + \mathsf{S}\ \frac{\partial \mathcal{H}}{\partial \mathsf{x}}\ \delta s \ , \tag{78}$$

where x and X are defined in (3) and (36). Since this is an infinitesimal transformation, results should retain terms only to first order in δs. We can determine whether (78) is a canonical transformation by computing its Jacobian matrix and testing it against the symplectic condition (50).

$$\mathsf{M} = \frac{\partial \mathsf{X}}{\partial \mathsf{x}} = \mathsf{I} + \mathsf{S}\ \frac{\partial^2 \mathcal{H}}{\partial \mathsf{x} \partial \mathsf{x}}\ \delta s \ , \tag{79}$$

where the second derivatives,

$$\left(\frac{\partial^2 \mathcal{H}}{\partial \mathsf{x} \partial \mathsf{x}} \right)_{kj} = \frac{\partial^2 \mathcal{H}}{\partial x_k \partial x_j} \ , \tag{80}$$

form a symmetric matrix. Applying the symplectic test,

$$\widetilde{\mathsf{M}}\mathsf{S}\mathsf{M} = \left(\mathsf{I} - \frac{\partial^2 \mathcal{H}}{\partial \mathsf{x} \partial \mathsf{x}} \mathsf{S} \delta s \right) \mathsf{S} \left(\mathsf{I} + \mathsf{S} \frac{\partial^2 \mathcal{H}}{\partial \mathsf{x} \partial \mathsf{x}} \delta s \right) = \mathsf{S} \tag{81}$$

to first order in δs, and we conclude that the infinitesimal transformation (78) is a canonical transformation.

On the basis of that finding, and recognizing that we can build up the motion over a finite interval from s_0 to s by adding up or integrating infinitesimal transformations, we conclude that the evolution with s of the solution from its initial value $\mathbf{x}_0$ to the present value is itself an s-dependent canonical transformation, $\mathbf{x}(\mathbf{x}_0, s)$. This conclusion will not surprise students of the Hamilton-Jacobi technique of solving the equations of motion, because that technique is based on finding just such canonical transformations. It is however an important and powerful result. It means that if we have found a solution to the equations of motion, $\mathbf{X}(\mathbf{x}_0, s)$, where the elements of $\mathbf{x}_0$ are the initial conditions, the Jacobian, $(\partial\mathbf{X}/\partial\mathbf{x}_0)$, must satisfy the symplectic condition.

4 Symplectic Matrices

In a purely mathematical sense, (50) defines a set of even-dimensioned square matrices which form a group (called appropriately the symplectic group) just as the associated canonical transformations form a group. Since for every canonical transformation, there exists a corresponding symplectic (Jacobian) matrix, the group property of the canonical transformations implies the group property of the matrices. However it is straightforward to demonstrate the group properties of the matrices directly by verifying that they satisfy the following conditions.

1. Multiplication is associative: $(\mathbf{M}_1\mathbf{M}_2)\mathbf{M}_3 = \mathbf{M}_1(\mathbf{M}_2\mathbf{M}_3)$, true for matrices in general.

2. Every product of two elements and the square of each element is a member of the set.
$$(\widetilde{\mathbf{M}_1\mathbf{M}_2})\mathbf{S}(\mathbf{M}_1\mathbf{M}_2) = \widetilde{\mathbf{M}}_2\widetilde{\mathbf{M}}_1\mathbf{S}\mathbf{M}_1\mathbf{M}_2 = \mathbf{S}$$

3. The set contains the unit element. The unit matrix is symplectic.

4. Every element has an inverse such that $\mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$. Since their determinants do not vanish, all symplectic matrices have inverses.

The symplectic condition (50) constitutes a fairly substantial set of constraints, one of the most important of which concerns the determinant of the matrix. Since the determinant of the product of two square matrices of equal dimension is equal to the product of their determinants, the square of the determinant of a symplectic matrix, and therefore of the Jacobian matrix of a canonical transformation, is equal to unity.

$$|\mathbf{M}|^2 = 1 \tag{82}$$

Symplectic matrices also obey the following relations.

$$\mathbf{M}^{-1} = -\mathbf{S}\widetilde{\mathbf{M}}\mathbf{S}\ , \qquad\qquad \mathbf{M}\mathbf{S}\widetilde{\mathbf{M}} = \mathbf{S} \tag{83}$$

Equation (82) is a main factor in demonstrating Liouville's Theorem, to which we shall turn in Section 6. Although this constraint allows the determinant to be either +1 or -1, we are interested in the former case, because as we saw in the preceding section, he evolution of the motion of a Hamiltonian system from some starting point s_0 to some other point s_f is itself a canonical transformation. Clearly such transformations must approach identity as s_f approaches s_0, and the determinant of $\mathbf{M}$ must be unity. We shall assume from this point forward that

$$|\mathbf{M}| = 1 \ . \tag{84}$$

For a two-by-two symplectic matrix, the determinant condition is the sole constraint; it completely determines the symplecticity of the matrix. In other words, for a single degree of freedom, uncoupled to any other, the determinant condition is equivalent to the symplectic condition.

For matrices of higher order than two, the symplectic condition implies more constraints among the matrix elements than the one expressed by (82). It turns out that the number of independent matrix elements in a symplectic matrix of order $2N$ is $N(2N+1)$, so the number of constraints among the $(2N)^2$ elements is

$$\text{number of constraints}: \quad N(2N-1) \tag{85}$$

The assertion that there are $N(2N+1)$ independent elements in the matrix rests on the following considerations.[5] We first define the exponential matrix and the logarithm of a matrix by analogy with the corresponding functions.

$$\exp(\mathbf{A}) = \sum_{n=0}^{\infty} \mathbf{A}^n/n! \ , \quad \log(\mathbf{A}) = -\sum_{n=1}^{\infty} \frac{(\mathbf{I}-\mathbf{A})^n}{n} \tag{86}$$

where $\mathbf{A}$ is a square matrix. Clearly these functions are also square matrices of the same order as their arguments. Since $\widetilde{(\mathbf{A}^n)} = (\widetilde{\mathbf{A}})^n$, we have $\widetilde{\exp \mathbf{A}} = \exp \widetilde{\mathbf{A}}$ in view of which it can be shown that any real symplectic matrix $\mathbf{M}$ sufficiently near the identity can be written

$$\mathbf{M} = \exp(\mathbf{S}\mathbf{A}_s) \ , \tag{87}$$

where $\mathbf{A}_s$ is a real *symmetric* matrix and $\mathbf{S}$ is of the same order as $\mathbf{A}_s$. For matrices near the identity, this establishes a one-to-one relationship between real symplectic matrices and real symmetric matrices. Since a real symmetric matrix of order $2N$ has $N(2N+1)$ independent elements, so has a real symplectic matrix sufficiently near the identity. To put it in terms of constraints, the number of constraints created by the symplectic condition is, as stated above, $4N^2 - N(2N+1) = N(2N-1)$: for one degree of freedom, one constraint; for two, six constraints; for three, fifteen constraints; *etc*. Finally, although these conclusions apply only to matrices "sufficiently close to the identity" (and which therefore have determinants of unity), those matrices form a group, and the group can be extended to a global group by repeatedly multiplying those already obtained. Consequently (85) applies to all symplectic matrices with determinant plus unity—those of primary interest to beam dynamicists.

Another way of representing symplectic matrices in exponential form is through polar decomposition. Any real nonsingular matrix $\mathbf{A}$ can be written uniquely in the form $\mathbf{A} = \mathbf{PO}$, where $\mathbf{P}$ is real positive definite symmetric matrix, and $\mathbf{O}$ is a real orthogonal matrix. (An orthogonal matrix is one that obeys $\widetilde{\mathbf{O}}\mathbf{O} = \mathbf{I}$.) The uniqueness of the decomposition implies that to a given matrix $\mathbf{A}$, there corresponds only one $\mathbf{P}$ and only one $\mathbf{O}$. Thus we can express $\mathbf{M}$ in polar terms.

$$\mathbf{M} = \mathbf{PO}. \tag{88}$$

Applying the symplectic condition and appealing to the uniqueness of the decomposition, we can show that both $\mathbf{P}$ and $\mathbf{O}$ are themselves symplectic.

As a consequence, it turns out that $\mathbf{O}$ and $\mathbf{P}$ can be written

$$\mathbf{O} = \exp(\mathbf{S}\mathbf{A}_s^c), \qquad \mathbf{P} = \exp(\mathbf{S}\mathbf{A}_s^a), \tag{89}$$

where $\mathbf{A}_s^c$ is a real symmetric matrix that commutes with $\mathbf{A}_s$, and $\mathbf{A}_s^a$ is a real symmetric matrix that anticommutes with $\mathbf{A}_s$.

$$\mathbf{A}_s^c\mathbf{S} - \mathbf{S}\mathbf{A}_s^c = \mathbf{0}, \qquad \mathbf{A}_s^a\mathbf{S} + \mathbf{S}\mathbf{A}_s^a = \mathbf{0}, \tag{90}$$

Any symmetric matrix $\mathbf{S}$ can be decomposed into a sum of such commuting and anticommuting matrices, $\mathbf{S} = \mathbf{A}_s^c + \mathbf{A}_s^a$, as follows.

$$\mathbf{A}_s^c = (\mathbf{S} - \mathbf{S}^{-1}\mathbf{S}\mathbf{S})/2, \qquad \mathbf{A}_s^a = (\mathbf{S} + \mathbf{S}^{-1}\mathbf{S}\mathbf{S})/2. \tag{91}$$

The conclusion is that any symplectic matrix can be written in the form

$$\mathbf{M} = \exp(\mathbf{S}\mathbf{A}_s^a)\exp(\mathbf{S}\mathbf{A}_s^c), \tag{92}$$

where $\mathbf{A}_s$ is a real symmetric matrix.

In beam dynamics, we are primarily interested in *real* coordinate systems, and the Jacobian matrices of transformations between real systems are real matrices. Consequently the symplectic matrices we deal with are also real, and the eigenvalues of real, symplectic matrices form quite a restricted set.[5] The eigenvalues are the roots of the characteristic polynomial,

$$P(\lambda) = |\mathbf{M} - \lambda\mathbf{I}| \ , \tag{93}$$

a polynomial whose degree is equal to the dimension of $\mathbf{M}$ $(2N)$. The characteristic equation is

$$P(\lambda) = 0. \tag{94}$$

We observe immediately that $\lambda = 0$ cannot be a root. This equation has $2N$ roots, and the product of the roots is equal to the constant term of the polynomial.

$$\lambda_1\lambda_2\cdots\lambda_{2N} = P(0) = \det(\mathbf{M}) = 1 \tag{95}$$

Either the eigenvalues are real or they occur in complex conjugate pairs. Appealing only to the properties of symplectic matrices, it can also be shown that

$$\lambda^{-N}P(\lambda) = \lambda^N P(1/\lambda) \ , \tag{96}$$

with the consequence that if λ is a root, so is its reciprocal $1/\lambda$, and furthermore, these roots have the same multiplicities. If either +1 or -1 is a root, then that root has even multiplicity. To reiterate, *the properties of the eigenvalues of real, symplectic matrices are the following.*

1. They are real or they occur in complex conjugate pairs.
2. They occur in reciprocal pairs, each member of a pair having the same multiplicity.
3. If either +1 or -1 is an eigenvalue, it has even multiplicity.

These properties have important consequences in particle beam dynamics, especially for the stability of the motions in storage rings. Considering only transverse motion (x and y), the four eigenvalues must lie on the unit circle, because if any eigenvalue lies off it, there must be two such eigenvalues, one of which lies outside the unit circle and represents unstable motion. This is a very general conclusion and it applies whether there is coupling or not. For a thorough explanation of the case of linear optics with coupling see Reference [2].

