10 ANALYTICAL MECHANICS: CANONICAL FORMALISM

10.1 Symplectic structure of the Hamiltonian phase space

Consider the real $2l \times 2l$ matrix

$$\mathcal{I} = \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \tag{10.1}$$

(with $\mathbf{1}$ and $\mathbf{0}$ we henceforth denote the identity and the null matrix, with the obvious dimensions, e.g. $l \times l$ in (10.1)). Note that $\mathcal{I}$ is orthogonal and skew-symmetric, i.e.

$$\mathcal{I}^{-1} = \mathcal{I}^{T} = -\mathcal{I} \tag{10.2}$$

and that $\mathcal{I}^2 = -\mathbf{1}$. As observed in Chapter 8, setting $\mathbf{x} = (\mathbf{p}, \mathbf{q})$, the Hamilton equations can be written in the form

$$\dot{\mathbf{x}} = \mathcal{I}\nabla_{\mathbf{x}} H(\mathbf{x}, t). \tag{10.3}$$

Example 10.1
Let S be a real symmetric constant $2l \times 2l$ matrix. A *linear Hamiltonian system* with constant coefficients is a system of $2l$ ordinary differential equations of the form (10.3), where

$$H(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T S\mathbf{x}. \tag{10.4}$$

The Hamiltonian is then a quadratic form in $\mathbf{x}$ and (10.3) takes the form

$$\dot{\mathbf{x}} = \mathcal{J} S\mathbf{x}.$$

The solution of this system of differential equations with the initial condition $\mathbf{x}(0) = \mathbf{X}$ is given by

$$\mathbf{x}(t) = \mathrm{e}^{tB}\mathbf{X}, \tag{10.5}$$

where we set

$$B = \mathcal{J} S.$$

The matrices with this structure deserve special attention. ■

DEFINITION 10.1 *A real $2l \times 2l$ matrix B is called* Hamiltonian *(or* infinitesimally symplectic*) if*

$$B^T \mathcal{J} + \mathcal{J} B = 0. \tag{10.6}$$

■

THEOREM 10.1 *The following conditions are equivalent:*

(1) the matrix B is Hamiltonian;
(2) $B = \mathcal{J}S$, with S a symmetric matrix;
(3) $\mathcal{J}B$ is a symmetric matrix.

In addition, if B and C are two Hamiltonian matrices, B^T, βB (with $\beta \in \mathbf{R}$), $B \pm C$ and $[B, C] = BC - CB$ are Hamiltonian matrices.

Proof
From the definition of a Hamiltonian matrix it follows that

$$\mathcal{J}B = -B^T\mathcal{J} = (\mathcal{J}B)^T,$$

and hence (1) and (3) are equivalent. The equivalence of (2) and (3) is immediate, as $S = -\mathcal{J}B$.

The first three statements of the second part of the theorem are obvious (for the first, note that $B^T = -S\mathcal{J} = \mathcal{J}S'$, with $S' = \mathcal{J}S\mathcal{J}$ symmetric). Setting $B = \mathcal{J}S$ and $C = \mathcal{J}R$ (with S and R symmetric matrices) we have

$$[B, C] = \mathcal{J}(S\mathcal{J}R - R\mathcal{J}S)$$

and

$$(S\mathcal{J}R - R\mathcal{J}S)^T = -R\mathcal{J}S + S\mathcal{J}R.$$

It follows that the matrix $[B, C]$ is Hamiltonian. ■

Remark 10.1
Writing B as a $2l \times 2l$ block matrix

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where a, b, c, d are $l \times l$ matrices, (10.6) becomes

$$B^T\mathcal{J} + \mathcal{J}B = \begin{pmatrix} -c + c^T & -a^T - d \\ a + d^T & b - b^T \end{pmatrix},$$

and hence B is Hamiltonian if and only if b and c are symmetric matrices and $a^T + d = 0$. If $l = 1$, B is Hamiltonian if and only if it has null trace. ■

Remark 10.2
From Theorem 10.1 it follows that the Hamiltonian matrices form a group (with

respect to matrix sum) called $\mathrm{sp}(l, \mathbf{R})$. If we identify the vector space of real $2l \times 2l$ matrices with $\mathbf{R}^{4l^2}$, the Hamiltonian matrices form a linear subspace, of dimension $l(2l + 1)$ (indeed, from what was previously discussed we may choose $l(l + 1)/2$ elements of the matrices b and c and, for example, l^2 elements of the matrix a). In addition, since the *Lie product* (or *commutator*) $[\ ,\ ]$ preserves the group of Hamiltonian matrices, $\mathrm{sp}(l, \mathbf{R})$ has a *Lie algebra* structure (see Arnol'd 1978a). ■

DEFINITION 10.2 *A real $2l \times 2l$ matrix A is called* symplectic *if*

$$A^T\mathcal{J}A = \mathcal{J}. \tag{10.7}$$

■

THEOREM 10.2 *Symplectic $2l \times 2l$ matrices form a group under matrix multiplication, denoted by $\mathrm{Sp}(l, \mathbf{R})$. The transpose of a symplectic matrix is symplectic.*

Proof
Evidently the $2l \times 2l$ identity matrix is symplectic, and if A satisfies (10.7) then it is necessarily non-singular, since from (10.7) it follows that

$$(\det(A))^2 = 1. \tag{10.8}$$

In addition, it can be easily seen that

$$A^{-1} = -\mathcal{J}A^T\mathcal{J}, \tag{10.9}$$

so that

$$(A^{-1})^T\mathcal{J}A^{-1} = (A^T)^{-1}\mathcal{J}(-\mathcal{J}A^T\mathcal{J}) = (A^T)^{-1}A^T\mathcal{J} = \mathcal{J},$$

i.e. A^{-1} is symplectic. If C is another symplectic matrix, we immediately have that

$$(AC)^T\mathcal{J}AC = C^TA^T\mathcal{J}AC = C^T\mathcal{J}C = \mathcal{J}.$$

In addition, $A^T = -\mathcal{J}A^{-1}\mathcal{J}$, from which it follows that

$$A\mathcal{J}A^T = AA^{-1}\mathcal{J} = \mathcal{J}.$$

■

Example 10.2
The group of symplectic 2×2 matrices with real coefficients, $\mathrm{Sp}(1, \mathbf{R})$, coincides with the group $\mathrm{SL}(2, \mathbf{R})$ of matrices with determinant 1. Indeed, if

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

the symplecticity condition becomes

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

Hence A is symplectic if and only if $\det(A) = \alpha\delta - \beta\gamma = 1$. It follows that every symplectic 2×2 matrix defines a linear transformation preserving area and orientation. The orthogonal unit matrices (with determinant equal to 1) are a subgroup of $\mathrm{SL}(2,\mathbf{R})$, and hence also of $\mathrm{Sp}(1,\mathbf{R})$. ■

Remark 10.3
Let A be a symplectic $2l \times 2l$ matrix. We write it as an $l \times l$ block matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{10.10}$$

The condition that the matrix is symplectic then becomes

$$A^T \mathcal{J} A = \begin{pmatrix} -a^Tc + c^Ta & -a^Td + c^Tb \\ -b^Tc + d^Ta & -b^Td + d^Tb \end{pmatrix} = \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix}, \tag{10.11}$$

and hence A is symplectic only if a^Tc and b^Td are $l \times l$ symmetric matrices and $a^Td - c^Tb = \mathbf{1}$. The symplecticity condition is therefore more restrictive in dimension $l > 1$ than in dimension $l = 1$, when it becomes simply $\det(A) = 1$. It is not difficult to prove (see Problem 1) that *symplectic matrices have determinant equal to 1 for every* l (we have already seen that $\det(A) = \pm 1$, see (10.8)). ■

Remark 10.4
Symplectic matrices have a particularly simple inverse: from (10.9) and (10.10) it follows immediately that

$$A^{-1} = \begin{pmatrix} d^T & -b^T \\ -c^T & a^T \end{pmatrix}. \tag{10.12}$$

■

Remark 10.5
If we identify the vector space of the $2l\times 2l$ matrices with $\mathbf{R}^{4l^2}$, the group $\mathrm{Sp}(l,\mathbf{R})$ defines a regular submanifold of $\mathbf{R}^{4l^2}$ of dimension $l(2l+1)$ (this can be verified immediately in view of the conditions expressed in Remark 10.3; indeed, starting from the dimension of the ambient space, $4l^2$, we subtract $2(l(l-1))/2$, since the matrices a^Tc and b^Td must be symmetric, and l^2 since $a^Td - c^Tb = \mathbf{1}$.) ■

PROPOSITION 10.1 *The tangent space to* $\mathrm{Sp}(l,\mathbf{R})$ *at* $\mathbf{1}$ *is the space of Hamiltonian matrices:*

$$T_{\mathbf{1}}\mathrm{Sp}(l,\mathbf{R}) = \mathrm{sp}(l,\mathbf{R}). \tag{10.13}$$

Proof
Let $A(t)$ be a curve in $\mathrm{Sp}(l,\mathbf{R})$ passing through $\mathbf{1}$ when $t = 0$, and hence such that

$$A(t)^T \mathcal{J} A(t) = \mathcal{J} \tag{10.14}$$

for every t and $A(0) = \mathbf{1}$.

By differentiating (10.14) with respect to t we find

$$\dot{A}^T \mathcal{J} A + A^T \mathcal{J} \dot{A} = 0,$$

from which, setting $B = \dot{A}(0) \in T_{\mathbf{1}}\mathrm{Sp}(l,\mathbf{R})$

$$B^T \mathcal{J} + \mathcal{J} B = 0,$$

and hence $B \in \mathrm{sp}(l,\mathbf{R})$. ■

Conversely, to every Hamiltonian matrix there corresponds a curve in $\mathrm{Sp}(l,\mathbf{R})$, as shown in the following.

PROPOSITION 10.2 *Let B be a Hamiltonian matrix. The matrix $A(t) = \mathrm{e}^{tB}$ is symplectic for every $t \in \mathbf{R}$.*

Proof
We must show that $A(t)$ satisfies (10.7) for every t, i.e.

$$(\mathrm{e}^{tB})^T \mathcal{J} \mathrm{e}^{tB} = \mathcal{J}.$$

It follows immediately from the definition

$$\mathrm{e}^{tB} = \sum_{n=0}^{\infty} \frac{t^n}{n!} B^n$$

that $(\mathrm{e}^{tB})^T = \mathrm{e}^{tB^T}$, and $(\mathrm{e}^{tB})^{-1} = \mathrm{e}^{-tB}$.

Hence the condition for the matrix to be symplectic becomes

$$\mathrm{e}^{tB^T} \mathcal{J} = \mathcal{J} \mathrm{e}^{-tB}.$$

But

$$\mathrm{e}^{tB^T} \mathcal{J} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \left(B^T\right)^{n-1} B^T \mathcal{J} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \left(B^T\right)^{n-1} (-\mathcal{J} B).$$

Iterating, we find

$$\mathrm{e}^{tB^T} \mathcal{J} = \mathcal{J} \sum_{n=0}^{\infty} \frac{t^n}{n!} (-1)^n B^n = \mathcal{J} \mathrm{e}^{-tB}.$$

■

DEFINITION 10.3 *The* symplectic product *on a real vector space V of dimension $2l$ is a skew-symmetric, non-degenerate bilinear form $\omega : V \times V \to \mathbf{R}$. The* *space V endowed with a symplectic product has a* symplectic structure *and V is a* symplectic space.

We recall that a bilinear skew-symmetric form is *non-degenerate* if and only if $\omega(v_1, v_2) = 0$ for every $v_2 \in V$ implies $v_1 = 0$. We note also that only vector spaces of even dimension admit a symplectic structure. Indeed, all bilinear skew-symmetric forms are necessarily degenerate in a space of odd dimension.

Consider the canonical basis $\mathbf{e}_1, \ldots, \mathbf{e}_{2l}$ in $\mathbf{R}^{2l}$. The symplectic product ω has a matrix representation W obtained by setting

$$W_{ij} = \omega(\mathbf{e}_i, \mathbf{e}_j).$$

Evidently the representative matrix W is skew-symmetric and the non-degeneracy condition is equivalent to $\det(W) \neq 0$. Moreover, for every $\mathbf{x}$, $\mathbf{y} \in \mathbf{R}^{2l}$ We have

$$\omega(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{2l} W_{ij} x_i y_j = \mathbf{x}^T W \mathbf{y}. \tag{10.15}$$

By choosing the matrix $W = \mathfrak{I}$ we obtain the so-called *standard symplectic product* (henceforth simply referred to as symplectic product unless there is a possibility of confusion) and correspondingly the standard symplectic structure.

Remark 10.6
The standard symplectic product has an interesting geometric characterisation. Given two vectors $\mathbf{x}$, $\mathbf{y}$ we have

$$\begin{aligned}\mathbf{x}^T \mathfrak{I} \mathbf{y} &= -x_1 y_{l+1} - \ldots - x_l y_{2l} + x_{l+1} y_1 + \ldots + x_{2l} y_l \\ &= (x_{l+1} y_1 - x_1 y_{l+1}) + \ldots + (x_{2l} y_l - x_l y_{2l}),\end{aligned}$$

corresponding to the sum of the (oriented) areas of the projection of the parallelogram with sides $\mathbf{x}$, $\mathbf{y}$ on the l planes $(x_1, x_{l+1}), \ldots, (x_l, x_{2l})$. Hence, if $\mathbf{p}$ is the vector constructed with the first l components of $\mathbf{x}$, and $\mathbf{q}$ is the one constructed with the remaining components, we have $\mathbf{x} = (\mathbf{p}, \mathbf{q})$, and analogously if $\mathbf{y} = (\mathbf{p}', \mathbf{q}')$, we have

$$\omega(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathfrak{I} \mathbf{y} = (q_1 p'_1 - p_1 q'_1) + \ldots + (q_l p'_l - p_l q'_l). \tag{10.16}$$

Note that in $\mathbf{R}^2$ the symplectic product of two vectors coincides with the unique non-zero scalar component of their vector product. ■

Definition 10.4 *Suppose we are given a symplectic product in $\mathbf{R}^{2l}$. A* symplectic basis *is a basis of $\mathbf{R}^{2l}$ with respect to which the symplectic product takes the standard form (10.16), and hence it has as representative matrix the matrix $\mathfrak{I}$.* ■

Given a symplectic product ω, a symplectic basis $\mathbf{e}_1, \ldots, \mathbf{e}_{2l} = \mathbf{e}_{p_1}, \ldots, \mathbf{e}_{p_l}, \mathbf{e}_{q_1}, \ldots, \mathbf{e}_{q_l}$ satisfies

$$\omega(\mathbf{e}_{q_i}, \mathbf{e}_{q_j}) = \omega(\mathbf{e}_{p_i}, \mathbf{e}_{p_j}) = 0, \tag{10.17}$$

for every $i, j = 1, \ldots, l$ and

$$\omega(\mathbf{e}_{q_i}, \mathbf{e}_{p_j}) = \delta_{ij}. \tag{10.18}$$

Remark 10.7
It follows that the choice of standard symplectic structure for $\mathbf{R}^{2l}$ coincides with the choice of the canonical basis of $\mathbf{R}^{2l}$ as symplectic basis. ■

Using a technique similar to the Gram–Schmidt orthonormalisation for the basis in an inner product space, it is not difficult to prove the following theorem.

