THE HERGLOTZ LECTURES ON CONTACT TRANSFORMATIONS AND HAMILTONIAN SYSTEMS

LECTURE NOTES IN NONLINEAR ANALYSIS
VOL. 1

THE HERGLOTZ LECTURES ON CONTACT TRANSFORMATIONS AND HAMILTONIAN SYSTEMS

R. B. Guenther C. M. Guenther
J. A. Gottsch

Juliusz Schauder Center for Nonlinear Studies
Nicholas Copernicus University

ISBN 83-231-0736-X

Uniwersytet Mikołaja Kopernika
87-100 Toruń, ul. Gagarina 11, tel. 14295, fax. 21224

Wyd. I, nakład 300 egz.
Objętość 11,7 ark. wyd.

Skład: Tomasz Wolniewicz
Druk: Zakład Poligraficzno Wydawniczy
POZKAL, Inowrocław
ul. Cegielna 12

Hans Schwedtfeger (1952)
Born, September 12, 1902 in Göttingen, Germany
Died, June 26, 1990 in Adelaide, Australia

We would like to dedicate this work to Professor Hans Schwerdtfeger. Professor Schwerdtfeger attended the University of Göttingen during the years 1927–1933, when mathematics and physics blossomed. He was particularly taken with the exposition and insights to be found in the lectures of Gustav Herglotz and he maintained this admiration throughout his life. The student-teacher relationship ripened into a warm personal relationship between Gustav Herglotz, Hans Schwerdtfeger, and his wife, Hanna Schwerdtfeger.

In 1979 after Professor Schwerdtfeger retired from the mathematics department at McGill, he supervised the publication of Herglotz's collected works (**Gesammelte Schriften**, Vandenhoeck & Ruprecht, Göttingen). Immediately thereafter, Professor Scherdtfeger and R.B. Guenther published Herglotz's **Vorlesungen über die Mechanik der Kontinua** which appeared in the series Teubner-Archive zur Mathematik, B.G. Teubner Verlagsgesellschaft, Leipzig, 1985. We decided then to make available the lecture course on contact transformations, which Professor Herglotz gave in the summer semester of 1932.

These lectures were based on developments of ideas stemming from S. Lie and later researchers. They also contain many insights due to G. Herglotz himself. For historical remarks through 1935, see C. Carathéodory's **Variationsrechnung und Partielle Differentialgleichungen Erster Ordnung**, Teubner Verlagsgesellschaft, Leipzig 1956.

Professor Herglotz emphasized techniques for constructing contact transformations and gave general methods for doing so. He also emphasized the physical applications as well as the connections to geometry. Now, with the increasing interest in Hamiltonian systems, these lectures seem as relevant as ever. We have completely reworked the notes which Professor Schwerdtfeger took as a student and modernized the notation. Certain geometric examples of contact transformations which were taken almost verbatim from the book, **Höhere Geometrie** by Felix Klein, were dropped. Gustav Herglotz did not give many concrete examples in his lectures and to make the material usable, we have added a number of them. Professor Schwerdtfeger paid careful attention to details and strove to make his work understandable to a broad audience. We hope he would be happy with the outcome of our endeavors.

We gratefully acknowledge the support of the **Office of Naval Research**, (ONR-ARI No. N 00014-92-J-1226).

RBG JAG CMG

Contents

I

Contact Transformations in the Plane

§ 1.1 Differential Equations

A differential equation of the first order

$$dy/dx = y' = f(x, y) \tag{1.1}$$

associates with each point, (x, y), in the domain of definition of a real valued function, f, a direction, $p = f(x, y)$. This direction is the slope of the function $y = y(x)$ satisfying (1.1), and passing through the point (x, y). The differential equation (1.1), therefore, defines a direction field, (see Figure 1.1). Integrating (1.1) corresponds to determining a one parameter family of curves, whose tangents are given by f. If a specific point, (a, b), is chosen, through which the curve must pass, it seems reasonable to expect that $y(x)$ is uniquely determined by (1.1).