5 Dimensions in Canonical Transformations

We now have three ways of knowing that the set of coordinates and momenta (P, Q) are canonically conjugate:

1. If we begin with a Lagrangian couched in a generalized coordinate system in which the coordinates are the Q's, and we define the momenta according to $P_k = \partial L/\partial \dot{Q}_k$ and the Hamiltonian by $H = \sum_k P_k \dot{Q}_k - L$ following the standard prescription, then we can be sure that the Q's and P's are canonically conjugate. In other words, we have created the canonical momentum from its basic definition.
2. Starting with a canonical description of the system in terms of a set of canonically conjugate variables (q, p), if we carry out a canonical transformation *via* the generating-function formalism described in Section 2 to a new set (Q, P), we may be sure the latter set is canonically conjugate. We go from one set of canonical variables to another *via* a canonical transformation.
3. Given a transformation $Q = Q(q, p, s), P = P(q, p, s)$, if the Jacobian matrix $\mathbf{M}$ (defined in (36) and (40)) satisfies the symplectic condition $\widetilde{\mathbf{M}}\mathbf{S}\mathbf{M} = \mathbf{S}$, then the Q's and P's are canonically conjugate. This procedure uses the direct condition.

What do these methods imply about the dimensional properties of canonical transformations?

In the case of method 1 the implications are clear. The Lagrangian function L is understood by definition to have the dimensions of energy, so the conjugate

momenta have the dimensions of energy divided by the generalized velocity, the time rate of change of the variable in question. We have had no need to consider the independent variable of the Lagrangian to be anything but time.

Method 2 however might not be so restrictive. Let us introduce the following notation for dimensions: $/x/$ means "the dimensions of x." Consulting (13), (14), (15) and (16), we derive the following rules about dimensions.

$$/q_k//p_k/ = /Q_k//P_k/ = /F/ = /\mathcal{H}//s/ \tag{97}$$

For any degree of freedom, the product of the coordinate dimensions and the momentum dimensions must be the same before and after the transformation; moreover that product must be equal to the product of the dimensions of the Hamiltonian (ordinarily, energy) and length. For example, for a single degree of freedom of a particle, for which $q = x$, the cartesian coordinate, and $p = m\dot{x}$, The product of the dimensions is ml^2/t, a momentum times a length. In the appendix we learn that $\mathcal{H}$ has the dimensions of a linear momentum, so $/qp/ = /\mathcal{H}s/$. In this case, the implication is that a canonical transformation will necessarily lead to variables, the product of whose dimensions will be momentum times length.

In Section 8, we shall express the symplectic condition in terms of the fundamental Lagrange and Poisson brackets, and from those, it will be immediately clear that method 3. leads to the same rules as those expressed in (97).

For the extended symplectic condition (49), $\widetilde{\mathbf{M}}\mathbf{S}\mathbf{M} = \lambda\mathbf{S}$, the rules are less strict, as we might expect.

$$/Q_k//P_k/ = /\lambda//q_k//p_k/ \tag{98}$$

Freer changes of dimension are tolerated, but restrictions remain. We saw an instance of such restrictions in the example at the end of Section 3, in which the transformations $(x, P_x) \to (\mathtt{x}, \mathtt{x}')$, $(y, P_y) \to (\mathtt{y}, \mathtt{y}')$ and $(\tau, -\epsilon) \to (\ell, \delta)$ are independent of each other so that the Jacobian matrix (64) is block diagonal. In that case, each transformation obeys (98), but note that λ must be the *same* in all three cases.

The identity transformation, $Q_k = q_k$, $P_k = p_k$ for all k, is generated by either $F_2 = \sum_k q_k P_k$ or $F_3 = -\sum_k p_k Q_k$, and these functions, both of the first degree in each variable, obviously satisfy the rule (97). But what of algebraic terms of higher degree as generating functions, terms of the form $aq_k^m P_k^n$ for example? In such terms, the constant coefficient a can be given appropriate dimensions to adjust the dimensions of the whole term to satisfy the rule.

6 Phase Space and Liouville's Theorem

The concept of phase space was introduced by J. W. Gibbs around the turn of the twentieth century to characterize the full dynamical state of a mechanical system such as a confined gas. For a system of N particles (each having no internal degrees of freedom), Gibbs' phase space has $6N$ dimensions, one for

each of the six canonically conjugate coordinates and momenta of each particle. Thus a system of 10^{10} particles is described by a *single point* in a phase space of 6×10^{10} dimensions, and each distinct point specifies a distinct dynamical state or "phase" of the system. The Hamiltonian of the system involves the coordinates and momenta of all the particles, and it dictates the motion of that single phase point in the phase space.

In Gibbs' phase space a collection of points, called an *ensemble*, represents a collection of possible *states of the total system*; and the distributions of such points as they move through phase space are used in probabilistic calculations in statistical mechanics. The Hamiltonian may include terms that account for interactions between the particles of the system—collision terms in a gas, for example. But it is important to remember that different points in Gibbs' phase space represent different possible realizations of the whole system, and that the points, being purely mathematical objects, do not in any sense "interact."

In the preceding section, we observed that the solution (77) of the equations of motion describes a trajectory in phase space which passes through the point defined by the initial conditions, and that furthermore that trajectory, and therefore that solution, is unique. Since each system trajectory is unique and distinct, different trajectories do not cross; they have no common points. To prove that assertion, suppose two trajectories did intersect so that they had at least one point in common. Then considering that common point as an initial condition, it must lead to at least two different trajectories from the same initial conditions—two different solutions from the same initial conditions. That cannot happen, because the solution is unique. Therefore the trajectories do not intersect.

Liouville's theorem states that *in any system governed by a Hamiltonian, the density of system points surrounding a particular system point in phase space must remain constant as the independent variable evolves.* As the system evolves, the "particular system point" moves through phase space, so the theorem refers to the density in a chunk of volume moving through phase space. The proof of the theorem depends upon the symplecticity of the trajectories.

Although originally applied to Gibbs' phase space in which the Hamiltonian depends on all $6N$ dynamic variables, the same reasoning can be applied to any phase space in which the motions of all the individual points are governed by the same Hamiltonian. An example is a phase space of six dimensions in which each of N non-interacting particles of a beam is represented by a different point. Each particle obeys a Hamiltonian of the same form expressed in terms of the six canonical variables of that particle. In this phase space, a given particle point is surrounded by a swarm of particle points representing particles physically near the given particle. Each of these particles moves in accordance with Hamilton's equations applied a Hamiltonian of the same form, and as we saw in Section 7, the motions are canonical transformations from the initial values of the coordinates and momenta to the present values, and these trajectories in phase space are symplectic.

Now consider a small region (a volume in the six-dimensional phase space) surrounding the given particle point, and imagine that the volume is bounded

by a surface made of points that move (like the given point) in accordance with the common Hamiltonian. Some number of actual particle points lies inside the volume, and since the points cannot leave the volume without intersecting the wall points, they cannot leave the volume: the number of points in the volume remains fixed.

The "walls" of the region may twist and change shape, of course, but as we learned in Section 4, the Jacobian determinant of a symplectic transformation is equal to one, and that fact ensures that the volume of the region will be preserved.[1] Therefore the density of points in the vicinity of the given point—each representing a different particle—is a constant. The theorem may be considered a consequence of the symplectic condition in that we appeal in its proof to the constancy of the Jacobian determinant, in turn a consequence of symplecticity.

In the physics of particle beams, this kind of six-dimensional space spanned by the canonically conjugate coordinates and momenta of a single particle is commonly called phase space, although that usage is at variance with the Gibbsian usage in statistical mechanics described above. In this phase space a beam is represented by a multiplicity of particle points, whose motions through the phase space are governed by the Hamiltonians of the individual particles. Now in the case of non-interacting particles, the Hamiltonians of all of the particles are of the same form and are expressed in term of the six dimensions of the phase space, so the reasoning of Liouville's theorem applies. If the particles do not interact, the particle points representing different particles and the points bounding the volume all move in accordance with the same Hamiltonian, so the density of particle points in the neighborhood of a particular particle remains constant.

This is an important result. The distribution of a beam in the beam dynamical phase space is typified by its compactness in at least some dimensions, a consequence of the highly organized nature of a beam, and the preservation of the density within the distribution is often of crucial importance in beam physics.

Clearly if the motions of the beam particles are governed by the Hamiltonian derived in the appendix which takes no account of mutual interactions among the particles, this reasoning applies. But what if the particles interact? If the interactions are represented in the Hamiltonian for each particle by a sum of terms in the canonical variables of all the other particles, the Hamiltonian has $2N$ variables, and the corresponding phase space has $2N$ dimensions. A single point in that phase space specifies the state of the whole system—all N particles—and we are back to a Gibbsian phase space. We cannot interpret multiple points as representing the physical density of particles in the beam.

If the beam particles emit appreciable radiation, the Hamiltonian must have

[1] transformation between N functions of N or more variables such as $U_k = U_k(u_1, u_2, \ldots, u_N, s)$ and its inverse transformation $u_k = u_k(U_1, U_2, \ldots, u_N, s)$ both assumed to be single-valued; infinitesimal volumes in the spaces of the two sets of variables are related by

$$dU_1 dU_1 \ldots dU_N = |\det(\mathbf{M})| du_1 du_2 \ldots du_N$$

where the vertical lines mean "the absolute value."

canonical variables representing the degrees of freedom of the electromagnetic field, and the corresponding symplectic condition involves those variables as well as the particle coordinates. Again there is no "Liouville theorem."

7 Linear Approximations

The properties of typical beam transport systems and the beams they transport often suggest using linear approximations in computing the motions—linear in the coordinates and their derivatives and in the energy deviation from that of the reference particle. The efficacy of these approximations rests on the following considerations.

1. The position variables of a particle are small in comparison to some scale distance characteristic of the physical dimensions of a magnet or an accelerating cavity. In a quadrupole magnet, for example, the scale distance is determined by the degree to which the field dependence on transverse coordinates departs from the desired linear dependence. It depends on the quality of the quadrupole. In good quadrupoles, the distance is typically many meters. In accelerating cavities, the scale distance of the electric and magnetic fields is of the order of the free-space wavelength corresponding to the operating frequency.

2. The ratio of the horizontal coordinate to the bending radius is very small compared to unity, or in the terms defined in the appendix, $hx \ll 1$ where h is the reciprocal of the reference trajectory's bending radius.

3. The derivatives x' and y' are very small compared to unity. This is assured by the nature of particle beams, in which the transverse velocities are much smaller than the longitudinal ones.

4. The energy of a particle deviates only slightly from the energy of the reference particle.

In fact linear approximations yield descriptions of the particle motion quite accurate enough for many purposes in accelerator physics. Moreover, results of higher accuracy are often obtained by perturbation techniques which employ the linear approximation as their point of departure.

There are several ways of "linearizing" the dynamical problem. Beginning with the exact Hamiltonian, (a) we might linearize the Hamiltonian itself, *i.e.*, expand it in a Taylor series in the six canonical variables and truncate the series appropriately; or (b) we might first apply Hamilton's equations of motion to the exact Hamiltonian and then linearize the resulting differential equations; or, finally, (c) we might find the exact solution for the exact Hamiltonian and then linearize that solution. We can see quickly that the first two procedures

lead to the same result by expanding the Hamiltonian as follows.

$$\mathcal{H} = \sum_{k=1}^{6} \left(h_k x_k + \frac{1}{2} \sum_{j=1}^{6} h_{kj} x_k x_j \right) + G(\mathbf{x}) \tag{99}$$

where x is defined in (3), $h_k = (\partial \mathcal{H}/\partial x_k)_0$, $h_{kj} = h_{jk} = \left(\partial^2 \mathcal{H}/\partial x_k \partial x_j\right)_0$ and the function G embodies the remainder of the terms of the expansion. The constant term is omitted, since it does not affect the equations of motion; it only sets the zero point of the momentum scale. Applying the equations of motion we get expressions like

$$x_1' = h_2 + \sum_{j=1}^{6} h_{2j} x_j + \partial G/\partial x_2 \ , \tag{100}$$

$$x_2' = -h_1 - \sum_{j=1}^{6} h_{1j} x_j - \partial G/\partial x_1 \ , \tag{101}$$

and so forth. But we can see immediately that the leading terms on the right in these equations of motion must be zero, for the equations apply to any particle including the reference particle. Since we have chosen the trajectory of the reference particle as the longitudinal axis, all coordinates and all slopes for the reference particle are zero, and it follows that $h_1 = h_2 = \cdots = 0$. Consequently the linearized forms of the Hamiltonian and the equations of motion are the following.

$$\mathcal{H} = \frac{1}{2} \sum_{k=1}^{6} \sum_{j=1}^{6} h_{kj} x_k x_j \tag{102}$$

$$x_1' = \sum_{j=1}^{6} h_{2j} x_j \ , \tag{103}$$

$$x_2' = -\sum_{j=1}^{6} h_{1j} x_j \ , \tag{104}$$

It clearly made no difference whether we dropped the term G in the Hamiltonian or its derivatives in the equations of motion; we get the same linearized equations of motion. It is worth noting, by the way, that the theory of linear differential equations assures us that the solutions of (103) and (104) are linear in the initial values of the variables.