Theorem 10.3 *In any space endowed with a symplectic product it is possible to construct a symplectic basis.* ■

As for inner product spaces, it is possible to choose as the first vector of the basis any non-zero vector.

Pursuing the analogy between an inner and a symplectic product, we can define a class of transformations that preserve the symplectic product, taking as a model the orthogonal transformations, which preserve the inner product.

Definition 10.5 *Given two symplectic spaces V_1, ω_1 and V_2, ω_2, a linear map $S : V_1 \to V_2$ is* symplectic *if $\omega_2(S(\mathbf{v}), S(\mathbf{w})) = \omega_1(\mathbf{v}, \mathbf{w})$ for every $\mathbf{v}$, $\mathbf{w} \in V_1$. If moreover S is an isomorphism, we say that S is a* symplectic isomorphism. ■

Remark 10.8
From Theorem 10.3 it follows, as an obvious corollary, that all symplectic spaces of the same dimension are symplectically isomorphic. A 'canonical' isomorphism can be obtained by choosing a symplectic basis in each space, and setting a correspondence between the basis elements with the same index. In particular, all symplectic spaces of dimension $2l$ are symplectically isomorphic to $\mathbf{R}^{2l}$ with its standard structure. ■

Theorem 10.4 *Let $\mathbf{R}^{2l}$ be considered with its standard structure. A linear map $S : \mathbf{R}^{2l} \to \mathbf{R}^{2l}$ is symplectic if and only if its representative matrix is symplectic.*

Proof
This is a simple check: given $\mathbf{x}, \mathbf{y} \in \mathbf{R}^{2l}$ we have

$$\omega(S\mathbf{x}, S\mathbf{y}) = (S\mathbf{x})^T \mathfrak{I} S\mathbf{y} = \mathbf{x}^T S^T \mathfrak{I} S \mathbf{y},$$

which is equal to

$$\omega(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathfrak{I} \mathbf{y}$$

for every $\mathbf{x}$, $\mathbf{y}$ if and only if

$$S^T \mathfrak{I} S = \mathfrak{I}.$$ ■

We conclude this section with the definition and characterisation of Hamiltonian vector fields (or symplectic gradient vector fields). These are useful in view of the fact that the Hamilton equations can be written in the form (10.3).

DEFINITION 10.6 *A vector field* $\mathbf{X}(\mathbf{x},t)$ *in* $\mathbf{R}^{2l}$ *is* Hamiltonian *if there exists a function* $f(\mathbf{x},t)$ *in* $\mathcal{C}^2$ *such that*

$$\mathbf{X}(\mathbf{x},t) = \mathcal{J}\nabla_{\mathbf{x}} f(\mathbf{x},t). \tag{10.19}$$

In this case f *is called the* Hamiltonian corresponding to the field $\mathbf{X}$ *and the field* $\mathbf{X}$ *is called the* symplectic gradient of f. *If* $\mathbf{X}$ *is Hamiltonian, the system of differential equations*

$$\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x},t) \tag{10.20}$$

is called Hamiltonian. ■

The system of Example 10.1 is Hamiltonian.

Remark 10.9
A Hamiltonian vector field determines the corresponding Hamiltonian f up to an arbitrary function $h(t)$ depending only on time t. This arbitrariness can be removed by requiring that the Hamiltonian associated with the field $\mathbf{X} = \mathbf{0}$ be zero. ■

Remark 10.10
In $\mathbf{R}^2$ the vector $\mathbf{w} = \mathcal{J}\mathbf{v}$ can be obtained by rotating $\mathbf{v}$ by $\pi/2$ in the positive direction. It is easy to check that, in $\mathbf{R}^{2l}$, $\mathcal{J}\mathbf{v}$ is normal to $\mathbf{v}$. It follows that in a Hamiltonian field, for every fixed t, the Hamiltonian is constant along the lines of the field (Fig. 10.1). If the field is independent of time the Hamiltonian is constant along its integral curves, i.e. along the Hamiltonian flow (recall equation (8.26)). ■

It is essential to characterise Hamiltonian vector fields. This is our next aim.

THEOREM 10.5 *A necessary and sufficient condition for a vector field* $\mathbf{X}(\mathbf{x},t)$ *in* $\mathbf{R}^{2l}$ *to be Hamiltonian is that the Jacobian matrix* $\nabla_{\mathbf{x}}\mathbf{X}(\mathbf{x},t)$ *is Hamiltonian for every* $(\mathbf{x},t)$.

Proof
The condition is necessary. Indeed, if f is the Hamiltonian corresponding to $\mathbf{X}$ we have that

$$\frac{\partial X_i}{\partial x_j} = \sum_{k=1}^{l} \mathcal{J}_{ik} \frac{\partial^2 f}{\partial x_k \partial x_j},$$

and hence the matrix $\nabla_{\mathbf{x}}\mathbf{X}$ can be written as the product of the matrix $\mathcal{J}$ and the Hessian matrix of f, which is evidently symmetric.

The condition is also sufficient: if $\nabla_{\mathbf{x}}\mathbf{X}(\mathbf{x},t)$ is Hamiltonian for every $(\mathbf{x},t)$, setting $\mathbf{Y}(\mathbf{x},t) = \mathcal{J}\mathbf{X}(\mathbf{x},t)$, by (3) of Theorem 10.1, we have that

$$\frac{\partial Y_i}{\partial x_j} = \frac{\partial Y_j}{\partial x_i}.$$

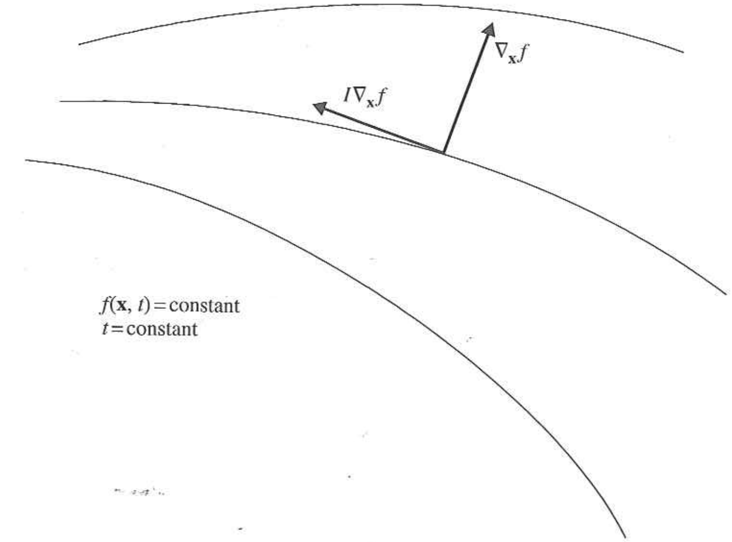

Fig. 10.1

Consequently, there exists a function $f(\mathbf{x},t)$ such that

$$\mathbf{Y}(\mathbf{x},t) = -\nabla_{\mathbf{x}} f(\mathbf{x},t).$$

From this it follows that

$$\mathbf{X}(\mathbf{x},t) = -\mathcal{J}\mathbf{Y}(\mathbf{x},t) = \mathcal{J}\nabla_{\mathbf{x}} f(\mathbf{x},t).$$ ■

Example 10.3
Consider the system of differential equations

$$\dot{p} = -p^{\alpha+1}q^{\delta}, \quad \dot{q} = p^{\alpha}q^{\beta},$$

and compute for which values of the real constants α, β and δ this is a Hamiltonian system. Find the corresponding Hamiltonian $H(q,p)$.

Consider the second equation; if there exists a Hamiltonian $H(p,q)$ such that $\dot{q} = \partial H/\partial p$, by integrating with respect to p we find:

(a) $H = q^{\beta}\log p + f(q)$ if $\alpha = -1$;
(b) $H = p^{\alpha+1}q^{\beta}/(\alpha+1) + g(q)$ if $\alpha \neq -1$.

By substituting in the equation $\dot{p} = -\partial H/\partial q$ and comparing with the equation given for p, we find that, if $\alpha = -1$, necessarily $\beta = 0$ and

(a′) $H = \log p + \{q^{\delta+1}/(\delta+1) + c\}$ if $\delta \neq -1$, where c is an arbitrary constant;
(a″) $H = \log p + \log q + c$ if $\delta = -1$, where c is an arbitrary constant.

If on the other hand $\alpha \neq -1$ we find $H = \{(qp)^{\alpha+1}/(\alpha+1)\} + c$, where as usual c is an arbitrary integration constant. ■

10.2 Canonical and completely canonical transformations

A method which can sometimes be applied to integrate differential equations is to use an appropriate change of variables which makes it possible to write the equation in a form such that the solution (or some property of the solution) can be immediately obtained. The study of particular classes of coordinate transformations in the phase space for the Hamilton equations is of great importance and will be carried out in this and the next sections. In Chapters 11 and 12 we will show how, through these transformations, it is possible to solve (exactly or approximately) the Hamilton equations for a large class of systems.

Given a system of ordinary differential equations

$$\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}, t), \tag{10.21}$$

where $\mathbf{x} \in \mathbf{R}^n$ (or a differentiable manifold of dimension n), consider an invertible coordinate transformation (possibly depending on time t)

$$\mathbf{x} = \mathbf{x}(\mathbf{y}, t), \tag{10.22}$$

with inverse

$$\mathbf{y} = \mathbf{y}(\mathbf{x}, t). \tag{10.23}$$

If the function $\mathbf{y}(\mathbf{x}, t)$ has continuous first derivatives, the system (10.21) is transformed into

$$\dot{\mathbf{y}} = \mathbf{w}(\mathbf{y}, t), \tag{10.24}$$

where

$$\mathbf{w}(\mathbf{y}, t) = J\mathbf{v} + \frac{\partial \mathbf{y}}{\partial t},$$

J is the Jacobian matrix of the transformation, $J_{ik} = \partial y_i/\partial x_k$, and the right-hand side is expressed in terms of the variables $(\mathbf{y}, t)$ using (10.22). Likewise we consider the system of canonical equations with Hamiltonian $H(\mathbf{x}, t)$, where $\mathbf{x} = (\mathbf{p}, \mathbf{q}) \in \mathbf{R}^{2l}$,

$$\dot{\mathbf{x}} = \mathcal{J}\nabla_{\mathbf{x}} H(\mathbf{x}, t), \tag{10.25}$$

and make the coordinate transformation

$$\mathbf{x} = \mathbf{x}(\mathbf{X}, t), \tag{10.26}$$

with $\mathbf{X} = (\mathbf{P}, \mathbf{Q}) \in \mathbf{R}^{2l}$, subject to the invertibility condition

$$\mathbf{X} = \mathbf{X}(\mathbf{x}, t), \tag{10.27}$$

and to the condition of continuity of the first derivatives. Then the system of canonical equations (10.25) is transformed into a new system of $2l$ differential equations

$$\dot{\mathbf{X}} = \mathbf{W}(\mathbf{X}, t), \tag{10.28}$$

where

$$\mathbf{W}(\mathbf{X}, t) = J\mathcal{J}\nabla_{\mathbf{x}} H + \frac{\partial \mathbf{X}}{\partial t}, \tag{10.29}$$

J is the Jacobian matrix of the transformation, with components $J_{ik} = \partial X_i/\partial x_k$, and the right-hand side is expressed in terms of the variables $\mathbf{X} = (\mathbf{P}, \mathbf{Q})$. In general, the system (10.28) does not have the canonical structure (10.25), as it is not necessarily true that a Hamiltonian $K(\mathbf{X}, t)$ exists such that

$$\mathbf{W} = \mathcal{J}\nabla_{\mathbf{X}} K. \tag{10.30}$$

Example 10.4
We go back to Example 10.1 with $H(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T S\mathbf{x}$, where S is a constant symmetric matrix. Let us consider how the Hamilton equation $\dot{\mathbf{x}} = \mathcal{J}S\mathbf{x}$ is transformed when passing to the new variables $\mathbf{X} = A\mathbf{x}$, with A a constant invertible matrix. We immediately find that $\dot{\mathbf{X}} = A\mathcal{J}SA^{-1}\mathbf{X}$ and in order to preserve the canonical structure we must have $A\mathcal{J}SA^{-1} = \mathcal{J}C$, with C symmetric. It is important to note that this must happen *for every symmetric matrix* S, and hence this is a genuine restriction on the class to which A must belong. We can rewrite this condition as $A^T\mathcal{J}A\mathcal{J}S = -A^TCA$. It follows that the existence of a symmetric matrix C is equivalent to the symmetry condition

$$A^T\mathcal{J}A\mathcal{J}S = S\mathcal{J}A^T\mathcal{J}A, \tag{10.31}$$

i.e. $\Lambda^T\mathcal{J}S + S\mathcal{J}\Lambda = 0$ with $\Lambda = A^T\mathcal{J}A = -\Lambda^T$, for every symmetric matrix S. If A is symplectic then $\Lambda = \mathcal{J}$ and the condition is satisfied. The same is true if $\Lambda = a\mathcal{J}$ (with $a \neq 0$ so that A is invertible). These conditions are also necessary. Indeed, using the $l \times l$ block decomposition we have $\Lambda = \begin{pmatrix} \lambda & \mu \\ -\mu^T & \nu \end{pmatrix}$ and $S = \begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix}$, with the conditions $\lambda^T = -\lambda$, $\nu^T = -\nu$, $\alpha^T = \alpha$, $\gamma^T = \gamma$. The equation $\Lambda\mathcal{J}S = S\mathcal{J}\Lambda$ leads to the system

$$-\lambda\beta^T + \mu\alpha = \alpha\mu^T + \beta\lambda,$$
$$-\lambda\gamma + \mu\beta = -\alpha\nu + \beta\mu,$$
$$\mu^T\beta^T + \nu\alpha = \beta^T\mu^T + \gamma\lambda,$$
$$\mu^T\gamma + \nu\beta = -\beta^T\nu + \gamma\nu.$$

Considering the particular case $\alpha = \gamma = 0$ we find that μ must commute with every $l \times l$ matrix, and therefore $\mu = a1$. Choosing $\alpha = \beta = 0$ we find $\lambda = 0$. From $\beta = \gamma = 0$ it follows that $\nu = 0$. Hence $\Lambda = a\mathcal{J}$, and in addition, from $A^T\mathcal{J}A = a\mathcal{J}$ it follows that $\mathcal{J}A\mathcal{J} = -a(A^{-1})^T$. We finally find that $C = a(A^{-1})^T SA^{-1}$ and the new Hamiltonian is $K(\mathbf{X}) = \frac{1}{2}\mathbf{X}^T C\mathbf{X}$. If A is symplectic it holds that $K(\mathbf{X}) = H(\mathbf{x})$, and if $a \neq 1$ we find $K(\mathbf{X}) = aH(\mathbf{x})$. ■

The necessity to preserve the canonical structure of the Hamilton equations, which has very many important consequences (see the following sections and Chapter 11), justifies the following definition.