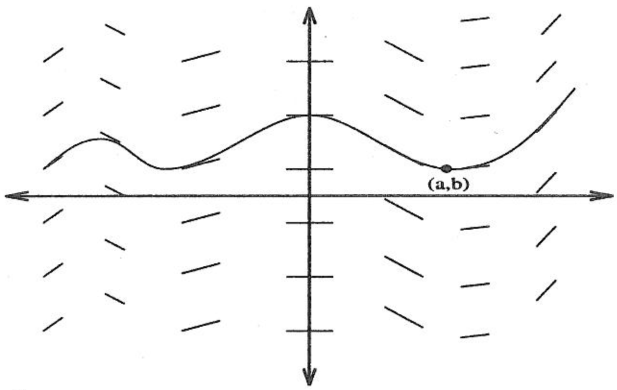

Figure 1.1

This expectation holds under very general circumstances, for example, if f is continuously differentiable. See e.g. Coddington and Levinson[2], or Kamke[1].

The proof of the existence and uniqueness theorem when f is continuously differentiable is usually based on an iteration scheme. The solution is then obtained, as the uniform limit of a sequence of functions. Alternatively, one can seek an explicit solution, or perhaps simplify the differential equation by a substitution,

$$X = X(x, y), \qquad Y = Y(x, y), \tag{1.2}$$

which transforms the variables (x, y), one-to-one, into the (X, Y) variables, and equation (1.1) into

$$dY/dX = F(X, Y). \tag{1.3}$$

The goal is to choose (1.2), so that (1.3) is simpler than the original equation. (1.3) is then solved, and the solution to (1.1) is obtained via the inverse to (1.2). The word simpler here is subjective, and finding the proper transformation, (1.2), is not always straightforward.

Example 1.1. Solve the *Bernoulli* equation

$$\frac{dy}{dx} + 2y = xy^n.$$

SOLUTION. Assume $n \neq 0$, and $n \neq 1$, otherwise the equation is linear.

Let

$$x = X, \qquad y = Y^\alpha,$$

where α is a parameter which will be chosen. Then

$$\frac{dy}{dx} = \alpha Y^{\alpha-1}\frac{dY}{dX},$$

and the differential equation becomes

$$\alpha Y^{\alpha-1}\frac{dY}{dX} + 2Y^\alpha = XY^{n\alpha},$$

or

$$\frac{dY}{dX} + \frac{2}{\alpha}Y = \frac{1}{\alpha}XY^{[(n-1)\alpha+1]}.$$

Choose $\alpha = \frac{-1}{(n-1)}$. The resulting linear equation has the solution

$$Y = \frac{1}{2}X + \frac{1}{4(n-1)} + Ce^{2(n-1)X},$$

and the reverse substitution yields

$$y = \left\{\frac{1}{2}x + \frac{1}{4(n-1)} + Ce^{2(n-1)x}\right\}^{-1/(n-1)}.$$

Example 1.2. A function f is said to be *homogeneous of degree k* if

$$f(tx, ty) = t^k f(x, y) \qquad \text{for all } t,\ x,\ y.$$

If f is homogeneous of degree zero we have

$$f(tx, ty) = f(x, y),$$

and we say the differential equation $y' = f(x, y)$ is *homogeneous*. Such equations can be reduced to ones where the variables separate.

Let

$$X = x, \qquad Y = \frac{y}{x}.$$

Then

$$\frac{dy}{dx} = Y + X\frac{dY}{dX},$$

and the differential equation becomes,

$$X\frac{dY}{dX} + Y = f(X, XY) = f(1, Y),$$

which separates to

$$\frac{dY}{f(1,Y) - Y} = \frac{dX}{X}.$$

One then integrates both sides to solve for Y in terms of X, and reverse substitutes to find y in terms of x.