It is natural now to ask, do we also get the same answer if we linearize the exact solution, case (c)? To answer the question, we must first clarify what is meant by "linearizing" it. The exact solution is not a function of the canonical variables *per se*. Rather it is a function of the initial values of those variables and of s, and when we speak of expanding it, we mean expanding it in the initial conditions. So the question becomes this: If the solutions are expanded in the

initial conditions, is the linear approximation of the exact solution of the exact equations of motion the same as the exact solution of the linearized equations of motion? The answer is: yes. To demonstrate why, we note first that the expansion of the equations of motion can be cast in matrix form.

$$\mathbf{x}' = \mathbf{V}\mathbf{x} + \text{higher order terms} \ , \tag{105}$$

where the matrix $\mathbf{V}$ is given by

$$\mathbf{V} = \begin{pmatrix} h_{21} & h_{22} & \cdots \\ -h_{11} & -h_{12} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \ . \tag{106}$$

The elements of $\mathbf{V}$ are functions of s. To expand the solution in the initial conditions, we designate the initial values by the letter z: $z_1 = x_1(0)$, $z_2 = x_2(0)$, *etc.*, and form them into a column matrix so that the initial value of $\mathbf{x}$ is $\mathbf{z}$. A Taylor expansion gives

$$\mathbf{x} = \mathbf{R}\mathbf{z} + \text{higher order terms} \ , \tag{107}$$

where the elements of $\mathbf{R}$, a square matrix, are functions of s. No leading constant term appears because the reference particle, for which $\mathbf{z} = 0$, follows the s-axis. Combining (105) and (107) and collecting terms, we obtain

$$(\mathbf{R}' - \mathbf{V}\mathbf{R})\mathbf{z} + \text{higher order terms} = 0 \ , \tag{108}$$

an equation which must be satisfied for any arbitrary choices of the z_k's. It is immediately evident that

$$\mathbf{x} = \mathbf{R}\mathbf{z} \tag{109}$$

is the linear approximation to the exact solution with $\mathbf{R}$ satisfying the linear differential equation,

$$\mathbf{R}' = \mathbf{V}\mathbf{R} \ , \tag{110}$$

with the initial condition $\mathbf{R}(0) = \mathbf{I}$. Clearly it is also true that (109) is the exact solution of the linearized equations of motion $\mathbf{x}' = \mathbf{V}\mathbf{x}$, so we conclude that all three methods of linearizing the problem yield the same solution, and moreover, the linearized solution satisfies the symplectic condition since it is the exact solution of a Hamiltonian, *viz.*, (102).

The Jacobian matrix of the linear approximation is simply the matrix we have called $\mathbf{R}$,

$$\mathbf{M} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} = \mathbf{R} \ , \tag{111}$$

so $\mathbf{R}$ is a symplectic matrix.

8 Lagrange and Poisson Brackets

The symplectic condition (50) is closely related to the fundamental Lagrange brackets. Both Lagrange brackets and the more useful Poisson brackets are

canonical invariants; that is, they are invariant under canonical transformations.[2] Consider a set of $2N$ well-behaved functions $u(q,p), v(q,p), w(q,p), \cdots$ of a set of canonical variables, where the expression (q,p) symbolizes the whole argument list $(q_1, p_1, q_2, p_2, \cdots, q_N, p_N)$. In principle, these relations can be inverted to give the $2N$ canonical variables in terms of u, v, etc. Then the definition of the Lagrange bracket $\{u,v\}$ of u and v is

$$\{u,v\}_{qp} = \sum_{i=1}^{N} \left(\frac{\partial q_i}{\partial u}\frac{\partial p_i}{\partial v} - \frac{\partial p_i}{\partial u}\frac{\partial q_i}{\partial v} \right) . \tag{112}$$

The Lagrange brackets are antisymmetric in their arguments

$$\{u,v\}_{qp} = -\{v,u\}_{qp},$$

and of course $\{u,u\} = 0$. If $\{u,v\}_{qp}$ is evaluated in another system (Q,P) related by a canonical transformation to (q,p), it yields the same result:

$$\{u,v\}_{QP} = \{u,v\}_{qp},$$

so it makes no difference in what system it is calculated. Let us consider the Lagrange brackets of the set of coordinates (Q,P):

$$\{Q_m, P_n\}_{qp} = \sum_{i=1}^{N} \left(\frac{\partial q_i}{\partial Q_m}\frac{\partial p_i}{\partial P_n} - \frac{\partial p_i}{\partial Q_m}\frac{\partial q_i}{\partial P_n} \right) . \tag{113}$$

Thanks to the invariance of the Lagrange brackets, these brackets can equally well be evaluated in the (QP) system where the calculation is trivial. The result is given the special name, *fundamental Lagrange brackets*, and their values are as follows.

$$\{Q_m, Q_n\} = 0, \qquad \{P_m, P_n\} = 0, \qquad \{Q_m, P_n\} = -\{P_n, Q_m\} = \delta_{mn}. \tag{114}$$

Since the subscripts are immaterial, we omit them. These equations give the fixed values of the fundamental Lagrange brackets in any system of canonical coordinates whatsoever.

We can cast the fundamental Lagrange brackets in a symplectic matrix form as follows. If $\mathbf{M}$ is taken to be the Jacobian matrix of the transformation $q_1 = q_1(Q,P,s),\ p_1 = p_1(Q,P,s), \cdots$, *i.e.*,

$$M_{11} = \frac{\partial q_1}{\partial Q_1}, \quad M_{12} = \frac{\partial q_1}{\partial P_1}, \quad M_{13} = \frac{\partial q_1}{\partial Q_2}, \cdots$$

the inverse transformation to that referred to in (53), then by straightforward

[2] Goldstein remarks in his Second Edition (1980) [1] that the Lagrange brackets are now "mainly of historical interest," although they served as an introduction to Poisson brackets in the first edition (1951).

multiplication,

$$\widetilde{\mathsf{M}}\mathsf{S}\mathsf{M} = \begin{pmatrix} \{Q_1,Q_1\} & \{Q_1,P_1\} & \{Q_1,Q_2\} & \cdots & \cdots \\ \{P_1,Q_1\} & \{P_1,P_1\} & \{P_1,Q_2\} & \cdots & \cdots \\ \{Q_2,Q_1\} & \{Q_2,P_1\} & \{Q_2,Q_2\} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \{P_N,Q_N\} & \{P_N,P_N\} \end{pmatrix} . \quad (115)$$

Since the brackets are invariant, we need not specify system in which they were evaluated, *i.e.*, $\{Q_1,Q_1\} = \{Q_1,Q_1\}_{qp} = \{Q_1,Q_1\}_{QP}$ *etc.* If we evaluate them in the (QP) system—the system in which the matrix elements are just the fundamental Lagrange brackets—we find the matrix equation,

$$\widetilde{\mathsf{M}}\mathsf{S}\mathsf{M} = \mathsf{S} \ . \quad (116)$$

We discover that the symplectic condition is the matrix form of the fundamental Lagrange brackets.

We can proceed similarly with the Poisson brackets:

$$[u,v]_{qp} = \sum_{i=1}^{N} \left(\frac{\partial u}{\partial q_i}\frac{\partial v}{\partial p_i} - \frac{\partial u}{\partial p_i}\frac{\partial v}{\partial q_i} \right) \ . \quad (117)$$

These have very similar properties to the Lagrange brackets. They are canonical invariants and they are antisymmetric in the arguments. The *fundamental Poisson brackets* are defined the same way as the fundamental Lagrange brackets are:

$$[q_m,q_n] = 0, \qquad [p_m,p_n] = 0, \qquad [q_m,p_n] = -\,[p_n,q_m] = \delta_{mn}. \quad (118)$$

If we reverse the order of the matrix product on the left of (115), forming the product $\mathsf{M}\mathsf{S}\widetilde{\mathsf{M}}$, we find that the elements of the product matrix are the Poisson brackets of q and p in the (QP) system, arranged as in (115). If we then evaluate in the (qp) system, where the elements become the fundamental Poisson brackets, we get

$$\mathsf{M}\mathsf{S}\widetilde{\mathsf{M}} = \mathsf{S} \ ,$$

which we learned in Section 4 is equivalent to the symplectic condition. In other words, if we form a square $2N \times 2N$ matrix of the Poisson brackets,

$$\begin{pmatrix} [q_1,q_1] & [q_1,p_1] & [q_1,q_2] & \cdots & \cdots \\ [p_1,q_1] & [p_1,p_1] & [p_1,q_2] & \cdots & \cdots \\ [q_2,q_1] & [q_2,p_1] & [q_2,q_2] & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & [p_N,q_N] & [p_N,p_N] \end{pmatrix} = \mathsf{M}\mathsf{S}\widetilde{\mathsf{M}} = \mathsf{S}, \quad (119)$$

it is equal to the constant matrix S in any canonically related system of variables.

Having established the symplectic condition in Section 3 as a necessary and sufficient condition for a canonical transformation, we may consider that the

fundamental Lagrange brackets and Poisson brackets follow from it. Conversely, if we consider the fundamental brackets as given, they lead immediately to the symplectic condition.

Poisson brackets are quite useful in studying the dynamics of Hamiltonian systems. For example, Hamilton's equations of motion can be written in terms of Poisson brackets, and Poisson brackets are critical components of the Lie algebraic structure of Hamiltonian mechanics. Consequently, we list here some of their algebraic properties.
They obey a distributive law.

$$[(af + bg), h] = a[f, h] + b[g, h]. \tag{120}$$

They are antisymmetric in the arguments.

$$[f, g] = -[g, f]. \tag{121}$$

They obey the Jacobi identity.

$$[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0. \tag{122}$$

Finally, the Poisson brackets satisfy a product rule analogous to that for differentiation.

$$[f, gh] = [f, g]h + g[f, h]. \tag{123}$$

It is sometimes useful to express Poisson brackets in terms of the x's, the elements of the column vector $\mathbf{x}$.

$$[f, g] = \sum_i \sum_j \left(\frac{\partial f}{\partial x_i} \; (\mathbf{S})_{ij} \; \frac{\partial g}{\partial x_j} \right). \tag{124}$$

The fundamental Poisson brackets become

$$[x_i, x_j] = (\mathbf{S})_{ij}. \tag{125}$$

9 Lie Algebra, Lie Operators and Lie Transformations

In this section and the next we shall exploit the canonical invariance of Poisson brackets using Lie algebraic structures to devise new methods of finding the motions of Hamiltonian systems. The development in these sections follows the lines of parts of Reference [5] which will serve as a comprehensive reference and elaboration for the ideas presented here.

Most of us were introduced to algebra in high school as the manipulation of symbols (x and y) according to the rules for addition, subtraction, multiplication and division of real numbers. In modern mathematics, the term "algebra" has taken on a more general and more abstract meaning. Where the most elementary algebra applies to the set of real numbers, an algebra can refer

in general to any appropriate set of mathematical objects together with an appropriate set of operations. Thus we can talk about an algebra of matrices or an algebra of the rotations of a solid body for example.

The algebras in which we are interested are linear algebras which deal with generalized vectors in linear vector spaces.[1] An algebra A over a field of numbers F is defined as a linear vector space, with a multiplication law defined whereby any two vectors yield a product vector. The product, however it is defined, must satisfy three distributive conditions. Indicating multiplication by the symbol $\bullet$, to every ordered pair of elements x, y belonging to A, there must correspond a third unique element of A denoted by $x \bullet y$ and called the product of x and y.

$$(cx) \bullet y = x \bullet cy = c(x \bullet y) \tag{126}$$

$$(x + y) \bullet z = x \bullet z + y \bullet z \tag{127}$$

$$x \bullet (y + z) = x \bullet y + x \bullet z \tag{128}$$

for any x, y, z in A and c in F. Note that associativity, $(x \bullet y) \bullet z = x \bullet (y \bullet z)$, while allowed, is not required.