DEFINITION 10.7 *A coordinate transformation* $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ *which is differentiable and invertible (for every fixed t)* preserves the canonical structure of Hamilton equations *if for any Hamiltonian* $H(\mathbf{x}, t)$ *there exists a corresponding function* $K(\mathbf{X}, t)$, *the* new Hamiltonian, *such that the system of transformed equations (10.28) coincides with the system of Hamilton equations (10.30) for K:*

$$\begin{aligned} \dot{P}_i &= -\frac{\partial K}{\partial Q_i}(\mathbf{Q}, \mathbf{P}, t), \quad i = 1, \ldots, l, \\ \dot{Q}_i &= \frac{\partial K}{\partial P_i}(\mathbf{Q}, \mathbf{P}, t), \quad i = 1, \ldots, l. \end{aligned} \tag{10.32}$$

■

Remark 10.11
The new Hamiltonian $K(\mathbf{Q}, \mathbf{P}, t)$ is not necessarily obtained by substituting into $H(\mathbf{q}, \mathbf{p}, t)$ the transformation (10.26). This is illustrated in the following examples. ■

Example 10.5
The translations of $\mathbf{R}^{2l}$ preserve the canonical structure of the Hamilton equations. The rotations $\mathbf{X} = R\mathbf{x}$, where R is an orthogonal matrix $R^T = R^{-1}$, preserve the structure if and only if R is a symplectic matrix (see Theorem 10.6 below). This is always true for $l = 1$, if R preserves the orientation of the plane (see Example 10.2), and hence if $\det(R) = 1$. ■

Example 10.6
The transformations

$$\begin{aligned} P_i &= \nu_i p_i, \quad i = 1, \ldots, l, \\ Q_i &= \mu_i q_i, \quad i = 1, \ldots, l, \end{aligned} \tag{10.33}$$

where $\mu_1, \ldots, \mu_l$ and $\nu_1, \ldots, \nu_l$ are $2l$ real arbitrary non-zero constants satisfying the condition $\mu_i\nu_i = \lambda$ for every $i = 1, \ldots, l$, are called *scale transformations* and preserve the canonical structure of the Hamilton equations. Indeed, it can be verified that the new Hamiltonian K is related to the old one H through

$$K(\mathbf{P}, \mathbf{Q}, t) = \lambda H(\nu_1^{-1}P_1, \ldots, \nu_l^{-1}P_l, \mu_1^{-1}Q_1, \ldots, \mu_l^{-1}Q_l, t).$$

Note that K is the transform of H only in the case that $\mu_i\nu_i = 1$, $i = 1, \ldots, l$, and hence if $\lambda = 1$ (in this case the Jacobian matrix of the transformation is symplectic). When $\lambda \neq 1$ we say that the scale transformation is *not natural.* Note that the Jacobian determinant of (10.33) is λ^l, and hence the transformation (10.33) preserves the measure if and only if $\lambda = 1$. The scale transformations are commonly used to change to dimensionless coordinates. ■

Example 10.7
Let $a(t)$ be a differentiable non-zero function. The transformation

$$\mathbf{Q} = a(t)\mathbf{q}, \quad \mathbf{P} = \frac{1}{a(t)}\mathbf{p}$$

preserves the canonical structure of the Hamilton equations. Indeed, the Hamilton equations become

$$\begin{aligned} \dot{\mathbf{P}} &= -\frac{1}{a(t)}\nabla_{\mathbf{q}} H - \frac{\dot{a}(t)}{a^2(t)}\mathbf{p}, \\ \dot{\mathbf{Q}} &= a(t)\nabla_{\mathbf{p}} H + \dot{a}(t)\mathbf{q}, \end{aligned}$$

corresponding to the Hamilton equations for the function

$$K(\mathbf{P}, \mathbf{Q}, t) = H\left(a(t)\mathbf{P}, \frac{\mathbf{Q}}{a(t)}, t\right) + \frac{\dot{a}(t)}{a(t)}\mathbf{P} \cdot \mathbf{Q}.$$ ■

Example 10.8
The transformation exchanging (up to sign) the coordinates q_i with the corresponding kinetic moments p_i preserves the canonical structure of the Hamilton equations

$$\mathbf{P} = -\mathbf{q}, \quad \mathbf{Q} = \mathbf{p}. \tag{10.34}$$

The new Hamiltonian is related to the old Hamiltonian through

$$K(\mathbf{P}, \mathbf{Q}, t) = H(\mathbf{Q}, -\mathbf{P}, t).$$

This transformation shows how, within the Hamiltonian formalism, there is no essential difference between the role of the coordinates $\mathbf{q}$ and of the conjugate momenta $\mathbf{p}$. ■

Example 10.9
The *point transformations* preserve the canonical structure of the Hamilton equations. Indeed, let

$$\mathbf{Q} = \mathbf{Q}(\mathbf{q}) \tag{10.35}$$

be an invertible Lagrangian coordinate transformation. The generalised velocities are transformed linearly:

$$\dot{Q}_i = \frac{\partial Q_i}{\partial q_j}(\mathbf{q})\dot{q}_j = J_{ij}(\mathbf{q})\dot{q}_j,$$

where $i = 1, \dots, l$ and we have adopted the convention of summation over repeated indices. Here $J(\mathbf{q}) = (J_{ij}(\mathbf{q}))$ is the Jacobian matrix of the transformation (10.35). If $L(\mathbf{q}, \dot{\mathbf{q}}, t)$ is the Lagrangian of the system, we denote by

$$\hat{L}(\mathbf{Q}, \dot{\mathbf{Q}}, t) = L(\mathbf{q}(\mathbf{Q}), J^{-1}(\mathbf{q}(\mathbf{Q}))\dot{\mathbf{Q}}, t)$$

the Lagrangian expressed through the new coordinates, and by $\mathbf{P}$ the corresponding kinetic momentum, whose components are given by

$$P_i = \frac{\partial \hat{L}}{\partial \dot{Q}_i} = J_{ji}^{-1} \frac{\partial L}{\partial \dot{q}_j} = J_{ji}^{-1} p_j,$$

for $i = 1, \dots, l$. The transformation (10.35) induces a transformation of the conjugate kinetic momenta:

$$\mathbf{P} = (J^T)^{-1}\mathbf{p}, \tag{10.36}$$

and Hamilton's equations associated with the Hamiltonian $H(\mathbf{p}, \mathbf{q}, t)$ become

$$\begin{aligned} \dot{P}_i &= -J_{ji}^{-1} \frac{\partial H}{\partial q_j} + p_j \frac{\partial J_{ji}^{-1}}{\partial Q_k} J_{kn} \frac{\partial H}{\partial p_n}, \\ \dot{Q}_i &= J_{ij} \frac{\partial H}{\partial p_j}, \end{aligned} \tag{10.37}$$

where $i = 1, \dots, l$.

Point transformations necessarily preserve the canonical structure. For the Hamiltonian systems originating from a Lagrangian, the proof is easy. Indeed, starting from the new Lagrangian $\hat{L}(\mathbf{Q}, \dot{\mathbf{Q}}, t)$ we can construct the Legendre transform $\hat{H}(\mathbf{P}, \mathbf{Q}, t)$ to take the role of the Hamiltonian in the equations thus obtained. It is easy to check that $\hat{H}$ is the transform of H:

$$\hat{H}(\mathbf{P}, \mathbf{Q}, t) = H(J^T(\mathbf{q}(\mathbf{Q}))\mathbf{P}, \mathbf{q}(\mathbf{Q}), t).$$

Indeed, to obtain the Legendre transform (8.19) of $\hat{L}(\mathbf{Q}, \dot{\mathbf{Q}}, t)$ we must compute

$$\hat{H}(\mathbf{P}, \mathbf{Q}, t) = \mathbf{P}^T\dot{\mathbf{Q}} - \hat{L}(\mathbf{Q}, \dot{\mathbf{Q}}, t),$$

and reintroducing the variables $(\mathbf{p}, \mathbf{q})$ we note that $\hat{L}$ goes to L, while $\mathbf{P}^T\dot{\mathbf{Q}} = \mathbf{p}^T J^{-1} J\dot{\mathbf{q}} = \mathbf{p}^T\dot{\mathbf{q}}$. It follows that $\hat{H}(\mathbf{P}, \mathbf{Q}, t) = H(\mathbf{p}, \mathbf{q}, t)$. We leave it to the reader to verify that (10.37) are the Hamilton equations associated with $\hat{H}$. ∎

DEFINITION 10.8 *A differentiable and invertible coordinate transformation* $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ *(for every fixed t) is called* canonical *if the Jacobian matrix*

$$J(\mathbf{x}, t) = \nabla_{\mathbf{x}}\mathbf{X}(\mathbf{x}, t)$$

is symplectic for every choice of $(\mathbf{x}, t)$ *in the domain of definition of the transformation. A time-independent canonical transformation* $\mathbf{X} = \mathbf{X}(\mathbf{x})$ *is called* completely canonical. ∎

We systematically assume in what follows that the matrix J is sufficiently regular (at least C^1). All arguments are local (i.e. are valid in an open connected subset of $\mathbf{R}^{2l}$).

Example 10.10
It can immediately be verified that the transformation considered in Example 10.7 is canonical, and those considered in Examples 10.22, 10.25 and 10.26 are completely canonical. The scale transformations (Example 10.5) are not canonical, except when $\lambda = 1$. ∎

Remark 10.12
Recall that symplectic matrices form a group under matrix multiplication. Then we immediately deduce that *the canonical transformations form a group.* The completely canonical transformations form a subgroup, usually denoted by $\mathrm{SDiff}(\mathbf{R}^{2l})$. We also note that $\det J = 1$, and hence *canonical transformations preserve the Lebesgue measure* in phase space. ∎

THEOREM 10.6 *The canonical transformations preserve the canonical structure of the Hamilton equations.* ∎

Before proving Theorem 10.6 it is convenient to digress and introduce a short lemma frequently used in the remainder of this chapter. We define first of all a class of $2l \times 2l$ matrices that generalises the class of symplectic matrices, by replacing the equation $J^T \mathfrak{I} J = \mathfrak{I}$ by

$$J^T \mathfrak{I} J = a\mathfrak{I}, \tag{10.38}$$

where a is a *constant* different from zero. It is immediately verified that these matrices have as inverse $J^{-1} = -(1/a)\mathfrak{I}J^T\mathfrak{I}$. This inverse belongs to the analogous class with a^{-1} instead of a. Therefore $J^T = -a\mathfrak{I}J^{-1}\mathfrak{I}$ and we can verify that J^T belongs to the same class of J, i.e. $J\mathfrak{I}J^T = a\mathfrak{I}$. Obviously the class (10.38) includes as a special case (for $a = 1$) the symplectic matrices. An important property of the time-dependent matrices that satisfy the property (10.38) (with a constant) is the following.

LEMMA 10.1 *If* $J(\mathbf{X}, t)$ *is a matrix in the class (10.38) then the matrix* $B = (\partial J/\partial t)J^{-1}$ *is Hamiltonian.*

Proof
Recalling Theorem 10.1, it is sufficient to prove that the matrix

$$A = \mathfrak{I}\frac{\partial J}{\partial t}J^{-1} \tag{10.39}$$

is symmetric. Differentiating with respect to t the two sides of (10.38) we obtain

$$\frac{\partial J^T}{\partial t}\mathcal{J}J + J^T\mathcal{J}\frac{\partial J}{\partial t} = 0. \tag{10.40}$$

Multiplying this on the left by $(J^{-1})^T$ and on the right by J^{-1} then yields

$$A^T = -(J^{-1})^T\frac{\partial J^T}{\partial t}\mathcal{J} = \mathcal{J}\frac{\partial J}{\partial t}J^{-1} = A.$$

■

We now turn to Theorem 10.6.

Proof of Theorem 10.6
Let $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ be a canonical transformation.

By differentiating $\mathbf{X}$ with respect to t and using $\dot{\mathbf{x}} = \mathcal{J}\nabla_{\mathbf{x}}H(\mathbf{x}, t)$ we find

$$\dot{\mathbf{X}} = \frac{\partial \mathbf{X}}{\partial t} + J\mathcal{J}\nabla_{\mathbf{x}}H. \tag{10.41}$$

Setting

$$\hat{H}(\mathbf{X}, t) = H(\mathbf{x}(\mathbf{X}, t), t), \tag{10.42}$$

we have that

$$\nabla_{\mathbf{x}}H = J^T\nabla_{\mathbf{X}}\hat{H}, \tag{10.43}$$

from which it follows that equation (10.41) can be written as

$$\dot{\mathbf{X}} = \frac{\partial \mathbf{X}}{\partial t} + J\mathcal{J}J^T\nabla_{\mathbf{X}}\hat{H}. \tag{10.44}$$

But J is by hypothesis symplectic, and therefore we arrive at the equation

$$\dot{\mathbf{X}} = \frac{\partial \mathbf{X}}{\partial t} + \mathcal{J}\nabla_{\mathbf{X}}\hat{H}, \tag{10.45}$$

which stresses the fact that the field $\mathcal{J}\nabla_{\mathbf{X}}\hat{H}$ is Hamiltonian.