In both of these examples, substitutions like (1.2) were used to simplify the differential equation. We can gain some additional flexibility by exploiting the geometric properties of the differential equation. We had observed that (1.1) relates the slope p, of the solution curve, to the point (x, y), through which the curve passes. Thus, there are three variables, (x, y, p), which are of importance here. If we transform all three of these variables by a transformation of the form

$$(1.4) \qquad X = X(x, y, p), \qquad Y = Y(x, y, p), \qquad P = P(x, y, p),$$

we should be able to integrate much more general classes of nonlinear differential equations. However, these transformations cannot be completely arbitrary. They have to be chosen in such a way that a curve is transformed into a curve, and the slope, p of the original curve at the point (x, y), is transformed into the slope, P of the image curve at the point (X, Y). The transformations which accomplish this are the **contact transformations**.

§ 1.2 Point Transformations and Extended Point Transformations

A **point transformation** is a mapping which assigns to each point, (x, y) in a domain, (an open, connected set), contained in the Cartesian xy–plane, a point (X, Y) in the XY–plane, which may or may not coincide with the original plane. Thus, a point transformation is determined by the pair of functions

$$X = X(x, y), \qquad Y = Y(x, y), \tag{2.1}$$

where we use the same symbols to denote both the functions as well as the function values.

We shall assume that the functions $X(x, y)$, $Y(x, y)$ are sufficiently smooth so that the computations in the sequel can be carried out. Also, we assume that in a simply connected domain D, contained in the xy–plane, the Jacobian

$$J(x, y) = \frac{\partial(X, Y)}{\partial(x, y)} = \begin{vmatrix} X_x(x, y) & X_y(x, y) \\ Y_x(x, y) & Y_y(x, y) \end{vmatrix} \neq 0, \tag{2.2}$$

so that the transformation (2.1) maps the domain D, one-to-one, onto a domain D' in the XY–plane, and the inverse mapping is continuously differentiable on D'.

Example 2.1. Consider the transformation

$$X = x^2 + y^2, \qquad Y = 2xy,$$

which is defined in the entire xy–plane. Addition and subtraction of these equations yields

$$X + Y = (x + y)^2, \quad \text{and} \quad X - Y = (x - y)^2,$$

and adding again gives

$$2X = (x + y)^2 + (x - y)^2.$$

Thus the image domain D', lies between the rays $Y = X$, and $Y = -X$, with $X > 0$. The domain D is not uniquely determined. One could take; for example, $D = \{ (x, y) : -x < y < x,\ x > 0 \}$. The Jacobian is

$$J(x, y) = \begin{vmatrix} 2x & 2y \\ 2y & 2x \end{vmatrix} = 4(x^2 - y^2),$$

which vanishes on the diagonals $y^2 = x^2$. Normally, the choice of D is dictated by the problem at hand.

Example 2.2. Let a, and b, be fixed real numbers, and let λ be real. Define the transformation

$$X = x + \lambda a, \qquad Y = y + \lambda b.$$

This transformation represents a translation by $(\lambda a, \lambda b)$. We can choose for both D and D', the entire plane, since $J(x, y) = 1$ for all x and y.

Example 2.3. For a given nonzero λ , define

$$X = \lambda x, \qquad Y = \lambda y.$$

This transformation represents a dilatation by the factor λ, where the dilatation may be either a magnification or a contraction depending on the value of λ. Here the Jacobian equals λ^2, so the transformation is valid in the entire plane.

Example 2.4. For λ real, let

$$X = x\cos\lambda - y\sin\lambda, \qquad Y = x\sin\lambda + y\cos\lambda.$$

This transformation represents a counterclockwise rotation of (x, y) about the origin, through the angle λ. The Jacobian is 1.

Example 2.5. The *Lorentz* transformation is defined by

$$X = x\cosh\lambda + y\sinh\lambda, \qquad Y = x\sinh\lambda + y\cosh\lambda.$$

Again the Jacobian is 1.