We may pause here and note that linear vector space of ordinary vectors with the scalar product defined as the multiplication law, $\mathbf{v} \bullet \mathbf{u} = \mathbf{v} \cdot \mathbf{u}$, does not qualify as an algebra, because the product is not a member of the set of vectors. However, if we replace the scalar product with the vector (cross) product as the multiplication law, $\mathbf{v} \bullet \mathbf{u} = \mathbf{v} \times \mathbf{u}$, we do get an algebra, and interestingly, the algebra we get is a Lie algebra!

A *Lie algebra* is an algebra, as defined above, in which the multiplicative rule (Lie product) satisfies, in addition to (126), (127) and (128), two further properties.

1. Antisymmetry

$$x \bullet y = -y \bullet x. \tag{129}$$

[1] In general, a non-empty set, the elements of which are called vectors, forms a *linear vector space* over a field F if addition and multiplication by elements of F of vectors and of vectors are defined and satisfy the following 10 postulates.

1. The sum $\mathbf{V}_1 + \mathbf{V}_2$ of any two vectors is a uniquely defined vector in the set.
2. Addition of vectors is commutative: $\mathbf{V}_1 + \mathbf{V}_2 = \mathbf{V}_2 + \mathbf{V}_1$.
3. Addition of vectors is associative: $(\mathbf{V}_1 + \mathbf{V}_2) + \mathbf{V}_3 = \mathbf{V}_1 + (\mathbf{V}_2 + \mathbf{V}_3)$.
4. There is a zero vector $\mathbf{0}$ such that $\mathbf{V} + \mathbf{0} = \mathbf{V}$ for all $\mathbf{V}$ in the set.
5. Every vector $\mathbf{V}$ has a unique negative $-\mathbf{V}$ such that $\mathbf{V} + (-\mathbf{V}) = \mathbf{0}$.
6. For every element c of F and any vector $\mathbf{V}$ in the set the product $c\mathbf{V}$ is a uniquely defined vector in the set.
7. $c(\mathbf{V}_1 + \mathbf{V}_2) = c\mathbf{V}_1 + c\mathbf{V}_2$.
8. $(c_1 + c_2)\mathbf{V} = c_1\mathbf{V} + c_2\mathbf{V}$.
9. $c_1(c_2\mathbf{V}) = (c_1 c_2)\mathbf{V}$.
10. $1\mathbf{V} = \mathbf{V}$.

In our case, F is the field of scalars.

2. The Jacobi condition

$$x \bullet (y \bullet z) + y \bullet (z \bullet x) + z \bullet (x \bullet y) = 0. \tag{130}$$

Any algebra that satisfies these five conditions is a Lie algebra. An example of a Lie algebra, as remarked above, is the set of all 3-vectors with the multiplication rule chosen to be the cross product. The antisymmetry in the exchange of operands is obvious, and the Jacobi condition can be shown to be satisfied by using the "BAC to CAB" expansion of the vector triple product.

The properties of Poisson brackets enumerated in the preceding section include both antisymmetry in the arguments and obedience to the Jacobi identity, facts which suggest that Poisson brackets might be used as the multiplication rule to define a Lie algebra. If that is to be done, the arguments of the Poisson brackets—functions $f, g \ldots$ of the canonical variables q, p, t—must form a linear vector space. Applying the criteria listed in the footnote on the preceding page, we find that these functions do indeed qualify as the vectors of a linear vector space. Thus we define the Lie product of any two such ordered functions f and g to be the Poisson bracket of those functions,

$$f \bullet g = [f, g], \tag{131}$$

and, according to (120) through (122), (129) and (130), the conditions for a Lie algebra are satisfied. The set of all functions of the variables q, p, t, together with (131) forms a Lie algebra which is called the *Poisson bracket Lie algebra of dynamical variables*.

It is also possible to define a *Lie algebra of symplectic matrices*. In order to do so, we recall that any real symplectic matrix sufficiently near the identity can be expressed *via* the exponential function in terms of a small real symmetric matrix $\mathbf{A}_s$. [See (87) in Section 4.]

$$\mathbf{M} = \exp(\mathbf{S}\mathbf{A}_s). \tag{132}$$

Conversely, every small real symmetric matrix $\mathbf{A}_s$ generates a real symplectic matrix *via* this equation, so at least for symplectic matrices near the identity, we have established a one-to-one correspondence between the symplectic matrix $\mathbf{M}$ and a particular symmetric matrix.

The set of matrices $(\mathbf{S}\mathbf{A}_s)$ fulfills all the requirements to constitute a linear vector field. Now if we define Lie multiplication of two matrices $\mathbf{A}$ and $\mathbf{B}$, in that order, to be their commutator,

$$\mathbf{A} \bullet \mathbf{B} \;=\; \mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}, \tag{133}$$

then the set of matrices $(\mathbf{S}\mathbf{A}_s)$ meets the conditions (129) and (130) and forms a Lie algebra.

For small $\mathbf{A}_{s1}$ and $\mathbf{A}_{s2}$, the product $\exp(\mathbf{S}\mathbf{A}_{s1})\exp(\mathbf{S}\mathbf{A}_{s2})$, yields another matrix, $\exp(\mathbf{S}\mathbf{A}_{s3})$, whose argument matrix $\mathbf{S}\mathbf{A}_{s3}$ belongs to the same Lie group. From that fact and the fact, already known to us, that symplectic matrices form

a group, we can conclude that these symplectic matrices near the identity form a Lie group, *i.e.,* a group that obeys a Lie algebra. It can be shown that this group may be extended to form a global group and, therefore, that symplectic matrices themselves form a Lie group.

Now we turn to the subject of Lie operators. As we have just seen, the use of Poisson brackets to define Lie multiplication is a cornerstone of the Lie algebra of dynamic variables. In order to exploit the Lie algebra, it is profitable to define an operator formalism to express Poisson brackets. Let $f(x,t)$ and $g(x,t)$ be functions of the canonical variables, $x_1, x_2 \cdots x_{2N}$ and t. The Lie operator associated with the function f will be denoted by enclosing f between two colons thus, $:f:$, and its action is defined in terms of Poisson brackets by the rule

$$:f:\, g = [f,g]. \tag{134}$$

Powers of the operator $:f:$ are defined as nested Poisson brackets,

$$:f:^2 g = [f,[f,g]], \quad :f:^3 g = [f,[f,[f,g]]], \ \ldots \tag{135}$$

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

$$:f:^0 g = g. \tag{136}$$

Because of the distributive properties of Poisson brackets, a Lie operator and its powers are linear operators, and a linear combination of two Lie operators is a Lie operator.

$$a:f: + b:g: \ = \ :(af + bg): \ , \tag{137}$$

the constants a and b being scalars. These properties qualify the set of Lie operators as a linear vector space.

The product rule for Poisson brackets (123), applies to a Lie operator.

$$:f:(gh) = (:f:g)h + g(:f:h). \tag{138}$$

Moreover a rule of the same structure applies to Poisson brackets, *i.e.,*

$$:f:[g,h] = [:f:g,h] + [g,:f:h]. \tag{139}$$

This rule follows from the Jacobi identity for Poisson brackets.

Building on the Lie operator and its properties we can now introduce *Lie transformations.* As in the case of the matrix exponential function, we refer to the exponential series, and define the Lie transformation associated with f operating on g as follows.

$$\exp\left(:f:\right) g = \sum_{n=0}^{\infty} \frac{1}{n!} :f:^n \ \ g. \tag{140}$$

The expanding this definition, we get

$$\exp\left(:f:\right) g = g + \ [f,g] \ + \ [f,[f,g]]/2! \ + \cdots \ . \tag{141}$$

In consequence of the product rule for Lie operators (138), the Lie transformation $\exp(:f:)$ has the property

$$\exp(:f:)(gh) = [\exp(:f:)g]\ [\exp(:f:)h]. \tag{142}$$

A Lie transformation acting on the product of two functions is exactly equivalent to the product of the results of the transformation acting on each function separately. The action of a Lie transformation on a function $g(x)$ is to perform a Lie transformation on its arguments.

$$\exp(:f:)\ g(x) = g(\exp(:f:)x). \tag{143}$$

And finally, an important property of Lie operators is the following.

$$\exp(:f:)\ [g,h] = [\exp(:f:)\ g, \exp(:f:)\ h]. \tag{144}$$

This property is useful in dealing with symplectic maps.

10 Symplectic Maps

A canonical transformation from one set of canonical variables $x_1, \cdots x_{2N}$ to a new set of canonical variables $\overline{x}_1, \cdots \overline{x}_{2N}$ is called a *mapping* or simply a map, and we shall denote it by $\mathcal{M}$. The expression,

$$\mathcal{M}\ :\quad x \to \overline{x}(x,t), \tag{145}$$

symbolizes what the mapping does. So a canonical transformation is a mapping, and the Jacobian matrix $\mathsf{M}(x,t)$ of the transformation is the Jacobian matrix of the map.

$$(\mathsf{M})_{ab}(x,t) = \frac{\partial \overline{x}_a}{\partial x_b}. \tag{146}$$

The map $\mathcal{M}$ is said to be symplectic—just as the transformation is—if its Jacobian matrix is a symplectic matrix.

$$\widetilde{\mathsf{M}}\mathsf{S}\mathsf{M} = \mathsf{S} \quad \text{or} \quad \mathsf{M}\mathsf{S}\widetilde{\mathsf{M}} = \mathsf{S}. \tag{147}$$

As we emphasized in Section 3, while M depends on x and t, the combinations $\widetilde{\mathsf{M}}\mathsf{S}\mathsf{M}$ and $\mathsf{M}\mathsf{S}\widetilde{\mathsf{M}}$ must be independent of both x and t—they must be constant. A symplectic map must have very special properties. As we found in Section 8, The fundamental Poisson brackets are invariants of the transformation and their invariance is a sufficient condition to establish that the transformation is canonical.

$$[\overline{x}_a, \overline{x}_b] = (\mathsf{M}\mathsf{S}\widetilde{\mathsf{M}})_{ab} = (\mathsf{S})_{ab} = [x_a, x_b]. \tag{148}$$

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

all dynamic variables. *Symplectic maps and canonical transformations are the same things.*

Now turning to the group properties of the map $\mathcal{M}$, we add the supposition that the map has a unique inverse $\mathcal{M}^{-1}$.

$$\mathcal{M}^{-1} \ : \overline{x} \to x. \tag{149}$$

That is, the inverse transformation sends us back exactly where we started. An increment in $\overline{x}$ is given by $d\overline{x} = \mathbf{M}\ dx$, and since the Jacobian matrix $\mathbf{M}$ is symplectic, and has an inverse $\mathbf{M}^{-1}$, we can write

$$dx = \mathbf{M}^{-1}\ d\overline{x}. \tag{150}$$

In view of (149) then, it is clear that the Jacobian matrix of $\mathcal{M}^{-1}$ is $\mathbf{M}^{-1}$. Moreover, since $\mathbf{M}$ is symplectic, we know that its inverse $\mathbf{M}^{-1}$ is also symplectic and that therefore $\mathcal{M}^{-1}$ is also a symplectic map.

Next if $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ are symplectic mappings that carry x to $\overline{x}$ and $\overline{x}$ to $\hat{x}$ respectively

$$\mathcal{M}^{(1)} \ : x \to \overline{x} \quad \text{and} \quad \mathcal{M}^{(2)} \ : \overline{x} \to \hat{x}, \tag{151}$$

then their product $\mathcal{M} = \mathcal{M}^{(1)}\mathcal{M}^{(2)}$ sends x to $\hat{x}$ and the Jacobian matrix of the composite mapping $\mathcal{M}$ is the product of the Jacobian matrices of the two mappings

$$\mathbf{M} = \mathbf{M}^{(1)}\mathbf{M}^{(2)}. \tag{152}$$

Furthermore since $\mathbf{M}$ is a symplectic matrix, the composite mapping $\mathcal{M}$ is also a symplectic map.