To complete the proof we must show that $\partial\mathbf{X}/\partial t$ is also a Hamiltonian vector field. By Theorem 10.5, a necessary and sufficient condition is that $B = \nabla_{\mathbf{X}}((\partial\mathbf{X}(\mathbf{x}(\mathbf{X}, t), t))/\partial t)$ is Hamiltonian.

We see immediately that

$$B_{ij} = \frac{\partial}{\partial X_j}\frac{\partial X_i}{\partial t} = \sum_{n=1}^{2l}\frac{\partial^2 X_i}{\partial t\partial x_n}\frac{\partial x_n}{\partial X_j},$$

and hence

$$B = \frac{\partial J}{\partial t}J^{-1}. \tag{10.46}$$

Now Lemma 10.1 ends the proof. ■

Remark 10.13
The new Hamiltonian K corresponding to the old Hamiltonian H is given by

$$K = \hat{H} + K_0, \tag{10.47}$$

where $\hat{H}$ is the old Hamiltonian expressed through the new variables (see (10.42)) and K_0 is the Hamiltonian of the Hamiltonian vector field $\partial\mathbf{X}/\partial t$, and hence satisfying

$$\frac{\partial \mathbf{X}}{\partial t} = \mathcal{J}\nabla_{\mathbf{X}}K_0. \tag{10.48}$$

It follows that K_0 depends only on the transformation $\mathbf{X}(\mathbf{x}, t)$ and it is uniquely determined by it, up to an arbitrary function $h(t)$ which we always assume to be identically zero (see Remark 10.9). Here K_0 can be identified with the Hamiltonian corresponding to $H \equiv 0$. If the transformation is completely canonical we have that $K_0 \equiv 0$, and the new Hamiltonian is simply obtained by expressing the old Hamiltonian in terms of the new coordinates (consistent with the interpretation of the Hamiltonian as the total mechanical energy of the system). ■

We then have the following.

COROLLARY 10.1 *For a completely canonical transformation the new Hamiltonian is simply the transformation of the original Hamiltonian. A time-dependent canonical transformation $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ is necessarily a Hamiltonian flow, governed by the equation $\partial\mathbf{X}/\partial t = \mathcal{J}\nabla_{\mathbf{X}}K_0(\mathbf{X}, t)$.* ■

We shall see that to every Hamiltonian flow $\mathbf{X} = S^t\mathbf{x}$ we can associate a canonical transformation. Hence we can identify the class of time-dependent canonical transformations with the class of Hamiltonian flows.

Example 10.11
Consider the *time-dependent* transformation

$$p = P - at, \quad q = Q + Pt - \frac{1}{2}at^2, \tag{10.49}$$

where a is a fixed constant. We can immediately check that the transformation is canonical, with inverse given by

$$P = p + at, \quad Q = q - pt - \frac{1}{2}at^2.$$

The Hamiltonian K_0 is the solution of (see (10.48))

$$\frac{\partial P}{\partial t} = a = -\frac{\partial K_0}{\partial Q}, \quad \frac{\partial Q}{\partial t} = -p - at = -P = \frac{\partial K_0}{\partial P},$$

from which it follows that

$$K_0(P, Q) = -\frac{P^2}{2} - aQ, \tag{10.50}$$

and the new Hamiltonian $K(P,Q,t)$ corresponding to $H(p,q,t)$ is:

$$K(P,Q,t) = H\left(P - at, Q + Pt - \frac{1}{2}at^2, t\right) + K_0(P,Q) = \hat{H}(P,Q,t) - \frac{P^2}{2} - aQ.$$

■

The next theorem includes Theorem 10.6, and characterises the whole class of transformations which preserve the canonical structure of the Hamilton equations. Moreover, it characterises how these transformations act on the Hamiltonian.

THEOREM 10.7 *A necessary and sufficient condition for a differentiable and invertible (for every fixed t) coordinate transformation $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ to preserve the canonical structure of the Hamilton equations is that its Jacobian matrix belongs to the class (10.38), i.e.*

$$J\mathfrak{I}J^T = J^T\mathfrak{I}J = a\mathfrak{I} \tag{10.51}$$

for some constant a different from zero. The transformation acts on the Hamiltonian as follows:

$$K(\mathbf{X},t) = a\hat{H}(\mathbf{X},t) + K_0(\mathbf{X},t), \tag{10.52}$$

where $\hat{H}(\mathbf{X},t) = H(\mathbf{x}(\mathbf{X},t),t)$ is the transform of the original Hamiltonian and K_0 (corresponding to $H = 0$) is the Hamiltonian of the vector field $\partial\mathbf{X}/\partial t$. The transformation is canonical if and only if $a = 1$. ■

COROLLARY 10.2 *The canonical transformations are the only ones leading to a new Hamiltonian of the form $K = \hat{H} + K_0$, and the completely canonical ones are the only ones for which $K = \hat{H}$.* ■

In addition, note that when $a \neq 1$ the transformation can be made into a canonical transformation by composing it with an appropriate scale change.

The proof of Theorem 10.7 makes use of a lemma. We present the proof of this lemma as given in Benettin *et al.* (1991).

LEMMA 10.2 *Let $A(\mathbf{x},t)$ be a regular function of $(\mathbf{x},t) \in \mathbf{R}^{2l+1}$ with values in the space of real non-singular $2l \times 2l$ matrices. If for any regular function $H(\mathbf{x},t)$, the vector field $A\nabla_\mathbf{x}H$ is irrotational, then there exists a function $a : \mathbf{R} \to \mathbf{R}$ such that $A = a(t)\mathbf{1}$.*

Proof

If $A\nabla_\mathbf{x}H$ is irrotational, for every i, $j = 1, \ldots, 2l$, we have that

$$\frac{\partial}{\partial x_i}(A\nabla_\mathbf{x}H)_j = \frac{\partial}{\partial x_j}(A\nabla_\mathbf{x}H)_i. \tag{10.53}$$

Let $H = x_i$. Then

$$\frac{\partial}{\partial x_i}A_{ji} = \frac{\partial}{\partial x_j}A_{ii} \tag{10.54}$$

(note that we are *not* using the convention of summation over repeated indices!), while if we let $H = x_i^2$ then

$$\frac{\partial}{\partial x_i}(A_{ji}x_i) = \frac{\partial}{\partial x_j}(A_{ii}x_i). \tag{10.55}$$

It follows using (10.54) that

$$A_{ji} = A_{ii}\delta_{ij},$$

i.e. the matrix A is diagonal. From (10.54) it also follows that

$$\frac{\partial A_{ii}}{\partial x_j} = 0, \quad \text{if } j \neq i,$$

and therefore A has the form

$$A_{ij}(\mathbf{x},t) = a_i(x_i,t)\delta_{ij},$$

for suitable functions a_i. Using (10.53) we find that

$$a_j\frac{\partial^2 H}{\partial x_i \partial x_j} = a_i\frac{\partial^2 H}{\partial x_i \partial x_j}, \quad \text{for } j \neq i,$$

from which it follows that $a_j = a_i = a(t)$. ■

Proof of Theorem 10.7

Suppose that the transformation preserves the canonical structure, so that

$$\dot{\mathbf{X}} = \mathfrak{I}\nabla_\mathbf{X}K(\mathbf{X},t). \tag{10.56}$$

Comparing (10.56) with the general form (10.44) of the transformed equation

$$\dot{\mathbf{X}} = \frac{\partial\mathbf{X}}{\partial t} + J\mathfrak{I}J^T\nabla_\mathbf{X}\hat{H} \tag{10.57}$$

we deduce

$$\frac{\partial\mathbf{X}}{\partial t} = \mathfrak{I}\nabla_\mathbf{X}K - J\mathfrak{I}J^T\nabla_\mathbf{X}\hat{H}. \tag{10.58}$$

We also know (by hypothesis) that to $H = 0$ there corresponds a Hamiltonian K_0, for which (10.58) becomes

$$\frac{\partial\mathbf{X}}{\partial t} = \mathfrak{I}\nabla_\mathbf{X}K_0. \tag{10.59}$$

By substituting (10.59) into (10.58) and multiplying by $\mathfrak{I}$ we find

$$\nabla_\mathbf{X}(K - K_0) = -\mathfrak{I}J\mathfrak{I}J^T\nabla_\mathbf{X}\hat{H}. \tag{10.60}$$

Hence the matrix $-\mathfrak{I}J\mathfrak{I}J^T$ satisfies the assumptions of Lemma 10.2 (because $\hat{H}$ is arbitrary). It follows that there exists a function $a(t)$ such that

$$-\mathfrak{I}J\mathfrak{I}J^T = a(t)\mathbf{1}. \tag{10.61}$$

Equation (10.61) shows clearly that J satisfies equation (10.51), with a possibly depending on time. To prove that a is constant we note that, since $\partial\mathbf{X}/\partial t$ is a Hamiltonian vector field (see (10.59)), its Jacobian matrix

$$B = \nabla_{\mathbf{X}}\frac{\partial \mathbf{X}}{\partial t} = \frac{\partial J}{\partial t}J^{-1}$$

is Hamiltonian (see Theorem 10.5 and equation (10.46)). Therefore we can write (Definition 10.1)

$$\left(\frac{\partial J}{\partial t}J^{-1}\right)^T \mathfrak{I} + \mathfrak{I}\frac{\partial J}{\partial t}J^{-1} = 0. \tag{10.62}$$

This is equivalent to the statement that $(\partial/\partial t)(J^T\mathfrak{I}J) = 0$, yielding $a = $ constant. Now from (10.57) and (10.59), we can deduce the expression (10.52) for the new Hamiltonian K.

Conversely, suppose that the matrix J satisfies the condition (10.51). Then (Lemma 10.1) $(\partial J/\partial t)J^{-1} = \nabla_{\mathbf{X}}\partial\mathbf{X}/\partial t$ is a Hamiltonian matrix. Therefore, the field $\partial\mathbf{X}/\partial t$ is Hamiltonian, and we can conclude that equation (10.57) takes the form

$$\dot{\mathbf{X}} = \mathfrak{I}\nabla_{\mathbf{X}}(K_0 + a\hat{H}).$$

It follows that the transformation preserves the canonical structure, and the new Hamiltonian K is given by (10.52). ■

For the case $l = 1$, Theorem 10.7 has the following simple interpretation.

Corollary 10.3 *For $l = 1$ the condition of Theorem 10.7 reduces to*

$$\det J = \text{constant} \neq 0. \tag{10.63}$$

Proof
It is enough to note that for $l = 1$ we have $J^T\mathfrak{I}J = \mathfrak{I}\det J$. ■

Example 10.12
The transformation

$$p = \alpha\sqrt{P}\cos\gamma Q, \quad q = \beta\sqrt{P}\sin\gamma Q, \quad \alpha\beta\gamma \neq 0,$$

with α, β, γ constants, satisfies condition (10.63), since $\det J = \frac{1}{2}\alpha\beta\gamma$. It is (completely) canonical if and only if $\frac{1}{2}\alpha\beta\gamma = 1$. ■

It is useful to close this section with a remark on the transformations which are inverses of those preserving the canonical structure. These inverse transformations clearly have the same property. If $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ is a transformation in the class (10.51), its inverse $\mathbf{x} = \mathbf{x}(\mathbf{X}, t)$ has Jacobian matrix $J^{-1} = -(1/a)\mathfrak{I}J^T\mathfrak{I}$, such that $(J^{-1})^T\mathfrak{I}J^{-1} = (1/a)\mathfrak{I}$ (as we have already remarked). The inverse transformation reverts the Hamiltonian (10.52) to the original Hamiltonian H. For the case of the inverse transformation, the same relation (10.52) is then applied as follows:

$$H(\mathbf{x}, t) = K_0'(\mathbf{x}, t) + \frac{1}{a}\left[\hat{K}_0(\mathbf{x}, t) + aH(\mathbf{x}, t)\right], \tag{10.64}$$

where $\hat{K}_0(\mathbf{x}, t)$ denotes the transform of $K_0(\mathbf{X}, t)$, and $K_0'(\mathbf{x}, t)$ is the Hamiltonian of the inverse flow $\partial\mathbf{x}/\partial t$. Equation (10.64) shows that K_0' and $\hat{K}_0$ are related by

$$K_0'(\mathbf{x}, t) = -\frac{1}{a}\hat{K}_0(\mathbf{x}, t). \tag{10.65}$$

Hence in the special case of the canonical transformations ($a = 1$) we have

$$K_0'(\mathbf{x}, t) = -\hat{K}_0(\mathbf{x}, t). \tag{10.66}$$

This fact can easily be interpreted as follows. To produce a motion that is retrograde with respect to the flow $\partial\mathbf{X}/\partial t = \mathfrak{I}\nabla_{\mathbf{X}}K_0(\mathbf{X}, t)$ there are two possibilities:

(a) reverse the orientation of time ($t \to -t$), keeping the Hamiltonian fixed;
(b) keep the time orientation, but change K_0 into $-K_0$.

The condition (10.66) expresses the second possibility.