Note that in the Examples 2.2–2.5, the transformations all depend on the parameter λ. In reality we have a one parameter family of transformations. Such one parameter families of transformations will play a significant role in the applications in the sequel.

We return to the general point transformation case and suppose the domains D, and D', have been determined. Let γ be a continuously differentiable curve lying in D which is described parametrically by

$$\gamma\colon\ x = x(t), \quad y = y(t). \tag{2.3}$$

The slope $p(t)$ of γ at the point $\big(x(t), y(t)\big)$ satisfies

$$p(t)\dot{x}(t) = \dot{y}(t), \tag{2.4}$$

where the dots represent derivatives with respect to t. The point transformation (2.1) maps γ into the curve Γ, which lies in D' and is given by,

$$\Gamma\colon\ X(t) = X\big(x(t), y(t)\big), \quad Y(t) = Y\big(x(t), y(t)\big). \tag{2.5}$$

The slope, $P(t)$, of Γ, satisfies

$$P(t)\,\dot{X}(t) = \dot{Y}(t). \tag{2.6}$$

When $\dot{x}(t) \neq 0$, $\dot{X}(t) \neq 0$, we obtain from equation (2.5), and the chain rule

$$\dot{X} = X_x\dot{x} + X_y\dot{y}, \qquad \dot{Y} = Y_x\dot{x} + Y_y\dot{y},$$

and then by (2.4) and (2.6) the requirement that

$$P = \frac{Y_x + pY_y}{X_x + pX_y}. \tag{2.7}$$

Equation (2.7) associates the slope p, at the point (x, y) of a curve lying in D, with the slope P, at the point (X, Y) of the image curve lying in D'. Its derivation shows that (2.1) together with (2.7) describes a transformation

$$T\colon (x, y, p) \longrightarrow (X, Y, P),$$

having the following property.

If p is the slope of the tangent to a curve γ at the point (x, y), then P is the slope of the tangent to the image curve Γ of γ at the point (X, Y). Two curves, γ' and γ'', passing through (x, y) and having the same tangent there, are mapped by T onto curves, Γ' and Γ'', respectively, having the same tangent at (X, Y).

Thus, Γ' and Γ'' touch each other, or are in contact with each other, in the same manner that γ' and γ'' are. Let us observe also that $X_x(x, y) + pX_y(x, y) \neq 0$ in D, for if it were, we could conclude that since p varies independently of x and y, first that $X_x = 0$, and then that $X_y = 0$, which would violate our assumption (2.2), the non-vanishing of the Jacobian of the transformation. Furthermore, since

$$\frac{\partial(X, Y, P)}{\partial(x, y, p)} = \begin{vmatrix} X_x & X_y & X_p \\ Y_x & Y_y & Y_p \\ P_x & P_y & P_p \end{vmatrix} = \frac{(X_xY_y - X_yY_x)^2}{(X_x + pX_y)^2} \neq 0,$$

the transformation is one-to-one.

The transformation (2.1) extended by (2.7) :

$$\boxed{X = X(x, y), \qquad Y = Y(x, y), \qquad P = \frac{Y_x(x, y) + pY_y(x, y)}{X_x(x, y) + pX_y(x, y)},} \tag{2.8}$$

is called the **extended point transformation** .

In the next section we shall characterize the most general contact transformations.

Example 2.6. The extension of the transformation of Example 1.1 is given by

$$X = x^2 + y^2, \qquad Y = 2xy, \qquad P = \frac{y + px}{x + py}.$$

If one chooses for γ the ray $x = t$, $y = pt$, $t > 0$, $-1 < p < 1$, then Γ is given by the ray $X = (1 + p^2)t^2$, $Y = 2pt^2$, and the slope is $P = 2p/(1 + p^2)$, with $-1 < P < 1$.

As a final remark, if the curve γ is given by $y = f(x)$, it can be trivially parameterized by $x = t$, $y = f(t)$, in which case $p = f'(x)$. Under an extended point transformation, Γ becomes $Y = F(X)$, and $P = F'(X)$.