Finally, it is also obvious that the identity matrix and the corresponding identity mapping are symplectic. We conclude that the set of all symplectic maps is a group. This comes as no surprise, since we learned in Section 2 that canonical transformations form a group under composition, and there is a one-to-one correspondence between canonical transformations and maps.

At this point, we introduce the concept of *Hamiltonian flow.* By adding a time axis to (Gibbsian) phase space, thus creating a Euclidean space with $2N+1$ dimensions, we construct a *state space.* In this state space, the set of all trajectories $x_1(t), \cdots x_{2N}(t)$ is called a Hamiltonian flow.

If $H(x,t)$ is the Hamiltonian for some dynamical system, and $x^{(0)}$ denotes a set of initial conditions at some initial time $t^{(0)}$, and if $x^{(f)}$ denotes the coordinates at some later time $t^{(f)}$ of the trajectory with the initial conditions $x^{(0)}$, then the mapping $\mathcal{M}$ from $x^{(0)}$ to $x^{(f)}$ obtained by following the Hamiltonian flow specified by $H(x,t)$,

$$\mathcal{M} \ : x^{(0)} \to x^{(f)}, \tag{153}$$

is symplectic. This finding has the same content as the conclusion reached in Section 3 that the motions of Hamiltonian systems are themselves canonical transformations.

Returning to the Lie transformation, we shall show the important fact that *Lie transformations generate symplectic maps and any symplectic map can be*

generated by a Lie transformation. Consider the Lie transformation $\exp(:f:)$, where $f(x,t)$ can be any function, and define a map $\mathcal{M}$ as an *operator,*

$$\mathcal{M} = \exp(:f:). \tag{154}$$

That is, $\mathcal{M}$ operates on x to produces new variables $\overline{x}(x,t)$ by the rule,

$$\overline{x}_a(x,t) = \mathcal{M}\ x = \exp(:f:)\ x_a, \qquad a = 1, 2, \ldots 2n. \tag{155}$$

The Poisson brackets of various $\overline{x}$'s with each other are

$$\begin{aligned}[\overline{x}_a, \overline{x}_b] &= [\exp(:f:)\ x_a, \exp(:f:)\ x_b] \\ &= \exp(:f:)\ [x_a, x_b] \\ &= \exp(:f:)\ (\mathbf{S})_{ab} = (\mathbf{S})_{ab}.\end{aligned} \tag{156}$$

The fundamental Poisson brackets are preserved, so $\mathcal{M}$ is a symplectic map: *every Lie transformation generates a symplectic map.*

An important and useful feature of the symplectic map generated by the Lie transformation $\exp(:f:)$ is that the argument f is an invariant function of the map.

$$f(\overline{x}, t) = f(x, t). \tag{157}$$

This follows from the fact that the action of a Lie transformation on a function of the dynamic variables is to perform a Lie transformation on its arguments, so that

$$\exp(:f:)\ f(x,t) = f(\exp(:f:)\ x) = f(\overline{x}, t). \tag{158}$$

Also

$$\exp(:f:)\ f(x,t) = f + [f,f] + \cdots = f(x,t), \tag{159}$$

and (157) is proven.

Applying the symplectic map $\exp(-:f:)$ to both sides of (155) above,

$$\exp(-:f:)\ \overline{x}_a(x,t) = \exp(-:f:)\exp(:f:)\ x_a\ , \tag{160}$$

and using the properties of Lie operators and the definition of the Lie transformation, we can show that the product $\exp(-:f:)\exp(:f:)$ results in a net identity operation, just as $e^{-x}e^{x} = 1$. Consequently, the inverse mapping is

$$x_a = \exp(-:f:)\ \overline{x}_a, \tag{161}$$

or in other words,

$$\mathcal{M}^{-1} = \exp(-:f:), \tag{162}$$

and as we already know, its Jacobian matrix is $\mathbf{M}^{-1}$, which is a symplectic matrix.

The next question that arises is this: can any given symplectic map $\mathcal{M}$ whatever be written in exponential form? The answer lies in the Factorization theorem. The proof of the theorem involves the concepts we have explored, but

the mathematics involved is beyond the scope of this monograph, so we just state the theorem.

Factorization theorem— *Let $\mathcal{M}$ be an analytic symplectic mapping which maps the origin into itself. That is, the relationship between the $\overline{x}$'s and the x's is assumed to be expressible in a power series of the form*

$$\overline{x}_a = \sum_b L_{ab}\, x_b \;+\; \text{higher order terms in } x. \tag{163}$$

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

$$\mathcal{M} = \exp\left(:f_2^c:\right) \exp\left(:f_2^a:\right) \exp\left(:f_3:\right) \exp\left(:f_4:\right) \cdots, \tag{164}$$

where the functions f_m are homogeneous polynomials of degree m in the variables $x_1 \cdots x_{2n}$. In particular, the polynomials f_2^c and f_2^a are quadratic polynomials of the form,

$$f_2^c = -\frac{1}{2} \sum_i \sum_j A_{ij}^c\, x_i\, x_j, \tag{165}$$

$$f_2^a = -\frac{1}{2} \sum_i \sum_j A_{ij}^a, x_i\, x_j, \tag{166}$$

where A_{ij}^c and A_{ij}^a are the elements of real symmetric matrices which commute and anticommute with S respectively.

$$\mathsf{A}^c\mathsf{S} - \mathsf{S}\mathsf{A}^c = \mathbf{0}, \qquad\qquad \mathsf{A}^a\mathsf{S} + \mathsf{S}\mathsf{A}^a = \mathbf{0}. \tag{167}$$

We had seen that all Lie transformations generate symplectic maps, and the Factorization theorem tells us that all symplectic maps can be expressed in terms of Lie transformations. This important theorem opens a new door to finding solutions to Hamilton's equations of motion, because, as we have learned, the solutions of Hamilton's equations are symplectic maps.

There are two particular features of the factored product representation (164) that merit emphasis. First, if the product (164) is truncated at any point, the result of the truncation is still a symplectic map, because each term in the product is a symplectic map. Furthermore, if we drop all of the factors beyond $\exp\left(:f_m:\right)$ for some particular m, then the power series expansion (163) for the truncated map agrees with the exact map through terms of degree $(m-1)$. The truncated map provides a symplectic approximation to the exact map.

Second, if the product (164) is factored into two partial products, the first involving only the quadratic polynomials and the second involving cubic and higher degree polynomials,

$$\mathcal{M} = \mathcal{M}_\ell\; \mathcal{M}_{n\ell}, \tag{168}$$

where

$$\mathcal{M}_\ell = \exp\left(:f_2^c:\right) \exp\left(:f_2^a:\right) \quad \text{and} \quad \mathcal{M}_{n\ell} = \exp\left(:f_3:\right) \exp\left(:f_4:\right) \cdots, \tag{169}$$

it can be shown that retaining only $\mathcal{M}_\ell$ leads in beam dynamics to the linear approximation for $\mathcal{M}$. The remaining factors $\mathcal{M}_{n\ell}$ then represent the non-linear effects in the beam dynamical system, such as chromaticity and non-linear guide fields.

The Factorization theorem tells us that the map we are seeking can be written as a product of Lie transformations, but it does not tell us explicitly how to find the transformations. Finding the Lie transformations for particular Hamiltonians is the subject of a whole field of study. For examples of this work the reader is directed to Reference [5].

The General Map for a "Time-independent" Hamiltonian

In the frequently encountered case in beam dynamics that the Hamiltonian does not depend explicitly on the independent variable, whether it be time or the longitudinal coordinate s, we can write a general formula for the map $\mathcal{M}$ as a single Lie transformation formed in terms of the Hamiltonian. Taking t as the independent variable, the map for this Hamiltonian flow carries the system from its state at t_0 to that at t_f.

$$\mathcal{M} : x(t_0) \rightarrow x(t_f), \tag{170}$$

where, of course, x symbolizes the canonical variables $x_1 \cdots x_{2N}$, and the operator $\mathcal{M}$ is

$$\mathcal{M} = \exp\{-(t_f - t_0)\ :H:\}. \tag{171}$$

For proof we first expand the transformation.

$$\overline{x}(t_f) = \sum_{n=0}^{\infty} \frac{1}{n!}(t_f - t_0)^n\ (: - H:)^n\ x. \tag{172}$$

Then we compare this expression with the Taylor expansion of $x(t_f)$.

$$x(t_f) = x(t_0) + \sum_{n=1}^{\infty} \frac{1}{n!}(t_f - t_0)^n \left(\frac{d^n x}{dt^n}\right)_{t_0}. \tag{173}$$

If the Hamiltonian does not depend on the time,

$$\frac{dx}{dt} = [x, H] = [-H, x] = : - H:\ x,$$

$$\frac{d^2x}{dt^2} = [-H, \dot{x}] = [-H, [-H, x]] + \frac{\partial}{\partial t}[-H, x] = (: - H:)^2\ x,$$

$$\frac{d^n x}{dt^n} = (: - H:)^n\ x, \tag{174}$$

and (172) is proven.

Example 6—To follow the process of generating symplectic maps using Lie transformations, let us apply the method to the case of particle motion through

an ideal quadrupole magnet. As our starting point we choose the second-order Taylor expansion of the Hamiltonian (58) in Section 3:

$$K_{2nd} = -p_0 + \frac{P_x^2}{2p_0} + \frac{P_y^2}{2p_0} + \frac{m^2\epsilon^2}{2p_0^3} - \frac{H_0 hx\epsilon}{p_0 c^2} + eS(x,y,s),$$

where s is the independent variable, the canonical variables are x, P_x, y, P_y, τ and $-\epsilon$, and $S(x,y,s)$ is the longitudinal component of the vector potential, the other two components being zero. For a quadrupole magnet, $S = G(y^2 - x^2)/2$, where G is the magnetic field gradient $(\partial B_y/\partial x)$. For simplicity we consider a particle at the design energy, so $\epsilon = 0$, and we shall ignore longitudinal (τ) motion. We are left with the Hamiltonian,

$$H = -p_0 + \frac{P_x^2}{2p_0} + \frac{P_y^2}{2p_0} + \frac{eG}{2}(y^2 - x^2). \tag{175}$$

Since this Hamiltonian contains no linear terms in the canonical coordinates, the map for the Hamiltonian flow maps the origin onto itself.

The Hamiltonian does not depend on the independent variable s, so (172) applies and the map is given by the the formula,

$$\mathcal{M} = \exp\{-(s_f - s_0)\!:\!H\!:\} = \exp\{-L\!:\!H\!:\}, \tag{176}$$

where L is the length of the quadrupole, and the transformations are

$$x_f = (\mathcal{M}x)_i, \quad P_{xf} = (\mathcal{M}P_x)_i, \quad y_f = (\mathcal{M}y)_i, \quad P_{yf} = (\mathcal{M}P_y)_i.$$

where the subscripts i and f label the initial and final values respectively. Let us expand the first of these.

$$x_f = \left(\sum_{n=0}^{\infty} \frac{1}{n!} \!:\!(-LH)\!:^n x \right)_i . \tag{177}$$

For organizing the evaluation of this series, it is helpful to define a set of auxiliary quantities:

$$\alpha_m = \!:\! - LH\!:^{m-1} x, \quad m = 1, 2, \ldots \tag{178}$$

In terms of these quantities, the series for x may be written

$$x_f = \left(\sum_{m=1}^{\infty} \frac{\alpha_m}{(m-1)!} \right)_i , \tag{179}$$

and a systematic calculation can be carried out, evaluating first α_1, then α_2 and so on in terms of the recurrence relation,

$$\alpha_{m+1} = \!:\! - LH\!:\alpha_m. \tag{180}$$

We begin the evaluation with α_1 and continue with the recurrence relation.

$$\begin{aligned}
\alpha_1 &= :-LH:^0\, x = x \\
\alpha_2 &= :-LH:\alpha_1 = LP_x/p_0 \\
\alpha_3 &= :-LH:\alpha_2 = L^2 eGx/p_0 \\
\alpha_4 &= :-LH:\alpha_3 = L^3 eGP_x/p_0^2 \\
&\quad etc.
\end{aligned}$$

The resulting series is

$$x_f = x_i + \frac{LP_{xi}}{p_0} + \frac{L^2 eG\, x_i}{2\, p_0} + \frac{L^3 eGP_{xi}}{6\, {p_0}^2} + \frac{L^4 e^2 G^2\, x_i}{24\, {p_0}^2} + \frac{L^5 e^2 G^2 P_{xi}}{120\, {p_0}^3} + \frac{L^6 e^3 G^3\, x_i}{720\, {p_0}^3} + \cdots \qquad (181)$$

We can then identify the resulting terms that are even in L with the first terms of the series for the hyperbolic cosine and those that are odd in L with the leading terms of the series for the hyperbolic sine as they appear in the following expression,

$$x_f = x_i\ \cosh\left(\sqrt{\frac{eG}{p_o}}L\right) + \frac{P_{xi}}{p_0}\sqrt{\frac{p_o}{eG}}\ \sinh\left(\sqrt{\frac{eG}{p_o}}L\right), \qquad (182)$$

which of course is the x-part of the correct map for a quadrupole magnet in the case that the quantity eG is positive.