Example 10.13
The transformation

$$P = \alpha p\cos\omega t + \beta q\sin\omega t, \quad Q = -\frac{a}{\beta}p\sin\omega t + \frac{a}{\alpha}q\cos\omega t, \tag{10.67}$$

with α, β, ω, a non-zero constants, preserves the canonical structure of the Hamilton equations (check that $\det J = a$). It is canonical if and only if $a = 1$. In this case, it is the composition of a rotation with a 'natural' change of scale. The inverse of (10.67) is given by

$$p = \frac{1}{\alpha}P\cos\omega t - \frac{\beta}{a}Q\sin\omega t, \quad q = \frac{\alpha}{a}Q\cos\omega t + \frac{1}{\beta}P\sin\omega t. \tag{10.68}$$

By differentiating (10.67) with respect to time, and inserting (10.68) we find the equations for the Hamiltonian flow $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$:

$$\frac{\partial P}{\partial t} = \frac{\alpha\beta\omega}{a}Q, \quad \frac{\partial Q}{\partial t} = -\frac{a\omega}{\alpha\beta}P, \tag{10.69}$$

with which we associate the Hamiltonian

$$K_0 = -\frac{1}{2}\frac{\alpha\beta\omega}{a}Q^2 - \frac{1}{2}\frac{a\omega}{\alpha\beta}P^2. \tag{10.70}$$

Performing the corresponding manipulations for the inverse transformation (10.68) we find the equations for the retrograde flow:

$$\frac{\partial p}{\partial t} = -\frac{\beta\omega}{\alpha}q, \quad \frac{\partial q}{\partial t} = \frac{\alpha\omega}{\beta}p, \tag{10.71}$$

which is derived from the Hamiltonian

$$K_0' = \frac{1}{2}\frac{\beta\omega}{\alpha}q^2 + \frac{1}{2}\frac{\alpha\omega}{\beta}p^2. \tag{10.72}$$

Expressing K_0 in the variables (p, q) we obtain

$$\hat{K}_0 = -aK_0', \tag{10.73}$$

which is in agreement with equation (10.65). ■

10.3 The Poincaré–Cartan integral invariant. The Lie condition

In this section we want to focus on the geometric interpretation of canonical transformations. In the process of doing this, we derive a necessary and sufficient condition for a transformation to be canonical. This condition is very useful in practice, as we shall see in the next section.

Let us start by recalling a few definitions and results concerning differential forms.

DEFINITION 10.9 *A differential form ω in $\mathbf{R}^{2l+1}$*

$$\omega = \sum_{i=1}^{2l+1} \omega_i(\mathbf{x})\ \mathrm{d}x_i, \tag{10.74}$$

is non-singular *if the $(2l+1)\times(2l+1)$ skew-symmetric matrix $A(\mathbf{x})$, defined by*

$$A_{ij} = \frac{\partial \omega_i}{\partial x_j} - \frac{\partial \omega_j}{\partial x_i}, \tag{10.75}$$

has maximal rank $2l$. The kernel of $A(\mathbf{x})$, characterised by $\{\mathbf{v}\in\mathbf{R}^{2l+1} | A(\mathbf{x})\mathbf{v} = \mathbf{0}\}$, as $\mathbf{x}$ varies determines a field of directions in $\mathbf{R}^{2l+1}$ called characteristic directions. *The integral curves of the field of characteristic directions are called characteristics of ω.* ■

Remark 10.14

For $l = 1$, setting $\omega = (\omega_1, \omega_2, \omega_3)$ the matrix $A(\mathbf{x})$ is simply

$$A(\mathbf{x}) = \begin{pmatrix} 0 & -(\omega)_3 & (\omega)_2 \\ (\omega)_3 & 0 & -(\omega)_1 \\ -(\omega)_2 & (\omega)_1 & 0 \end{pmatrix}$$

and $A(\mathbf{x})\mathbf{v} = \boldsymbol{\omega}(\mathbf{x}) \times \mathbf{v}$. Therefore the characteristics of the form ω can be indentified with those of the field $\boldsymbol{\omega}$. ■

Example 10.14

The form $\omega = x_2\ \mathrm{d}x_1 + x_3\ \mathrm{d}x_2 + x_1\ \mathrm{d}x_3$ in $\mathbf{R}^3$ is non-singular. The associated characteristic direction is constant and is determined by the line $x_1 = x_2 = x_3$. ■

Example 10.15

The form $\omega = x_1\ \mathrm{d}x_2 + \frac{1}{2}(x_1^2 + x_2^2)\ \mathrm{d}x_3$ is non-singular. The associated field of characteristic directions is $(x_2, -x_1, 1)$. ■

Remark 10.15

The reader familiar with the notion of a differential 2-form (see Appendix 4) will recognise in the definition of the matrix A the representative matrix of the 2-form

$$-\mathrm{d}\omega = \sum_{i,j=1}^{2l+1} \frac{\partial \omega_i}{\partial x_j}\mathrm{d}x_i \wedge \mathrm{d}x_j. \quad ■$$

The following result can be easily deduced from Definition 10.9.

PROPOSITION 10.3 *Two non-singular forms differing by an exact form have the same characteristics.* ■

Consider any regular closed curve γ. The characteristics of ω passing through the points of γ define a surface in $\mathbf{R}^{2l+1}$ (i.e. a regular submanifold of dimension 2) called the *tube of characteristics.* The significance of non-singular differential forms, and of the associated tubes of characteristics, is due to the following property.

THEOREM 10.8 (Stokes' lemma) *Let ω be a non-singular differential form, and let γ_1 and γ_2 be any two homotopic closed curves belonging to the same tube of characteristics. Then*

$$\oint_{\gamma_1} \omega = \oint_{\gamma_2} \omega. \tag{10.76}$$

■

Equation (10.76) expresses the invariance of the *circulation* of the field $\mathbf{X}(\mathbf{x})$, whose components are the ω_i, along the closed lines traced on a tube of characteristics.

The previous theorem is a consequence of Stokes' lemma, discussed in Appendix 4. Note that this is natural generalisation of the Stokes formula, well known from basic calculus (see Giusti 1989).

We now consider a system with Hamiltonian $H(\mathbf{p}, \mathbf{q}, t)$ and its 'extended' phase space, where together with the canonical coordinates we consider the time $t : (\mathbf{p}, \mathbf{q}, t) \in \mathbf{R}^{2l+1}$.

THEOREM 10.9 *The differential form*

$$\omega = \sum_{i=1}^{l} p_i \, \mathrm{d}q_i - H(\mathbf{p}, \mathbf{q}, t) \, \mathrm{d}t \tag{10.77}$$

in $\mathbf{R}^{2l+1}$ is non-singular and it is called the Poincaré–Cartan form. *Its characteristics are the integral curves of the system of Hamilton's equations associated with the Hamiltonian H.*

Proof
The matrix associated with the form ω is

$$A(\mathbf{p}, \mathbf{q}, t) = \begin{pmatrix} \mathbf{0} & -\mathbf{1} & \nabla_{\mathbf{p}} H \\ \mathbf{1} & \mathbf{0} & \nabla_{\mathbf{q}} H \\ -(\nabla_{\mathbf{p}} H)^T & -(\nabla_{\mathbf{q}} H)^T & 0 \end{pmatrix}.$$

Evidently the rank of the matrix A is equal to $2l$ for every $(\mathbf{p}, \mathbf{q}, t)$ (note that one of its $2l \times 2l$ submatrices coincides with the matrix $\mathfrak{I}$). It follows that the form ω is non-singular. Moreover, the vector

$$\mathbf{v}(\mathbf{p}, \mathbf{q}, t) = (-\nabla_{\mathbf{q}} H, \nabla_{\mathbf{p}} H, 1)$$

is in the kernel of A for every $(\mathbf{p}, \mathbf{q}, t)$, and therefore it determines the characteristics of ω. The integral curves of $\mathbf{v}$ are the solutions of

$$\dot{\mathbf{p}} = -\nabla_{\mathbf{q}} H,$$
$$\dot{\mathbf{q}} = \nabla_{\mathbf{p}} H,$$
$$\dot{t} = 1,$$

and hence they are precisely the integral curves of Hamilton's system of equations for H, expressed in the extended phase space $\mathbf{R}^{2l+1}$. ■

The application of Stokes' lemma to the Poincaré–Cartan form (10.77) has a very important consequence.

THEOREM 10.10 (Integral invariant of Poincaré–Cartan) *Let γ_1 and γ_2 be any two homotopic closed curves in $\mathbf{R}^{2l+1}$ belonging to the same tube of characteristics relative to the form (10.77). Then*

$$\oint_{\gamma_1} \left(\sum_{i=1}^{l} p_i \, \mathrm{d}q_i - H(\mathbf{p}, \mathbf{q}, t) \, \mathrm{d}t \right) = \oint_{\gamma_2} \left(\sum_{i=1}^{l} p_i \, \mathrm{d}q_i - H(\mathbf{p}, \mathbf{q}, t) \, \mathrm{d}t \right). \tag{10.78}$$

■

Remark 10.16
Denote by γ_0 a closed curve belonging to the same tube of characteristics as γ, lying in the plane $t = t_0$, for fixed t_0. Then the result of Theorem 10.10 yields as a consequence the fact that

$$\oint_{\gamma} \left(\sum_{i=1}^{l} p_i \, \mathrm{d}q_i - H(\mathbf{p}, \mathbf{q}, t) \, \mathrm{d}t \right) = \oint_{\gamma_0} \sum_{i=1}^{l} p_i \, \mathrm{d}q_i. \tag{10.79}$$

We shall see how the integral (10.79) completely characterises the canonical transformations, highlighting the relation with the geometry of the Hamiltonian flow (i.e. of the tubes of characteristics of the Poincaré–Cartan form). Indeed, starting from a system of Hamilton's equations for a Hamiltonian H and going to a new system of Hamilton's equations for a new Hamiltonian K, the canonical transformations map the tubes of characteristics of the Poincaré–Cartan form (10.77) associated with H onto the tubes of characteristics of the corresponding form associated with K. ■

We can state the following corollary to Theorem 10.12.

COROLLARY 10.4 *A canonical transformation maps the tubes of characteristics of the Poincaré–Cartan form (10.80) into the tubes of characteristics of the corresponding form*

$$\Omega = \sum_{i=1}^{l} P_i \mathrm{d}Q_i - K(P, Q, t) \, \mathrm{d}t. \tag{10.80}$$

■

Example 10.16
Consider the transformation of Example 10.12, which we rewrite as

$$p = \alpha\sqrt{P} \cos \gamma Q, \quad q = \beta\sqrt{P} \sin \gamma Q.$$

For $\alpha\beta\gamma = 2$ this transformation is completely canonical. We compare the Poincaré–Cartan forms written in the two coordinate systems:

$$\omega = p \, \mathrm{d}q - H(p, q, t) \, \mathrm{d}t, \quad \Omega = P \, \mathrm{d}Q - \hat{H}(P, Q, t) \, \mathrm{d}t.$$

The difference is

$$\omega - \Omega = p \, \mathrm{d}q - P \, \mathrm{d}Q.$$

Expressing it in the variables P, Q we obtain

$$\omega - \Omega = \mathrm{d}\left(\frac{1}{2\gamma} P \sin 2\gamma Q \right).$$

Since ω and Ω differ by an exact differential, they have the same tubes of characteristics. ■

We now want to show that the result discussed in the previous example ($\omega - \Omega = \mathrm{d}f$) is entirely general and constitutes a necessary and sufficient condition for a transformation to be canonical. We start by analysing the difference $\omega - \Omega$ when we 'fix time' (freezing the variable t).

Consider a differentiable, invertible transformation $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ from the coordinates $\mathbf{x} = (\mathbf{p}, \mathbf{q})$ to $\mathbf{X} = (\mathbf{P}, \mathbf{Q})$:

$$p_i = p_i(\mathbf{P}, \mathbf{Q}, t), \quad q_i = q_i(\mathbf{P}, \mathbf{Q}, t), \tag{10.81}$$

where $i = 1, \ldots, l$. Consider the differential form

$$\tilde{\omega} = \sum_{i=1}^{l} p_i(\mathbf{P}, \mathbf{Q}, t)\ \tilde{\mathrm{d}} q_i(\mathbf{P}, \mathbf{Q}, t), \tag{10.82}$$

where, given any regular function $f(\mathbf{P}, \mathbf{Q}, t)$, we set

$$\tilde{\mathrm{d}} f = \mathrm{d}f - \frac{\partial f}{\partial t}\ \mathrm{d}t = \sum_{i=1}^{l} \left(\frac{\partial f}{\partial P_i}\ \mathrm{d}P_i + \frac{\partial f}{\partial Q_i}\ \mathrm{d}Q_i \right). \tag{10.83}$$

Here $\tilde{\mathrm{d}}$ is the so-called 'virtual differential' or 'time frozen differential' (see Levi-Civita and Amaldi 1927).

THEOREM 10.11 (Lie condition) *The transformation (10.81) is canonical if and only if the difference between the differential forms $\tilde{\omega}$ and $\tilde{\Omega}$ is exact, and hence if there exists a regular function $f(\mathbf{P}, \mathbf{Q}, t)$ such that*

$$\tilde{\omega} - \tilde{\Omega} = \sum_{i=1}^{l} (p_i\ \tilde{\mathrm{d}} q_i - P_i\ \tilde{\mathrm{d}} Q_i) = \tilde{\mathrm{d}} f. \tag{10.84}$$

Proof

Consider the difference

$$\tilde{\vartheta} = \tilde{\omega} - \tilde{\Omega}$$

and write it as

$$\begin{aligned} 2\tilde{\vartheta} &= \sum_{i=1}^{l} \left(p_i\ \tilde{\mathrm{d}} q_i - q_i\ \tilde{\mathrm{d}} p_i\right) - \sum_{i=1}^{l} \left(P_i\ \tilde{\mathrm{d}} Q_i - Q_i\ \tilde{\mathrm{d}} P_i\right) + \tilde{\mathrm{d}} \sum_{i=1}^{l} (p_i q_i - P_i Q_i) \\ &= \tilde{\eta} + \tilde{\mathrm{d}} \sum_{i=1}^{l} (p_i q_i - P_i Q_i). \end{aligned}$$

The form $\tilde{\eta}$ can be rewritten as

$$\tilde{\eta} = \mathbf{X}^T \mathfrak{I}\ \tilde{\mathrm{d}} \mathbf{X} - \mathbf{x}^T \mathfrak{I}\ \tilde{\mathrm{d}} \mathbf{x}.$$

Recalling that $\tilde{\mathrm{d}} \mathbf{X} = J\ \tilde{\mathrm{d}} \mathbf{x}$, we see that

$$\tilde{\eta} = \left(\mathbf{X}^T \mathfrak{I} J - \mathbf{x}^T \mathfrak{I}\right) \tilde{\mathrm{d}} \mathbf{x} = \mathbf{g}^T\ \tilde{\mathrm{d}} \mathbf{x},$$

with $\mathbf{g} = -J^T \mathfrak{I} \mathbf{X} + \mathfrak{I} \mathbf{x}$. Therefore, the form $\tilde{\eta}$ is exact if and only if $\partial g_i / \partial x_j = \partial g_j / \partial x_i$. We now compute (using the convention of summation over repeated indices)

$$\begin{aligned} \frac{\partial g_i}{\partial x_j} &= \mathfrak{I}_{ij} - \frac{\partial J_{ki}}{\partial x_j} \mathfrak{I}_{kh} X_h - J_{ki} \mathfrak{I}_{kh} J_{hj}, \\ \frac{\partial g_j}{\partial x_i} &= \mathfrak{I}_{ji} - \frac{\partial J_{kj}}{\partial x_i} \mathfrak{I}_{kh} X_h - J_{kj} \mathfrak{I}_{kh} J_{hi}, \end{aligned}$$

and note that

$$\frac{\partial J_{ki}}{\partial x_j} = \frac{\partial^2 X_k}{\partial x_i \partial x_j} = \frac{\partial J_{kj}}{\partial x_i},$$

and hence

$$\frac{\partial g_i}{\partial x_j} - \frac{\partial g_j}{\partial x_i} = (\mathfrak{I} - J^T \mathfrak{I} J)_{ij} - (\mathfrak{I} - J^T \mathfrak{I} J)_{ji} = 2(\mathfrak{I} - J^T \mathfrak{I} J)_{ij},$$

where $\mathfrak{I} - J^T \mathfrak{I} J$ is skew-symmetric. We can conclude that the form $\tilde{\eta}$, and therefore $\tilde{\omega} - \tilde{\Omega}$, is exact if and only if J is symplectic, or equivalently if and only if the transformation is canonical. ■

Remark 10.17

If the transformation is completely canonical, it is immediate to check that in the expression (10.84) $\tilde{\mathrm{d}} = \mathrm{d}$, and f can be chosen to be independent of t. ■

Example 10.17

Using the Lie condition it is easy to prove that point transformations (Example 10.9) are canonical. It follows from (10.35), (10.36) that

$$\begin{aligned} \sum_{i=1}^{l} (p_i\ \tilde{\mathrm{d}} q_i - P_i\ \tilde{\mathrm{d}} Q_i) &= \sum_{i=1}^{l} p_i\ \tilde{\mathrm{d}} q_i - \sum_{i,j,k=1}^{l} J_{ji}^{-1} p_j J_{ik}\ \tilde{\mathrm{d}} q_k \\ &= \sum_{i=1}^{l} p_i\ \tilde{\mathrm{d}} q_i - \sum_{j,k=1}^{l} p_j \delta_{jk}\ \tilde{\mathrm{d}} q_k = 0. \end{aligned}$$ ■

Example 10.18

Using the Lie condition let us check that the transformation (see Gallavotti 1986)

$$q_1 = \frac{P_1 P_2 - Q_1 Q_2}{P_1^2 + Q_2^2}, \qquad q_2 = \frac{P_2 Q_2 + P_1 Q_1}{P_1^2 + Q_2^2},$$

$$p_1 = -P_1 Q_2, \qquad p_2 = \frac{P_1^2 - Q_2^2}{2}$$

is completely canonical. Setting

$$\mathcal{P} = p_1 + \mathrm{i} p_2, \qquad \mathcal{Q} = q_1 + \mathrm{i} q_2,$$

where $\mathrm{i}=\sqrt{-1}$, note that

$$\mathcal{P}=\frac{\mathrm{i}}{2}(P_1+\mathrm{i}Q_2)^2, \qquad \mathcal{Q}=\frac{P_2+\mathrm{i}Q_1}{P_1-\mathrm{i}Q_2},$$

from which it follows that

$$p_1\ \mathrm{d}q_1+p_2\ \mathrm{d}q_2=\mathrm{Re}(\mathcal{P}\ \mathrm{d}\bar{\mathcal{Q}})=P_1\ \mathrm{d}Q_1+P_2\ \mathrm{d}Q_2-\frac{1}{2}\mathrm{d}(P_1Q_1+P_2Q_2);$$

hence the Lie condition is satisfied with $f=-\frac{1}{2}(P_1Q_1+P_2Q_2)$. ■

Remark 10.18
We can see that the Lie condition (10.84) is equivalent to the statement that there exists a regular function $f(\mathbf{P},\mathbf{Q},t)$, defined up to an arbitrary function of time, such that, for every $i=1,\ldots,l$,

$$\begin{aligned}\frac{\partial f}{\partial P_i}(\mathbf{P},\mathbf{Q},t)&=\sum_{j=1}^{l}p_j(\mathbf{P},\mathbf{Q},t)\frac{\partial q_j}{\partial P_i}(\mathbf{P},\mathbf{Q},t),\\ \frac{\partial f}{\partial Q_i}(\mathbf{P},\mathbf{Q},t)&=\sum_{j=1}^{l}p_j(\mathbf{P},\mathbf{Q},t)\frac{\partial q_j}{\partial Q_i}(\mathbf{P},\mathbf{Q},t)-P_i.\end{aligned} \tag{10.85}$$

■

The Lie condition has as a corollary an interesting result that characterises the canonical transformations through the Poincaré–Cartan integral invariant.

COROLLARY 10.5 *The transformation (10.81) is canonical if and only if, for every closed curve γ_0 in $\mathbf{R}^{2l+1}$ made of simultaneous states $(\mathbf{p},\mathbf{q},t_0)$, if Γ_0 is its image under the given transformation (in turn made of simultaneous states $(\mathbf{P},\mathbf{Q},t_0)$), then*

$$\oint_{\gamma_0}\sum_{i=1}^{l}p_i\ \mathrm{d}q_i=\oint_{\Gamma_0}\sum_{i=1}^{l}P_i\ \mathrm{d}Q_i. \tag{10.86}$$

Proof
From the definition of a fixed time differential, it follows that

$$\oint_{\gamma_0}\sum_{i=1}^{l}p_i\ \mathrm{d}q_i=\oint_{\Gamma_0}\tilde{\omega}, \qquad \oint_{\Gamma_0}\sum_{i=1}^{l}P_i\ \mathrm{d}Q_i=\oint_{\Gamma_0}\tilde{\Omega},$$

where $\tilde{\omega}$ and $\tilde{\Omega}$ are computed fixing $t=t_0$. Note that on Γ_0 we assume that $\tilde{\omega}$ is expressed in the new variables. Therefore the condition is necessary. Indeed, if the transformation is canonical, by the Lie condition the difference $\tilde{\omega}-\tilde{\Omega}$ is an exact form, whose integral along any closed path vanishes.

Evidently the condition is also sufficient. Indeed, if

$$\oint_{\Gamma_0}(\tilde{\omega}-\tilde{\Omega})=0$$

along any closed path Γ_0 then the form $\tilde{\omega}-\tilde{\Omega}$ is exact (see Giusti 1989, Corollary 8.2.1). ■

For $l=1$ equation (10.86) is simply the area conservation property, which we already know (in the form $\det J=1$) to be the characteristic condition for a transformation to be canonical.

We can now prove the important result, stated previously: the conservation of the Poincaré–Cartan integral invariant is exclusively a property of canonical transformations.

THEOREM 10.12 *If the transformation (10.81) is canonical, denote by*

$$\Omega=\sum_{i=1}^{l}P_i\ \mathrm{d}Q_i-K(\mathbf{P},\mathbf{Q},t)\ \mathrm{d}t \tag{10.87}$$

the new Poincaré–Cartan form. Then there exists a regular function $\mathcal{F}(\mathbf{P},\mathbf{Q},t)$ such that

$$\sum_{i=1}^{l}(p_i\ \mathrm{d}q_i-P_i\ \mathrm{d}Q_i)+(K-H)\ \mathrm{d}t=\omega-\Omega=\mathrm{d}\mathcal{F}. \tag{10.88}$$

Hence the difference between the two Poincaré–Cartan forms is exact. Conversely, if (10.81) is a coordinate transformation such that there exist two functions $K(\mathbf{P},\mathbf{Q},t)$ and $\mathcal{F}(\mathbf{P},\mathbf{Q},t)$ which, for Ω defined as in (10.87), satisfy (10.88), then the transformation is canonical and K is the new Hamiltonian.

Proof
We prove that if the transformation is canonical, then condition (10.88) is satisfied. Consider any regular closed curve γ in $\mathbf{R}^{2l+1}$, and let Γ be its image under the canonical transformation (10.81).

Since the transformation is canonical the tube of characteristics of ω through γ is mapped to the tube of characteristics of Ω through Γ (Corollary 10.74). Therefore it is possible to apply Stokes' lemma to write

$$\oint_{\Gamma}(\omega-\Omega)=\oint_{\Gamma_0}(\omega-\Omega)=\oint_{\gamma_0}\sum_{i=1}^{l}p_i\ \mathrm{d}q_i-\oint_{\Gamma_0}\sum_{i=1}^{l}P_i\ \mathrm{d}Q_i=0,$$

where γ_0,Γ_0 are the intersections of the respective tubes of characteristics with $t=t_0$ (Fig. 10.2). It follows that the integral of $\omega-\Omega$ along any closed path in $\mathbf{R}^{2l+1}$ is zero, and therefore the form is exact.

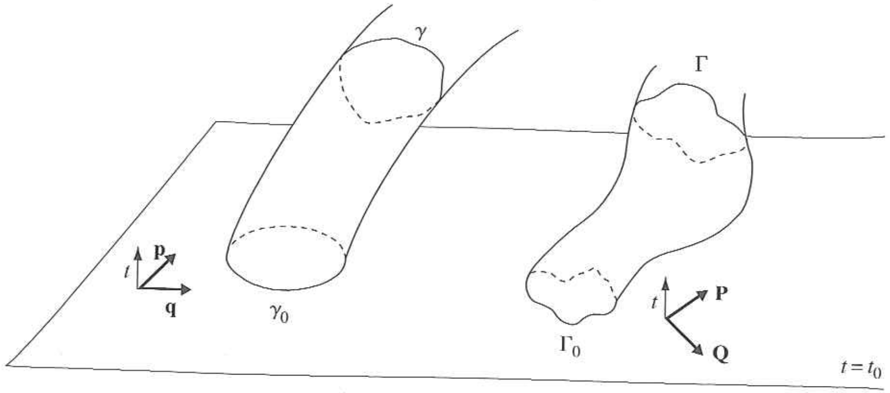

Fig. 10.2

We now prove the second part of the theorem. Since the difference $\omega - \Omega$ is exact we have

$$\oint_{\gamma_0} \sum_{i=1}^{l} p_i \, \mathrm{d}q_i - \oint_{\Gamma_0} \sum_{i=1}^{l} P_i \, \mathrm{d}Q_i = \oint_{\Gamma_0} (\omega - \Omega) = 0,$$

and the transformation is canonical. Therefore the characteristic directions of the form ω coincide, after the transformation, with those of the form $\Omega' = \sum_{i=1}^{l} P_i \, \mathrm{d}Q_i - K' \, \mathrm{d}t$, where K' is the new Hamiltonian. On the other hand, the characteristic directions of ω coincide with those of $\Omega + \mathrm{d}\mathcal{F}$, and hence of Ω. In addition $\Omega' - \Omega = (K' - K) \, \mathrm{d}t$ and the coincidence of characteristics implies that $K' - K$ may depend only on t. Hence, following our convention, $K' = K$. ∎

Example 10.19

We consider again Example 10.11 in the light of the results of this section. By equation (10.49), the Lie condition (10.84) can be written as

$$p \, \tilde{\mathrm{d}}q - P \, \tilde{\mathrm{d}}Q = (P - at)(\mathrm{d}Q + t \, \mathrm{d}P) - P \, \mathrm{d}Q = \tilde{\mathrm{d}} f(P, Q, t),$$

from which it follows that

$$f(P, Q, t) = t\frac{P^2}{2} - at^2 P - atQ + f_1(t),$$

where f_1 is an arbitrary function of time.

The condition (10.88) for the transformation (10.49), taking into account (10.50), can be written as

$$(P - at)(\mathrm{d}Q + P \, \mathrm{d}t + t \, \mathrm{d}P - at \, \mathrm{d}t) - P \, \mathrm{d}Q + \left(-\frac{P^2}{2} - aQ\right) \mathrm{d}t = \mathrm{d}\mathcal{F}(P, Q, t),$$

and after some simple manipulations we find

$$\mathcal{F}(P, Q, t) = \frac{1}{2}tP^2 - at^2 P - atQ + \frac{1}{3}a^2 t^3.$$ ∎

We conclude this section by proving that the Hamiltonian flow defines a canonical transformation.

Let $H(\mathbf{p}, \mathbf{q}, t)$ be a Hamiltonian function, and consider the associated Hamiltonian flow $\mathbf{x} = S^t \mathbf{X}$:

$$p_i = p_i(\mathbf{P}, \mathbf{Q}, t), \quad q_i = q_i(\mathbf{P}, \mathbf{Q}, t), \tag{10.89}$$

where $i = 1, \ldots, l$. Equations (10.89) are therefore the solutions of the system of equations

$$\frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial q_i}, \quad \frac{\partial q_i}{\partial t} = \frac{\partial H}{\partial p_i}, \tag{10.90}$$

with initial conditions $p_i(0) = P_i$, $q_i(0) = Q_i$, $i = 1, \ldots, l$. By the theorem of existence, uniqueness and continuous dependence on the initial data for ordinary differential equations (see Appendix 1) equation (10.89) defines a coordinate transformation which is regular and invertible.

THEOREM 10.13 *The Hamiltonian flow (10.89) is a time-dependent canonical transformation, that at every time instant t maps* $\mathbf{X}$ *to* $S^t\mathbf{X}$. *In addition, the new Hamiltonian associated with H in the variables* $\mathbf{X}$ *is* $K \equiv 0$.