We note that if the sign of eG were reversed ($eG \to -eG$), the new dependence on y and P_y would be the same as the previous dependence on x and P_x. In other words, the roles of x and y would be reversed. Therefore the y-part of the quadrupole map must be the same series as (181) but with the sign of eG reversed. That series corresponds to the expression,

$$y_f = y_i\ \cos\left(\sqrt{\frac{eG}{p_o}}L\right) + \frac{P_{yi}}{p_0}\sqrt{\frac{p_o}{eG}}\ \sin\left(\sqrt{\frac{eG}{p_o}}L\right). \qquad (183)$$

We can obtain the series for P_x by evaluating the α's and summing the series (179) as we did above. Then we can use the role-reversal argument to get P_y, and the quadrupole map is complete.

In fact the formula (172) for the general map for a time-independent Hamiltonian is even more useful than it might at first appear, because *a system that has a time-dependent Hamiltonian with N degrees of freedom can be described by a different Hamiltonian with $(N+1)$ degrees of freedom, which is time-independent* and to which, therefore, the formula applies. This maneuver is accomplished as follows.

Suppose we have system with N degrees of freedom described by the time-dependent Hamiltonian $H(q,p,t)$ where as usual q and p symbolize the whole sets of canonical coordinates and momenta. We define a new Hamiltonian,

$$H_+ = H(q,p,q_{N+1}) + p_{N+1}. \tag{184}$$

We have extended the phase space to include an additional coordinate q_{N+1} and its canonically conjugate momentum p_{N+1}, modified the Hamiltonian by the addition of the term p_{N+1} and substituted the new coordinate q_{N+1} for time. The equation of motion for the new coordinate is

$$\dot{q}_{N+1} = \frac{\partial H_+}{\partial p_{N+1}} = 1, \tag{185}$$

which, when integrated gives

$$q_{N+1}(t) = t. \tag{186}$$

As we shall see, we need not concern ourselves with the equation of motion for p_{N+1}. The remaining equations of motion—those for the original N degrees of freedom—are exactly as they were before, except that q_{N+1} appears in place of t. Clearly since $q_{N+1}(t)$ duplicates the flow of time, the solutions to these equations behave exactly like those of Hamilton's equations of motion for $H(q,p,t)$. The equation of motion for p_{N+1} tells us nothing further about the system we are interested in, and we may therefore ignore it. Symplecticity applies to the whole system of $(N+1)$ degrees of freedom, but it also applies to the original system governed by H.

Since H_+ is explicitly time-independent, the map for the system can be generated by the Lie transformation,

$$\mathcal{M} = \exp\{-(t_f - t_0)\ :H_+:\}, \tag{187}$$

in which the complete set of canonical coordinates for the system include q_{N+1} and p_{N+1}.

Example 7—To see how this technique works in a simple case, consider a particle of unit mass under the influence of a force that increases linearly with time; a one-dimensional Hamiltonian for such a system is

$$H = p^2/2 - qt, \tag{188}$$

where q is the coordinate of the particle. Hamilton's equations of motion yield the solutions,

$$q(t) = q_0 + p_0 t + t^3/6, \qquad p(t) = p_0 + t^2/2. \tag{189}$$

The extended Hamiltonian for the system is the following.

$$H_+ = p_1^2/2 + q_1 q_2 + p_2. \tag{190}$$

The Lie transformations are easily calculated finite sums:

$$q1(t) = q_1(0) + p_1(0)t + t^3/6, \qquad p_1(t) = p_1(0) + t^2/2, \tag{191}$$

in agreement with the foregoing solution. Of course, $q_2(t) = t$, and a map is generated for p_2, but it is of no interest to us.

The Lie-transformation method is not difficult to carry out on a simple problem like that of the quadrupole, especially with the help of a computer algebra program like *Maple* or *Mathematica*, and the computation is straightforward. However its real power lies beyond these simple problems. The Hamiltonian equations of motion for the quadrupole are linear and can be easily solved. But what of beam-transport elements with higher-order fields for which the equations of motion are non-linear and become rapidly less tractable as the non-linearity becomes stronger? For these elements, the Lie-transformation method retains the same systematic, formulaic procedure, even if the mathematical manipulations themselves become more complicated.

A great advantage of the Lie transformation comes into play when the motion can be determined only to an approximation. In that case, if the Lie transformation can be factored into a product of Lie transformations, as in the Factorization theorem, then the product can be truncated and the truncated product generates a symplectic approximation to the true map.

11 Symplectic Integration

In predicting or analyzing the behavior of a beam in an accelerator or storage ring—and especially in a costly machine yet to be built—we are often forced to use numerical integration to trace particles through complex electromagnetic fields that do not admit of analytic solution. Since we frequently have to rely on the long-term results of these integrations, the build-up of errors in the numerical integration steps becomes very important. It is not possible to eliminate all the errors, of course, because each step is intrinsically an approximation; however it is possible to eliminate one class of errors: that due to asymplecticity. Generally asymplectic or non-Hamiltonian integration steps display long-term damping or antidamping with concomitant concentration or dilution of particle-point density in phase space. We can eliminate those effects by making each step symplectic.

One method of creating symplectic steps is a "building-block" approach. to illustrate the method, consider the an idealized pendulum governed by the Hamiltonian,

$$H(x, p) = p^2/2 + \cos x. \tag{192}$$

(The mass of the bob is unity and the length of the pendulum is $1/g$ so that $mlg = 1$.) The system has one degree of freedom with angle coordinate x and canonical momentum p. The equations of motion are

$$\dot{x} = p, \qquad \dot{p} = \sin x. \tag{193}$$

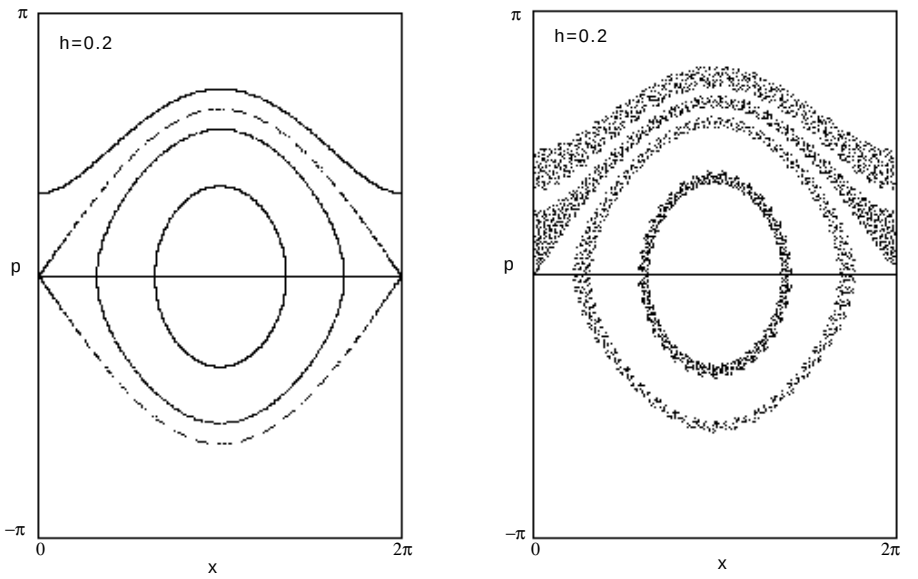

Figure 1: Phase-space trajectories for the pendulum generated by the building-block integrator (left) and a second-order Runge-Kutta integrator (right) of comparable computational cost. The time-step h and the initial conditions are the same for the two plots.

To construct an integration step, consider a short interval of time of duration h. The general approach is to concentrate the effect of the force ($\sin x$) in an infinitesimal instant at the center of the interval, and let the pendulum rotate freely during the first and last halves of the interval. We cut the interval h into two parts—building blocks—of duration $h/2$ each, during which we ignore the action of gravity so that the effective Hamiltonian is just $H_{\text{eff}} = p^2/2$; we call these blocks "drifts." Between these two parts the force of gravity is considered to change the momentum impulsively (instantaneously) by an amount that approximates the change that would take place in the entire interval h in the exact solution. This block is called an "impulse."

$$\Delta p = \int_0^h \sin x(t) dt \approx h \sin(x_0 + h p_0/2).$$

This is the final building block. Labeling the initial values of the coordinates x_0 and p_0, the transformations are these:

$$t \to t + h/2: \quad x_1 = x_0 + h p_0/2, \quad p_1 = p_0, \tag{194}$$

$$\text{Impulse}: \quad x_2 = x_1, \quad p_2 = p_1 + h \sin x_1, \tag{195}$$

$$t + h/2 \to t + h: \quad x_3 = x_2 + h p_2/2, \quad p_3 = p_2, \tag{196}$$

Since each of these is manifestly symplectic, the transformation through the whole interval is symplectic, and we may regard the composite transformation from (x_0, p_0) to (x_3, p_3) as a symplectic integration step for the system.

We can visualize the advantage of a symplectic integrator over an asymplectic integrator by comparing the phase-space trajectories generated by the two integrators for the same system with the same initial conditions. Figure 1 shows the comparison. On the left are trajectories for the pendulum calculated by the building-block integrator, and on the right those calculated by a second-order Runge-Kutta integrator of comparable computational cost to that of the building-block integrator. The time-step for the two examples is the same, $h = 0.2$; the period for small-amplitude motion is 2π; so the time-step is about one thirtieth of the small-amplitude period. The integrations extend over 1000 steps.

While the symplectic integrator gives a faithful representation of the pendulum's motion, the Runge-Kutta integrator shows anti-damping behavior, which manifests itself in an outward spiraling of the phase trajectories, creating bands instead of lines in the plot. This behavior is non-Hamiltonian; it clearly implies that Liouville's Theorem is not obeyed by the solutions generated by the Runge-Kutta integrator. A numerical-integration step is a map, so the phase-space plots of Figure 1 depict iterations of the map. If the iterations lead to a closed contour in phase space, the interior area of the contour is preserved since the integrator is a symplectic map.

In the case of the pendulum we are mapping a non-linear Hamiltonian system. In the left-hand plot of Figure 1, some typical features of such systems exhibit themselves: a region of stable motion—bounded in phase-space—exists, centered on the point $(\pi, 0)$; The region of bounded motion is delimited by a separatrix, which appears dashed in the plot; and regions of unbounded motion, corresponding to monotonic increase or decrease of x lie beyond the separatrix. The phase space is divided into regions of different behavior. Another feature that is typical of non-linear Hamiltonian systems is chaotic motion, a feature that is not visibly present in the plot. Chaotic motion can be made to appear in the motions generated by the building-block integrator simply by increasing the time step h, and thus increasing the effects of non-linearity on the map.

Figure 2 shows the result of increasing h by a factor of five, from 0.2, shown on the left, to 1.0, shown on the right. The stable trajectories are distorted somewhat, and the separatrix, if it exists at all, is contracted. Chaotic motion arises at larger amplitudes. The $h = 1.0$ map still produces stable, undamped motion at small amplitudes, which is not the case with the asymplectic Runge-Kutta step even at small amplitudes.

The building-block approach to constructing symplectic integrators may be used to construct more accurate integration steps by using more building blocks of the same sort, however care must be exercised in constructing them to assure that they are more accurate. Such integrators are guaranteed to be symplectic, but they are not guaranteed to be accurate.