Proof

We verify that the Lie condition (10.84) is satisfied, with

$$f(\mathbf{P}, \mathbf{Q}, t) = \int_0^t \left[\sum_{j=1}^{l} p_j(\mathbf{P}, \mathbf{Q}, \tau) \frac{\partial q_j}{\partial t}(\mathbf{P}, \mathbf{Q}, \tau) - H(\mathbf{p}(\mathbf{P}, \mathbf{Q}, \tau), \mathbf{q}(\mathbf{P}, \mathbf{Q}, \tau), \tau) \right] \mathrm{d}\tau. \tag{10.91}$$

By Remark 10.18, it is enough to show that for every $i = 1, \ldots, l$ we have

$$\frac{\partial f}{\partial P_i}(\mathbf{P}, \mathbf{Q}, t) = \sum_{j=1}^{l} p_j(\mathbf{P}, \mathbf{Q}, t) \frac{\partial q_j}{\partial P_i}(\mathbf{P}, \mathbf{Q}, t),$$

$$\frac{\partial f}{\partial Q_i}(\mathbf{P}, \mathbf{Q}, t) = \sum_{j=1}^{l} p_j(\mathbf{P}, \mathbf{Q}, t) \frac{\partial q_j}{\partial Q_i}(\mathbf{P}, \mathbf{Q}, t) - P_i.$$

We prove the second relation. The first one can be shown in an analogous manner. We have

$$\frac{\partial f}{\partial Q_i} = \int_0^t \sum_{j=1}^{l} \left[\frac{\partial p_j}{\partial Q_i} \frac{\partial q_j}{\partial t} + p_j \frac{\partial^2 q_j}{\partial t \partial Q_i} - \frac{\partial H}{\partial p_j} \frac{\partial p_j}{\partial Q_i} - \frac{\partial H}{\partial q_j} \frac{\partial q_j}{\partial Q_i} \right] \mathrm{d}\tau,$$

but since (10.89) is the transformation generated by the Hamiltonian flow, it follows from equations (10.90) that

$$\begin{aligned}\frac{\partial f}{\partial Q_i} &= \int_0^t \sum_{j=1}^l \left[\frac{\partial p_j}{\partial Q_i}\frac{\partial q_j}{\partial t} + p_j \frac{\partial^2 q_j}{\partial t \partial Q_i} - \frac{\partial q_j}{\partial t}\frac{\partial p_j}{\partial Q_i} + \frac{\partial p_j}{\partial t}\frac{\partial q_j}{\partial Q_i}\right] \mathrm{d}\tau \\ &= \int_0^t \frac{\partial}{\partial t}\sum_{j=1}^l p_j \frac{\partial q_j}{\partial Q_i}\,\mathrm{d}\tau \\ &= \sum_{j=1}^l p_j(\mathbf{P},\mathbf{Q},t)\frac{\partial q_j}{\partial Q_i}(\mathbf{P},\mathbf{Q},t) - \sum_{j=1}^l p_j(\mathbf{P},\mathbf{Q},0)\frac{\partial q_j}{\partial Q_i}(\mathbf{P},\mathbf{Q},0) \\ &= \sum_{j=1}^l p_j(\mathbf{P},\mathbf{Q},t)\frac{\partial q_j}{\partial Q_i}(\mathbf{P},\mathbf{Q},t) - \sum_{j=1}^l P_j\delta_{ji} \\ &= \sum_{j=1}^l p_j(\mathbf{P},\mathbf{Q},t)\frac{\partial q_j}{\partial Q_i}(\mathbf{P},\mathbf{Q},t) - P_i.\end{aligned}$$

By what we have just computed,

$$\tilde{\mathrm{d}} f = \sum_{i=1}^l \left(p_i\, \tilde{\mathrm{d}} q_i - P_i\, \mathrm{d}Q_i\right),$$

while from (10.83) it obviously follows that

$$\begin{aligned}\mathrm{d}f &= \tilde{\mathrm{d}} f + \frac{\partial f}{\partial t}\,\mathrm{d}t = \sum_{i=1}^l \left(p_i\, \tilde{\mathrm{d}} q_i - P_i\, \tilde{\mathrm{d}} Q_i\right) + \left[\sum_{j=1}^l p_j \frac{\partial q_j}{\partial t} - H\right]\mathrm{d}t \\ &= \sum_{i=1}^l (p_i\, \mathrm{d}q_i - P_i\, \mathrm{d}Q_i) - H\,\mathrm{d}t.\end{aligned}$$

Taking into account Theorem 10.6, it follows from this that the new Hamiltonian associated with H is exactly $K \equiv 0$. ■

Remark 10.19

From the expression (10.91) for f, since $\dot{p}_i = \partial p_i/\partial t$ and $\dot{q}_i = \partial q_i/\partial t$, we see that $f(\mathbf{P},\mathbf{Q},t)$ is the Hamiltonian action $A(\mathbf{P},\mathbf{Q},t)$ (see (9.43)) computed by an integration along the Hamiltonian flow (10.89), i.e. the natural motion. ■

Recalling the result of Corollary 10.1, we can now state that *the canonical transformations depending on time are all and exclusively the Hamiltonian flows.* If we apply the canonical transformation $\mathbf{x} = \mathbf{x}(\mathbf{x}^*, t)$ generated by the Hamiltonian $H(\mathbf{x}, t)$, to a system with Hamiltonian $H^*(\mathbf{x}^*, t)$, we obtain the new Hamiltonian $K^*(\mathbf{x}, t) = \hat{H}^*(\mathbf{x}, t) + H(\mathbf{x}, t)$ (here H plays the role of the function indicated by K_0 in the previous section). Consider now the Hamiltonian flow $\mathbf{x} = S^t\mathbf{X}$, with Hamiltonian $H(\mathbf{x}, t)$. The inverse transformation, mapping $S^t\mathbf{X}$ in $\mathbf{X}$ for every t, corresponds to the retrograde motion (with Hamiltonian $-H$) and it is naturally canonical. For the canonical transformation $\mathbf{x} = S^t\mathbf{X}$ the variables $\mathbf{X}$ play the role of *constant canonical coordinates* ($\dot{\mathbf{X}} = 0$). In agreement with this fact, we note that the composition of the two flows yields the Hamiltonian $K(\mathbf{X}, t) = 0$ and therefore precisely constant canonical coordinates. As an example, note that the transformation (10.49) is the flow with Hamiltonian $H = p^2/2 + aq$. This is independent of time, and hence it is a constant of the motion, implying that $p^2/2+aq = P^2/2+aQ$. This is the equation for the trajectories, travelled 'forwards' $(P, Q) \to (p, q)$ through the flow with Hamiltonian $H(p, q)$, and 'backwards' $(p, q) \to (P, Q)$ with Hamiltonian (10.50), i.e. $-H(P, Q)$. The superposition of the two yields $(P, Q) \to (P, Q)$ for every t, and hence $\dot{P} = \dot{Q} = 0$ (corresponding to the null Hamiltonian).

Remark 10.20

The apparent lack of symmetry between the condition

$$\sum_{i=1}^l (p_i\, \mathrm{d}q_i - P_i\, \mathrm{d}Q_i) = \mathrm{d}\mathcal{F},$$

where $\mathcal{F}$ is independent of t, for a transformation to be *completely* canonical, and the relation

$$\sum_{i=1}^l (p_i\, \mathrm{d}q_i - P_i\, \mathrm{d}Q_i) + (K - H)\,\mathrm{d}t = \mathrm{d}\mathcal{F},$$

where $\mathcal{F}$ depends also on t, for a time-dependent transformation to be canonical, can be eliminated by using a significant extension of the Hamiltonian formalism.

Indeed, given a non-autonomous Hamiltonian system $H(\mathbf{p}, \mathbf{q}, t)$, we consider, in addition to the canonical equations (10.90), the equations (see (8.26))

$$-\dot{H} = -\frac{\mathrm{d}H}{\mathrm{d}t} = -\frac{\partial H}{\partial t}, \quad \dot{t} = 1. \tag{10.92}$$

The system of equations (10.90), (10.92) corresponds to the canonical equations for the Hamiltonian $\mathcal{H} : \mathbf{R}^{2l+2} \to \mathbf{R}$,

$$\mathcal{H}(\mathbf{p}, \pi, \mathbf{q}, \tau) = H(\mathbf{p}, \mathbf{q}, \tau) + \pi, \tag{10.93}$$

where

$$\pi = -H, \quad \tau = t, \tag{10.94}$$

and hence the Hamiltonian and time are considered as a new pair of canonically conjugate variables. This is possible since $\nabla_{\mathbf{p}}\mathcal{H} = \nabla_{\mathbf{p}}H$, $\nabla_{\mathbf{q}}\mathcal{H} = \nabla_{\mathbf{q}}H$ and

$$\dot{\pi} = -\frac{\partial \mathcal{H}}{\partial \tau} = -\frac{\partial H}{\partial t}, \quad \dot{\tau} = \frac{\partial \mathcal{H}}{\partial \pi} = 1.$$

By (10.94) we also have that $\mathcal{H} = 0$, and the Poincaré–Cartan form (10.77) becomes

$$\sum_{i=1}^{l} p_i \, dq_i - H \, dt = \sum_{i=1}^{l} p_i \, dq_i + \pi \, d\tau = \sum_{i=1}^{l+1} p_i \, dq_i, \tag{10.95}$$

where we set $p_{l+1} = \pi$, $q_{l+1} = \tau$.

The canonical transformations (10.81) are therefore always completely canonical in $\mathbf{R}^{2l+2}$, and they associate with the variables $(\mathbf{p}, \pi, \mathbf{q}, \tau)$ new variables $(\mathbf{P}, \Pi, \mathbf{Q}, T)$, with the constraint $T = \tau$. The Hamiltonian $\mathcal{H}$ is always zero.

Conversely, transformations such as

$$\tau = a(T), \quad \pi = \frac{1}{a'(T)} \Pi \tag{10.96}$$

can be included in the canonical formalism, since

$$\pi \, d\tau = \frac{1}{a'(T)} \Pi \, a'(T) \, dT = \Pi \, dT.$$

The effect of equation (10.96) is a re-parametrisation of time, and by using the fact that it is canonical one can show that the canonical structure of Hamilton's equations is preserved, by appropriately rescaling the Hamiltonian $H = -\pi$. ■

10.4 Generating functions

In the previous sections we completely described the class of canonical transformations. We now study a procedure to generate *all* canonical transformations.

As we saw in the previous section, the Lie condition (10.84), or its equivalent formulation (10.88), is a necessary and sufficient condition for a coordinate transformation to be canonical. In the form (10.88), it allows the introduction of an efficient way to construct other canonical transformations.

Assume that

$$\mathbf{p} = \mathbf{p}(\mathbf{P}, \mathbf{Q}, t), \quad \mathbf{q} = \mathbf{q}(\mathbf{P}, \mathbf{Q}, t) \tag{10.97}$$

defines a canonical transformation in an open domain of $\mathbf{R}^{2l}$, with inverse

$$\mathbf{P} = \mathbf{P}(\mathbf{p}, \mathbf{q}, t), \quad \mathbf{Q} = \mathbf{Q}(\mathbf{p}, \mathbf{q}, t). \tag{10.98}$$

A canonical transformation of the type (10.97) satisfying

$$\det \left(\frac{\partial q_i}{\partial P_j} \right) \neq 0 \tag{10.99}$$

is called *free*. Applying the implicit function theorem to the second of equations (10.97), the condition (10.99) ensures that the variables $\mathbf{P}$ can be naturally expressed as functions of the variables $\mathbf{q}$, $\mathbf{Q}$, as well as of time. Therefore, if

$$\mathbf{P} = \hat{\mathbf{P}}(\mathbf{q}, \mathbf{Q}, t), \tag{10.100}$$

by substituting this relation into the first of equations (10.97) we find

$$\mathbf{p} = \hat{\mathbf{p}}(\mathbf{q}, \mathbf{Q}, t). \tag{10.101}$$

The condition (10.88)

$$\sum_{i=1}^{l} p_i \, dq_i - H \, dt - \left(\sum_{i=1}^{l} P_i \, dQ_i - K \, dt \right) = d\mathcal{F}$$

can therefore be written

$$\sum_{i=1}^{l} \hat{p}_i(\mathbf{q}, \mathbf{Q}, t) \, dq_i - H(\mathbf{q}, \hat{\mathbf{p}}(\mathbf{q}, \mathbf{Q}, t), t) \, dt$$
$$- \left(\sum_{i=1}^{l} \hat{P}_i(\mathbf{q}, \mathbf{Q}, t) \, dQ_i - K(\hat{\mathbf{P}}(\mathbf{q}, \mathbf{Q}, t), \mathbf{Q}, t) \, dt \right) = dF(\mathbf{q}, \mathbf{Q}, t), \tag{10.102}$$

where the variables $(\mathbf{q}, \mathbf{Q})$ are considered to be independent and $F(\mathbf{q}, \mathbf{Q}, t)$ is obtained from $\mathcal{F}(\mathbf{P}, \mathbf{Q}, t)$ through equation (10.100). From (10.102) it follows that

$$p_i = \frac{\partial F}{\partial q_i}, \tag{10.103}$$

$$P_i = -\frac{\partial F}{\partial Q_i}, \tag{10.104}$$

$$K = H + \frac{\partial F}{\partial t}, \tag{10.105}$$

where $i = 1, \ldots, l$.

Equation (10.104) shows that the matrix $-(\partial q_i / \partial P_j)$ is the inverse matrix of $(\partial^2 F / (\partial q_i \partial Q_j))$. Therefore the condition (10.99) is clearly equivalent to requiring that

$$\det \left(\frac{\partial^2 F}{\partial q_i \partial Q_j} \right) \neq 0. \tag{10.106}$$

We now follow the converse path, starting from the choice of a function of the type (10.106).

Definition 10.10 *A function $F(\mathbf{q}, \mathbf{Q}, t)$ satisfying condition (10.106) is called a* generating function (of the first kind, *and it is often denoted by $F = F_1$) of the canonical transformation defined implicitly by equations (10.103)–(10.105).* ■

Remark 10.21

Given the generating function F, equations (10.103)–(10.105) define the canonical transformation implicitly. However the condition (10.106) ensures that the variables $\mathbf{Q}$ can be expressed as functions of $(\mathbf{q}, \mathbf{p})$ and of time t, by inverting equation (10.103). The expression of $\mathbf{P}$ as a function of $(\mathbf{q}, \mathbf{p})$ and of the time t can be obtained by substituting the relation $Q_i = Q_i(\mathbf{q}, \mathbf{p}, t)$ into equation (10.104). The invertibility of the transformation thus obtained is again guaranteed by the implicit function theorem. Indeed, equation (10.106) also ensures that it is possible to express $\mathbf{q} = \mathbf{q}(\mathbf{Q}, \mathbf{P}, t)$ by inverting (10.104). Substituting these into equation (10.103) we finally find $\mathbf{p} = \mathbf{p}(\mathbf{Q}, \mathbf{P}, t)$. ■

Example 10.20
The function $F(q, Q) = m\omega/2q^2 \cot Q$ generates a canonical transformation

$$p = \sqrt{2P\omega m} \cos Q, \qquad q = \sqrt{\frac{2P}{\omega m}} \sin Q,$$

which transforms the Hamiltonian of the harmonic oscillator

$$H(p, q) = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}$$

into

$$K(P, Q) = \omega P.$$

■

Example 10.21
The identity transformation $p = P$, $q = Q$ is not free. Hence it does not admit a generating function of the first kind. ■

After setting $\mathbf{x} = (\mathbf{p}, \mathbf{q})$ and $\mathbf{X} = (\mathbf{P}, \mathbf{Q})$, we see that a generating function can also depend on $x_{m_1}, \ldots, x_{m_l}$, $X_{n_1}, \ldots, X_{n_l}$ for an arbitrary choice of the indices m_i and n_i (all different). We quickly analyse all possible cases.