There are other methods of constructing symplectic integrators. One method, which we shall call Ruth's method[7], that approaches the problem with a strat-

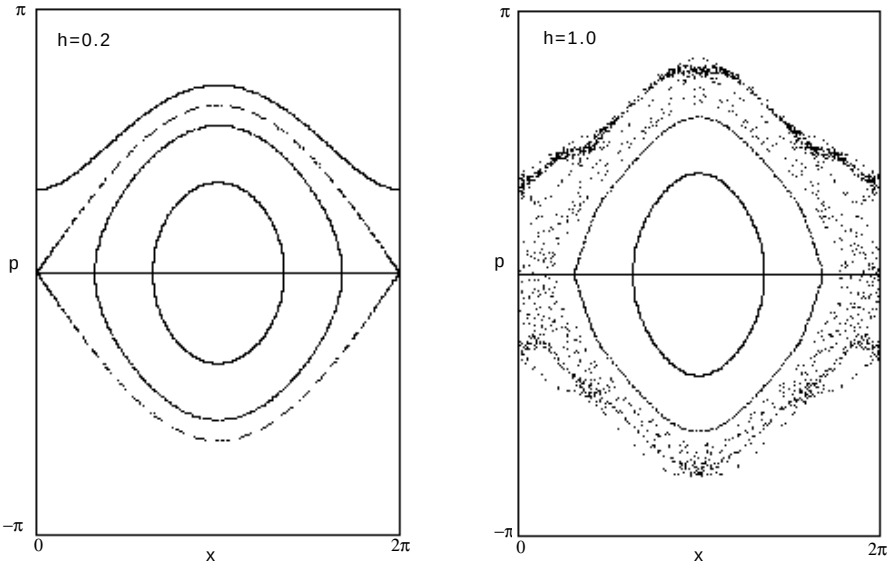

Figure 2: The effect of increasing the time step h is to distort the stable trajectories and to introduce chaotic motion in the vicinity of the separatrix and beyond. The plot on the left is the same as that of Figure 1. The plot on the right shows the phase-space trajectories generated by the building-block integrator for the same initial conditions but with a time step of 1.0.

egy that is a little reminiscent of the Hamilton-Jacobi method of solving Hamilton's equations. The strategy is to seek a solution to Hamilton's equations by finding a canonical transformation that transforms the Hamiltonian to an appropriate form. In the Hamilton-Jacobi method one tries to transform the Hamiltonian to zero. In Ruth's method the goal is to transform the Hamiltonian *almost* to zero.

Numerical integration steps are approximations that take advantage of the smallness of the step in the independent variable—t in the case of the pendulum. In Ruth's method we consider the transformed Hamiltonian as a power series in the time-step h and measure the merit of the transformation in terms of the lowest power of h that appears in the transformed Hamiltonian. In fact, the procedure is to make several successive transformations, each one of which eliminates the lowest remaining power of h in the resulting Hamiltonian. The result is that for a given step h, the Hamiltonian gets nearer zero in each transformation.

The rationale for this scheme, as it is in the Hamilton-Jacobi scheme, is the following. As we have already learned, the solutions of Hamilton's equations are themselves canonical transformations. Thus if the Hamiltonian is $H(x, p, t)$ and the solutions are $x(x_0, p_0, t)$ and $p(x_0, p_0, t)$, x_0 and p_0 being the initial conditions, we may think of the solutions as transformations carrying the system from the initial conditions to the state at time t. Conversely the inverse transformations carry the system from the state at t to the initial conditions, which are constants.

The inverse transformations transform the Hamiltonian H, according to (51), to a transformed Hamiltonian $K(x_0, p_0, t)$ in which the transformed canonical coordinates and momenta are constants. Then since $\dot{x}_0 = 0$ and $\dot{p}_0 = 0$, K must be independent of x_0 and p_0 and may be chosen to be zero. In summary the inverse of the solution transforms the Hamiltonian to zero.[3]

Now starting again with the Hamiltonian $H(x, p, t)$, suppose we have found not the solution but a canonical transformation that transforms the Hamiltonian to the form,

$$K = \sum_{n=k}^{\infty} a_n(X, P)t^n = a_k(X, P)t^k + a_{k+1}(X, P)t^{k+1} \cdots . \tag{197}$$

Hamilton's equations of motion for this system are

$$\dot{X} = \sum_{n=k}^{\infty} \left(\frac{\partial a_n}{\partial P} \right) t^n, \qquad \dot{P} = -\sum_{n=k}^{\infty} \left(\frac{\partial a_n}{\partial X} \right) t^n. \tag{198}$$

Integrating with respect to time we get the solutions,

$$X = X_0 + \mathcal{O}(t^{k+1}), \qquad P = P_0 - \mathcal{O}(t^{k+1}). \tag{199}$$

[3]In fact the Hamiltonian may be any function of t, but such a function has no effect on the equations of motion, and may be discarded.

Therefore the original canonical transformation differs from the exact solution only in the $(k+1)$th order and higher in the time.

In the case of the Ruth integrator, the successive transformations are designed to eliminate successively higher powers of t in the transformed Hamiltonian, and therefore to give progressively higher accuracy. And at every step the result is a symplectic transformation. The composition of the transformations may be thought of as carrying the system from the state at some time $t+h$ backward in time to the state t. To create the desired approximation to the solution itself, we use the *inverse* of this transformation as the integration step.

To find the appropriate canonical transformations to transform the Hamiltonian to the desired form, the generating-function method is used. For an explanation of the method the reader is referred to Reference [7]. To complete our discussion we give the second order (two transformations) integration step for a Hamiltonian of the form

$$H(x,p) = \frac{p^2}{2} + V(x,t).$$

Designating the time step h, the initial conditions of the motion (x_0, p_0, t_0) and the state of motion at the end of the step (x, p, t), the transformations are

$$\begin{array}{ll} p_1 = p_0, & x_1 = x_0 + hp_1/2, \\ p = p_1 + hf(x_1, t_0 + h/2), & x = x_1 + hp/2, \end{array} \tag{200}$$

where $f(x,t) = -\partial V/\partial x$. Interestingly this result coincides exactly for the pendulum (192) with that of the building-block method described earlier in this section.

12 Appendix: The Hamiltonian Formulation For Beam Dynamics

In this appendix we derive the Hamiltonian for non-interacting beam particles moving in the given electromagnetic fields of the accelerator. This Hamiltonian involves only the three degrees of freedom possessed by the charged particle.

The *coordinate system* we use is a curvilinear system based on the motion of an ideal particle called the "reference particle" which follows what we shall call the "reference trajectory" through the system. The reference trajectory is a curve in space consisting in general of a succession of straight segments and curved segments, and it may be an open curve as it is in a beam transport line or linear accelerator, or a closed curve as it is in a storage ring or synchrotron. Examples are shown in Figure 3. Letting the variable s measure the distance along the reference trajectory from some chosen origin, the reference trajectory is traced out by the vector $\mathbf{r}_0(s)$ whose own origin is an arbitrary point in the laboratory system. To every value of s there corresponds a $\mathbf{r}_0$.

We denote the curvature of the reference trajectory by $h(s)$; its magnitude is the reciprocal of the local bending radius of the reference trajectory. We make

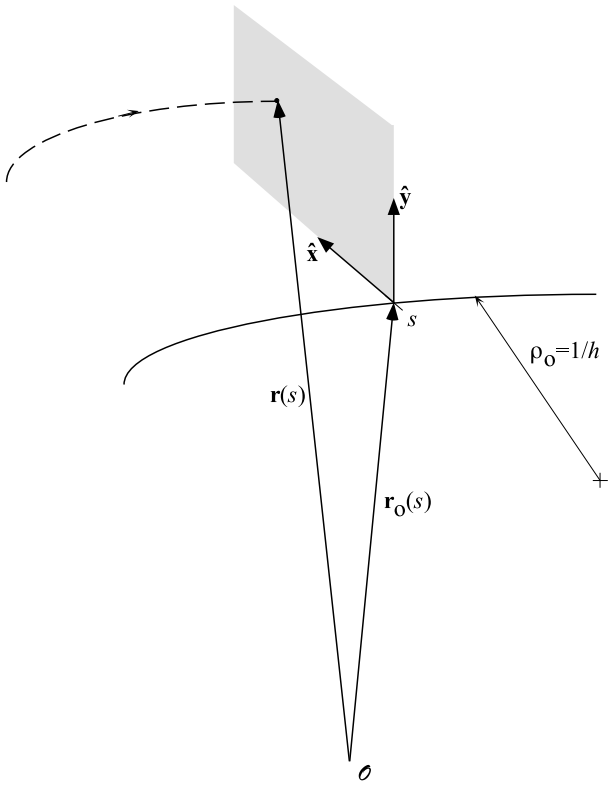

Figure 3: The curvilinear coordinate system. The origin of the **r**-coordinates is an arbitrary point $\mathcal{O}$ in the laboratory system. The origin of the curvilinear x-y-s system is chosen for convenience and does not coincide with that of the **r**-system.

the important restriction that the local reference trajectory must lie in a fixed plane—locally, the reference trajectory undergoes no torsion. This constraint is not unduly restrictive in practise, because the coordinate system can be rotated about the reference trajectory at any point, defining a new fixed plane for the next local segment of the reference trajectory.

We may think of the motion of the reference particle as defining the reference trajectory, its path being specified by the vector function, $\mathbf{r}_0(s)$, and its schedule along the path as a scalar function $t_0(s)$. This characterization implies s to be the independent variable. On the other hand, beginning on a more traditional note and acknowledging the special status of time as independent variable in classical mechanics, we could specify instead $\mathbf{r}_0(t)$, the time-dependent trajectory of the reference particle. Having done so, we could extract $s(t)$ for the reference particle and invert it to find $t_0(s)$ and from that $\mathbf{r}_0(s) = \mathbf{r}_0(t_0(s))$. Note that we can be sure that the inversion of $s(t)$ will lead to a single-valued $t_0(s)$, because $s(t)$ is always a smooth, monotonic function of t in particle beams.

The motion of a general particle, given by the position vector $\mathbf{r}(s)$, is de-

scribed relative to the reference trajectory. This is done by establishing a local coordinate system whose longitudinal axis is formed by the reference trajectory itself. The variable s is the longitudinal coordinate; it increases in the direction of the motion of the reference particle. The value of the longitudinal coordinate of a general particle is the perpendicular projection on the s-axis of the particle's position. The other two coordinates are perpendicular to the s-axis and to each other, forming a right-handed orthogonal set. They are called x and y with corresponding unit vectors, $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. The unit vectors form a right-handed system. See Figure 2.

$$\begin{aligned} \hat{\mathbf{x}} &= \hat{\mathbf{y}} \times \hat{\mathbf{s}} \\ \hat{\mathbf{y}} &= \hat{\mathbf{s}} \times \hat{\mathbf{x}} \\ \hat{\mathbf{s}} &= \hat{\mathbf{x}} \times \hat{\mathbf{y}} \ , \end{aligned} \tag{201}$$

where $\hat{\mathbf{s}} = d\mathbf{r}_0/ds$.

Wherever the reference trajectory is curved, the plane of bending is well defined. The coordinate x always lies in the bending plane where a bending plane exists. In other portions of the system, where the reference trajectory is straight, the x-direction may be chosen for convenience. Once x-direction is established, however that is done, the y-direction is uniquely defined by (201). The transverse (xy) coordinate system has its origin at the point where a plane perpendicular to the reference trajectory and containing the particle is pierced by the reference trajectory. It moves along with the particle. See Figure 2.

In terms of these coordinates the position vector of a general particle is given by

$$\mathbf{r} = \mathbf{r}_0(s) + \hat{\mathbf{x}}(s)x(s) + \hat{\mathbf{y}}y(s) \ , \tag{202}$$

where $\mathbf{r}_0(s), \hat{\mathbf{x}}(s), \hat{\mathbf{y}}$ are given *a priori* and x, y are to be determined by solving the equations of motion. The derivatives of the unit vectors are

$$\hat{\mathbf{x}}' = \hat{\mathbf{s}}h \tag{203}$$

$$\hat{\mathbf{y}}' = 0 \tag{204}$$

$$\hat{\mathbf{s}}' = -\hat{\mathbf{x}}h \ , \tag{205}$$

where primes denote total differentiation with respect to s. The magnitudes of the tangent vectors in this coordinate system are

$$\left|\frac{\partial \mathbf{r}}{\partial x}\right| = 1 \ , \qquad \left|\frac{\partial \mathbf{r}}{\partial y}\right| = 1 \ , \qquad \left|\frac{\partial \mathbf{r}}{\partial s}\right| = 1 + hx \ , \tag{206}$$

so the vector differential operations—gradient, divergence, curl and Laplacian—are the following.

$$\text{grad } \psi = \hat{\mathbf{x}}\frac{\partial \psi}{\partial x} + \hat{\mathbf{y}}\frac{\partial \psi}{\partial y} + \frac{\hat{\mathbf{s}}}{1+hx}\frac{\partial \psi}{\partial s} \ , \tag{207}$$

$$\text{div } \mathbf{A} = \frac{1}{1+hx}\frac{\partial}{\partial x}[(1+hx)A_x] + \frac{\partial A_y}{\partial y} + \frac{1}{1+hx}\frac{\partial A_s}{\partial s} \ , \tag{208}$$

$$\text{curl } \mathbf{A} =$$
$$\hat{\mathbf{x}}\left\{\frac{\partial A_s}{\partial y}-\frac{1}{1+hx}\frac{\partial A_y}{\partial s}\right\}+\frac{\hat{\mathbf{y}}}{1+hx}\left\{\frac{\partial A_x}{\partial s}-\frac{\partial}{\partial x}[(1+hx)A_s]\right\}+\hat{\mathbf{s}}\left\{\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right\}\ , \tag{209}$$

$$\text{div (grad } \psi) =$$
$$\frac{1}{1+hx}\left\{\frac{\partial}{\partial x}\left[(1+hx)\frac{\partial\psi}{\partial x}\right]+\frac{\partial}{\partial y}\left[(1+hx)\frac{\partial\psi}{\partial y}\right]+\frac{\partial}{\partial s}\left[\frac{1}{1+hx}\frac{\partial\psi}{\partial s}\right]\right\}\ , \tag{210}$$

where ψ and $\mathbf{A}$ are a scalar function and a vector function respectively.

In classical mechanics, whether in the Newtonian formulation or the Hamiltonian formulation, time occupies a special role as the inherently independent variable, with the consequence that the description of the motion of a general particle is ordinarily expressed by the three coordinate functions $x(t)$, $y(t)$, and $s(t)$. The corresponding velocities are $\dot{x}(t)$, $\dot{y}(t)$, and $\dot{s}(t)$. (*Overhead dots denote total differentiation with respect to time.*) In the case of a particle in a beam, the longitudinal velocity $\dot{s}$ never drops to zero or reverses. In other words, the variable $s(t)$ increases monotonically with time, $\dot{s} > 0$, and consequently we may invert $s(t)$ to obtain a unique $t(s)$ which will itself be monotonic in s. This fact opens the door to treating s as the independent variable in beam dynamics and considering t as a coordinate depending on s. This procedure is convenient in beam dynamics, so it customary to describe the motion of a beam particle in terms of the three functions $x(s)$, $y(s)$ and $t(s)$. The corresponding "velocities" are $x'(s)$, $y'(s)$ and $t'(s)$.

To derive the Hamiltonian formulation for particle beam dynamics, we begin with the relativistic Lagrangian for a charged particle moving in an electromagnetic field. The field is prescribed by its vector and scalar potentials Φ and $\mathbf{A}$ which are related to the electromagnetic fields by the MKS equations

$$\mathbf{E} = -\text{grad } \Phi - \partial\mathbf{A}/\partial t\ , \qquad\qquad \mathbf{B} = \text{curl } \mathbf{A}\ . \tag{211}$$

The Lagrangian is

$$L(\mathbf{r},\mathbf{v},t) = -mc^2\sqrt{1-(v/c)^2} - e\left[\Phi(\mathbf{r},t) - \mathbf{v}\cdot\mathbf{A}(\mathbf{r},t)\right]\ , \tag{212}$$

in which $\mathbf{r}$ is the physical position vector and $\mathbf{v}$ is the physical velocity. In our coordinate system, the velocity is

$$\mathbf{v} = \hat{\mathbf{x}}\dot{x} + \hat{\mathbf{y}}\dot{y} + \hat{\mathbf{s}}(1+hx)\dot{s}\ , \tag{213}$$

where $\dot{x}$, $\dot{y}$ and $\dot{s}$ are the generalized velocities. The vector potential is just

$$\mathbf{A} = \hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{s}}A_s\ , \tag{214}$$

where $A_x = (\hat{\mathbf{x}}\cdot\mathbf{A})$ and so on. The canonical momenta are

$$P_x = \frac{\partial L}{\partial\dot{x}} = \frac{m\dot{x}}{\sqrt{1-(v/c)^2}} + eA_x$$

$$P_y = \frac{\partial L}{\partial \dot{y}} = \frac{m\dot{y}}{\sqrt{1-(v/c)^2}} + eA_y \tag{215}$$
$$P_s = \frac{\partial L}{\partial \dot{s}} = \frac{m(1+hx)^2\dot{s}}{\sqrt{1-(v/c)^2}} + e(1+hx)A_s \ .$$

(In this section we denote the canonical coordinates and momenta by capital Q's and P's rather than lower case ones, because we reserve the lower case p for the kinetic momentum and its components. Of course $Q_1 = x$, $Q_2 = y$ and $Q_3 = s$.) The Hamiltonian is to be formed from the Lagrangian, the generalized coordinates and the canonical momenta according to the prescription,

$$H = \sum_k P_k \dot{Q}_k - L \ . \tag{216}$$

with all appearances of the generalized velocities removed by substituting their values in terms of the canonical momenta. The generalized velocities appear in (215) both explicitly and through the physical velocity v, so it is convenient to write the mechanical momentum vector $\mathbf{p}$ in terms of the canonical momenta.

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-(v/c)^2}} = \hat{\mathbf{x}}(P_x - eA_x) + \hat{\mathbf{y}}(P_y - eA_y) + \hat{\mathbf{s}}\left(\frac{P_s}{1+hx} - eA_s\right) \ , \tag{217}$$

from which we find

$$v = \frac{\pm\, cp}{\sqrt{p^2 + m^2c^2}} \ . \tag{218}$$

Choosing the positive sign so that a positive velocity corresponds to a positive momentum, we get the generalized velocities in terms of the canonical momenta.

$$\dot{x} = (P_x - eA_x)\frac{c}{\sqrt{p^2+m^2c^2}} \tag{219}$$
$$\dot{y} = (P_y - eA_y)\frac{c}{\sqrt{p^2+m^2c^2}} \tag{220}$$
$$\dot{s} = \left(\frac{P_s}{1+hx} - eA_s\right)\frac{c}{(1+hx)\sqrt{p^2+m^2c^2}} \tag{221}$$

Now we can carry out (216) and arrive at the Hamiltonian for our coordinate system, but still a Hamiltonian for which time is the independent variable.[4]

$$H = c\sqrt{(P_x - eA_x)^2 + (P_y - eA_y)^2 + \left(\frac{P_s}{1+hx} - eA_s\right)^2 + m^2c^2} \quad + \quad e\Phi \tag{222}$$

[4]Note that in the curvilinear coordinate system of beam dynamics, the Hamiltonian cannot be written in the oft-quoted vectorial form, $c[(\mathbf{P} - e\mathbf{A})^2 + m^2c^2]^{1/2} + e\Phi$, which is valid only in cartesian coordinates.

The Hamiltonian equations of motion are

$$\dot{x} = \frac{\partial H}{\partial P_x} \ , \qquad \dot{y} = \frac{\partial H}{\partial P_y} \ , \qquad \dot{s} = \frac{\partial H}{\partial P_s} \ , \tag{223}$$

$$\dot{P}_x = -\frac{\partial H}{\partial x} \ , \qquad \dot{P}_y = -\frac{\partial H}{\partial y} \ , \qquad \dot{P}_s = -\frac{\partial H}{\partial s} \ , \tag{224}$$

and from these it follows that the total time derivative of the Hamiltonian is $\dot{H} = \partial H/\partial t$.

The next step is to change the independent variable from t to s. This cannot be done unless s is a monotonic function of t so that $t(s)$ is a single-valued function. In beam dynamics, the dominant velocity is normally in the s-direction, and the s-motion never reverses. Consequently we are safe in making the change. To accomplish it we make repeated use of the third equation of (223) in the following procedure. Combining it with the first of those equations, we write

$$x' = \frac{\dot{x}}{\dot{s}} = \left(\frac{\partial H}{\partial P_x}\right)_{P_s} \Big/ \left(\frac{\partial H}{\partial P_s}\right)_{P_x} = -\left(\frac{\partial P_s}{\partial P_x}\right)_H \ . \tag{225}$$

In the rightmost expression of this equation, P_s is considered as that function of the variable pairs (x, P_x), (y, P_y) and (t, H) obtained by solving (222) for P_s. All variables unmentioned in the subscripts of the partial derivatives are held constant in the differentiations. We apply the same kind of operation to all but the last equations in each column of (223). To complete the change of variables, we apply it to $\dot{H}$ to obtain

$$H' = -\frac{\partial P_s}{\partial t} \ , \tag{226}$$

and note that

$$t' = \frac{1}{\dot{s}} = \frac{\partial P_s}{\partial H} \ . \tag{227}$$

The result is a Hamiltonian system of equations in which the pair $(s, -P_s)$ become the independent variable and the Hamiltonian respectively,

$$\mathcal{H} = -P_s(x, P_x, y, P_y, t, -H) = -(1 + hx)\Psi \ , \tag{228}$$

where

$$\Psi = \sqrt{\left(\frac{H - e\Phi}{c}\right)^2 - (P_x - eA_x)^2 - (P_y - eA_y)^2 - m^2c^2} \ + \ eA_s \tag{229}$$

and the pair $(t, -H)$ become a new canonical coordinate-momentum pair whose evolution is to be found, along with that of the other pairs by solving the following new equations of motion.

$$x' = \frac{\partial \mathcal{H}}{\partial P_x} \ , \qquad y' = \frac{\partial \mathcal{H}}{\partial P_y} \ , \qquad t' = \frac{\partial \mathcal{H}}{\partial(-H)} \ , \tag{230}$$

$$P_x' = -\frac{\partial \mathcal{H}}{\partial x} \ , \qquad P_y' = -\frac{\partial \mathcal{H}}{\partial y} \ , \qquad -H' = -\frac{\partial \mathcal{H}}{\partial t} \ . \tag{231}$$

We note that we can identify the value of the original Hamiltonian, and therefore the meaning of the symbol H, from (215) and (222), as the total energy of the particle.

$$H = \sqrt{p^2c^2 + m^2c^4} + e\Phi \ . \tag{232}$$

The last of the equations of motion gives the rate of change of the energy.

We call shall often refer to the generalized coordinates and the conjugate momenta as the "q's" and "p's" as they are usually referred to in textbooks. In these terms, $q_1 = x,\ p_1 = P_x,\ q_2 = y,\ p_2 = P_y,\ q_3 = t$ and $p_3 = -H$.

References

[1] Goldstein, H., *Classical Mechanics (Second Edition)*, Addison-Wesley, Reading, (1980).

[2] Courant, E. and Snyder, H., *Ann. of Phys.*, **3**, 1, (1958).

[3] Ludford, G. and Yannitell, D., *Am. J. Phys.*, **36**, 231 (1968).

[4] Brown, K. *et al., Transport, a Computer Program for Designing Beam Transport Systems,* Report SLAC-91, Rev. 2, (1977).

[5] Dragt, A. J. in *Physics of High Energy Particle Accelerators, Fermilab Summer School, 1981*, p. 147, American Institute of Physics, New York, (1982).

[6] Kolmogorov, A. N., *Elements of the Theory of Functions and Functional Analysis, Vol. 1*, Graylock Press, Rochester, (1957).

[7] Ronald D. Ruth, *IEEE Transactions on Nuclear Science*, **NS-30**, No 4, August 1983, and Forest, E. and Ruth, R. D., *Physica*, **D**, 43, 1990.