DEFINITION 10.11 *A function $F(\mathbf{q}, \mathbf{P}, t)$ satisfying the condition*

$$\det \left(\frac{\partial^2 F}{\partial q_i \partial P_j} \right) \neq 0 \tag{10.107}$$

is called a generating function of the second kind *(and it is often denoted by $F = F_2$) of the canonical transformation implicitly defined by*

$$p_i = \frac{\partial F}{\partial q_i}, \quad i = 1, \ldots, l, \tag{10.108}$$

$$Q_i = \frac{\partial F}{\partial P_i}, \quad i = 1, \ldots, l. \tag{10.109}$$

■

Example 10.22
Point transformations (see Example 10.9)

$$\mathbf{Q} = \mathbf{Q}(\mathbf{q}, t)$$

are generated by

$$F_2(\mathbf{q}, \mathbf{P}, t) = \sum_{i=1}^{l} P_i Q_i(\mathbf{q}, t).$$

Setting $\mathbf{Q} = \mathbf{q}$ we find that $F_2 = \sum_{i=1}^{l} P_i q_i$ is the generating function of the identity transformation. ■

DEFINITION 10.12 *A function $F(\mathbf{p}, \mathbf{Q}, t)$ which satisfies the condition*

$$\det \left(\frac{\partial^2 F}{\partial p_i \partial Q_j} \right) \neq 0 \tag{10.110}$$

is called a generating function of the third kind *(and it is often denoted by $F = F_3$) of the canonical transformation implicitly defined by*

$$q_i = -\frac{\partial F}{\partial p_i}, \quad i = 1, \ldots, l, \tag{10.111}$$

$$P_i = -\frac{\partial F}{\partial Q_i}, \quad i = 1, \ldots, l. \tag{10.112}$$

■

Example 10.23
It is immediate to check that the function $F(p, Q) = -p(e^Q - 1)$ generates the canonical transformation

$$P = p(1 + q), \qquad Q = \log(1 + q).$$

■

DEFINITION 10.13 *A function $F(\mathbf{p}, \mathbf{P}, t)$ which satisfies the condition*

$$\det \left(\frac{\partial^2 F}{\partial p_i \partial P_j} \right) \neq 0 \tag{10.113}$$

is called a generating function of the fourth kind *(and it is often denoted by $F = F_4$) of the canonical transformation implicitly defined by*

$$q_i = -\frac{\partial F}{\partial p_i}, \quad i = 1, \ldots, l, \tag{10.114}$$

$$Q_i = \frac{\partial F}{\partial P_i}, \quad i = 1, \ldots, l. \tag{10.115}$$

■

Example 10.24
The canonical transformation of Example 10.8, exchanging the coordinates and the kinetic momenta, admits as generating function $F(\mathbf{p}, \mathbf{P}) = \sum_{i=1}^{l} p_i P_i$. ■

THEOREM 10.14 *The generating functions of the four kinds F_1, F_2, F_3 and F_4 satisfy, respectively,*

$$\sum_{i=1}^{l} (p_i\, dq_i - P_i\, dQ_i) + (K - H)\, dt = dF_1(\mathbf{q}, \mathbf{Q}, t), \tag{10.116}$$

$$\sum_{i=1}^{l} (p_i\, dq_i + Q_i\, dP_i) + (K - H)\, dt = dF_2(\mathbf{q}, \mathbf{P}, t), \tag{10.117}$$

$$\sum_{i=1}^{l} (-q_i\, dp_i - P_i\, dQ_i) + (K - H)\, dt = dF_3(\mathbf{p}, \mathbf{Q}, t), \tag{10.118}$$

$$\sum_{i=1}^{l} (-q_i\, dp_i + Q_i\, dP_i) + (K - H)\, dt = dF_4(\mathbf{p}, \mathbf{P}, t). \tag{10.119}$$

If a canonical transformation admits more than one generating function of the previous kinds, then these are related by a Legendre transformation:

$$\begin{aligned} F_2 &= F_1 + \sum_{i=1}^{l} P_i Q_i, \\ F_3 &= F_1 - \sum_{i=1}^{l} p_i q_i, \\ F_4 &= F_1 - \sum_{i=1}^{l} p_i q_i + \sum_{i=1}^{l} P_i Q_i = F_2 - \sum_{i=1}^{l} p_i q_i = F_3 + \sum_{i=1}^{l} P_i Q_i. \end{aligned} \tag{10.120}$$

Proof

The first part of the theorem is a consequence of Definitions 10.10–10.13. The proof of the second part is immediate, and can be obtained by adding or subtracting $\sum_{i=1}^{l} P_i Q_i$ and $\sum_{i=1}^{l} p_i q_i$ from (10.116). ■

Remark 10.22

At this point it should be clear how, in principle, there exist $2\binom{2l}{l}$ different kinds of generating functions, each corresponding to a different arbitrary choice of l variables among $\mathbf{q}$, $\mathbf{p}$ and of l variables among $\mathbf{Q}$, $\mathbf{P}$. However, it is always possible to reduce it to one of the four previous kinds, by taking into account that the exchanges of Lagrangian coordinates and kinetic momenta are canonical transformations (see Example 10.8). ■

The transformations associated with generating functions exhaust all canonical transformations.

THEOREM 10.15 *It is possible to associate with every canonical transformation a generating function, and the transformation is completely canonical if and only if its generating function is time-independent. The generating function is of one of the four kinds listed above, up to possible exchanges of Lagrangian coordinates with kinetic moments.*

Proof

Consider a canonical transformation, and let $\mathcal{F}$ the function associated with it by Theorem 10.12. If it is possible to express the variables $\mathbf{p}$, $\mathbf{P}$ as functions of $\mathbf{q}$, $\mathbf{Q}$, and hence if (10.99) holds, then, as we saw at the beginnning of this section, it is enough to set

$$F_1(\mathbf{q}, \mathbf{Q}, t) = \mathcal{F}(\hat{\mathbf{P}}(\mathbf{q}, \mathbf{Q}, t), \mathbf{Q}, t)$$

and the conditions of Definition 10.10 are satisfied.

If, on the other hand, we have

$$\det\left(\frac{\partial q_i}{\partial Q_j}\right) \neq 0, \tag{10.121}$$

we can deduce $\mathbf{Q} = \hat{\mathbf{Q}}(\mathbf{q}, \mathbf{P}, t)$ from the second of equations (10.97) and, by substitution into the first of equations (10.97), we find that the variables $\mathbf{p}$ can also be expressed through $\mathbf{q}$, $\mathbf{P}$. Hence we set

$$F_2(\mathbf{q}, \mathbf{P}, t) = \mathcal{F}(\mathbf{P}, \hat{\mathbf{Q}}(\mathbf{q}, \mathbf{P}, t), t) + \sum_{i=1}^{l} P_i \hat{Q}_i(\mathbf{q}, \mathbf{P}, t).$$

The condition (10.107) is automatically satisfied, since $(\partial^2 F / \partial q_i \partial P_j)$ is the inverse matrix of $(\partial q_i / \partial Q_j)$.

Analogously, if

$$\det\left(\frac{\partial p_i}{\partial P_j}\right) \neq 0, \tag{10.122}$$

the variables $\mathbf{q}$, $\mathbf{P}$ can be expressed through $\mathbf{p}$, $\mathbf{Q}$, and we set

$$F_3(\mathbf{p}, \mathbf{Q}, t) = \mathcal{F}(\hat{\mathbf{P}}(\mathbf{p}, \mathbf{Q}, t), \mathbf{Q}, t) - \sum_{i=1}^{l} p_i \hat{q}_i(\mathbf{p}, \mathbf{Q}, t).$$

Then the conditions of Definition 10.12 are satisfied.

Finally, if

$$\det\left(\frac{\partial p_i}{\partial Q_j}\right) \neq 0, \tag{10.123}$$

by expressing $\mathbf{q}$, $\mathbf{Q}$ as functions of $\mathbf{p}$, $\mathbf{P}$, we find that the generating function is given by

$$F_4(\mathbf{p}, \mathbf{P}, t) = \mathcal{F}(\mathbf{P}, \hat{\mathbf{Q}}(\mathbf{p}, \mathbf{P}, t), t) - \sum_{i=1}^{l} p_i \hat{q}_i(\mathbf{p}, \mathbf{P}, t) + \sum_{i=1}^{l} P_i \hat{Q}_i(\mathbf{p}, \mathbf{P}, t).$$

It is always possible to choose l variables among $\mathbf{p}$, $\mathbf{q}$ and l variables among $\mathbf{P}$, $\mathbf{Q}$ as independent variables. As a matter of fact, the condition that the Jacobian matrix of the transformation is symplectic, and therefore non-singular, guarantees the existence an $l \times l$ submatrix with a non-vanishing determinant. If the selected independent variables are not in any of the four groups already considered, we can proceed in a similar way, and obtain a generating function of a different kind. On the other hand, it is always possible to reduce to one of the previous cases by a suitable exchange of variables. ■

Remark 10.23
An alternative proof of the previous theorem, that is maybe more direct and certainly more practical in terms of applications, can be obtained simply by remarking how conditions (10.99), (10.121)–(10.123) ensure that the Lie condition can be rewritten in the form (10.116)–(10.119), respectively. The functions $F_1, \dots, F_4$ can be determined by integration along an arbitrary path in the domain of definition and the invertibility of the transformation. ■

Example 10.25
Consider the canonical transformation

$$p = 2e^t \sqrt{PQ} \log P, \quad q = e^{-t}\sqrt{PQ},$$

defined in $D = \{(P,Q) \in \mathbf{R}^2 | P > 0, Q \geq 0\} \subset \mathbf{R}^2$. Evidently it is possible to choose (q, P) as independent variables and write

$$p = 2e^{2t} q \log P, \quad Q = \frac{e^{2t}q^2}{P}.$$

The generating function $F_2(q, P, t)$ can be found, for example, by integrating the differential form

$$\hat{p}(q,P,t)\ \mathrm{d}q + \hat{Q}(q,P,t)\ \mathrm{d}P$$

along the path $\gamma = \{(x,1) | 0 \leq x \leq q\} \cup \{(q,y) | 1 \leq y \leq P\}$ in the plane (q, P). Since along the first horizontal part of the path γ one has $p(x, 1, t) \equiv 0$ (this simplification motivates the choice of the integration path γ), we have

$$F_2(q,P,t) = e^{2t}q^2 \int_1^P \frac{\mathrm{d}y}{y} + \tilde{F}_2(t) = e^{2t}q^2 \log P + \tilde{F}_2(t),$$

where $\tilde{F}_2$ is an arbitrary function of time. ■

Remark 10.24
Every generating function F is defined up to an arbitrary additive term, a function only of time. This term does not change the transformation generated by F, but it modifies the Hamiltonian (because of (10.105)) and it arises from the corresponding indetermination of the difference between the Poincaré–Cartan forms associated with the transformation (see Remark 10.18). Similarly to what has already been seen, this undesired indetermination can be overcome by requiring that the function F does not contain terms that are only functions of t. ■

We conclude this section by proving a uniqueness result for the generating function (once the arbitrariness discussed in the previous remark is resolved).

PROPOSITION 10.4 *All the generating functions of a given canonical transformation, depending on the same group of independent variables, differ only by a constant.*

Proof
Consider as an example the case of two generating functions $F(\mathbf{q}, \mathbf{Q}, t)$ and $G(\mathbf{q}, \mathbf{Q}, t)$. The difference $F - G$ satisfies the conditions

$$\frac{\partial}{\partial q_i}(F-G) = 0, \quad \frac{\partial}{\partial Q_i}(F-G) = 0,$$

for every $i = 1, \dots, l$. Hence, since by Remark 10.24 we have neglected additive terms depending only on time, $F - G$ is necessarily constant. ■

10.5 Poisson brackets

Consider two funtions $f(\mathbf{x}, t)$ and $g(\mathbf{x}, t)$ defined in $\mathbf{R}^{2l} \times \mathbf{R}$ with sufficient regularity, and recall the definition (10.16) of a standard symplectic product.

DEFINITION 10.14 *The* Poisson bracket *of the two functions, denoted by $\{f, g\}$, is the function defined by the symplectic product of the gradients of the two functions:*

$$\{f,g\} = (\nabla_{\mathbf{x}} f)^T \mathfrak{J} \nabla_{\mathbf{x}} g. \tag{10.124}$$

■

Remark 10.25
If $\mathbf{x} = (\mathbf{p}, \mathbf{q})$, the Poisson bracket of two functions f and g is given by

$$\{f,g\} = \sum_{i=1}^{l} \left(\frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q_i}\frac{\partial f}{\partial p_i} \right). \tag{10.125}$$

■

Remark 10.26
Using the Poisson brackets, Hamilton's equations in the variables $(\mathbf{p}, \mathbf{q})$ can be written in a perfectly symmetric form as

$$\dot{p}_i = \{p_i, H\}, \qquad \dot{q}_i = \{q_i, H\}, \quad i = 1, \dots, l. \tag{10.126}$$

■

Remark 10.27
From equation (10.125) we derive the *fundamental Poisson brackets*

$$\{p_i, p_j\} = \{q_i, q_j\} = 0, \quad \{q_i, p_j\} = -\{p_i, q_j\} = \delta_{ij}. \tag{10.127}$$

■