§ 1.3 Contact Transformations

An **element** is an ordered triple, (x, y, p), where (x, y) defines a point, and p is the slope of a straight line passing through (x, y).

The element (x, y, p) is a **curve element** with respect to a curve γ, if p is the slope of the tangent line of γ at the point (x, y).

A one-to-one, continuously differentiable transformation

$$X = X(x, y, p), \qquad Y = Y(x, y, p), \qquad P = P(x, y, p), \tag{3.1}$$

from a cylindrical domain in the xyp–space,

$$\{\, (x, y, p) : (x, y) \in D,\ |p| \le \tilde{p},\ \tilde{p} \text{ a constant } \}$$

onto a domain in the XYP–space, with a non-vanishing Jacobian,

$$\frac{\partial(X, Y, P)}{\partial(x, y, p)} = \begin{vmatrix} X_x & X_y & X_p \\ Y_x & Y_y & Y_p \\ P_x & P_y & P_p \end{vmatrix} \ne 0$$

is called an **element transformation.**

An example of an element transformation is the extended point transformation (2.8).

An element transformation is a contact transformation if every curve element (x, y, p), is mapped into a curve element (X, Y, P).

Not every element transformation is a contact transformation. For if (x, y, p) are the elements of a curve γ, so that p is the slope of γ at (x, y), then P does not necessarily have to represent the slope of the image curve Γ at (X, Y). See Figure 3.1 for a pictorial example.

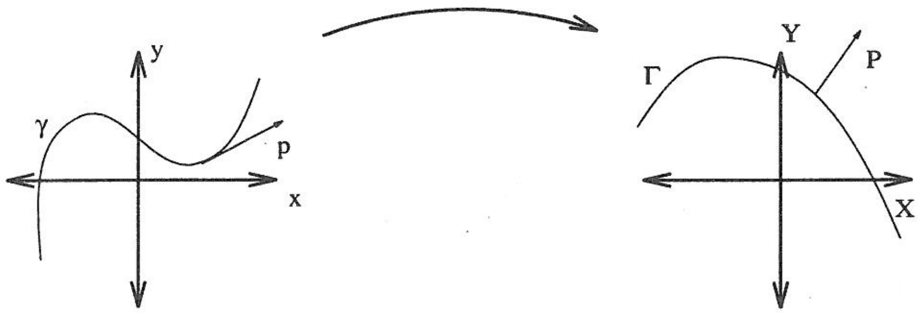

Figure 3.1

Before treating specific examples, it is useful to have an analytic characterization of contact transformations.

A system of three continuously differentiable functions $\big(x(t), y(t), p(t)\big)$, is called a **union of elements** if

$$p\dot{x} - \dot{y} = 0, \tag{3.2}$$

holds for all t.

A curve $\big(x(t), y(t)\big)$ together with the direction

$$p(t) = \dot{y}(t)/\dot{x}(t),$$

of the tangent, is a union of elements. Note, however, that certain degeneracies can occur. For example, a union of elements is also determined by a single point (x_0, y_0), together with an arbitrary coordinate direction, $p(t)$.

Example 3.1. Consider the hyperbola,

$$\gamma\colon y = \sqrt{x^2 + 1}.$$

Curve elements are triples of the form (x, y, p), where $p = x/y$. If we parameterize γ by $x(t) = \sinh t$, $y(t) = \cosh t$, then $p(t) = \tanh t$, and $(\sinh t, \cosh t, \tanh t)$ forms a union of elements for all real t.

The element transformation (3.1) is a contact transformation if it maps unions of elements into unions of elements, that is if

$$\boxed{p\dot{x} - \dot{y} = 0 \quad \text{implies} \quad P\dot{X} - \dot{Y} = 0,} \tag{3.3}$$

where $\big(X(t), Y(t), P(t)\big)$ is the image of $\big(x(t), y(t), p(t)\big)$ under (3.1).

This definition allows us to give an equivalent definition of contact transformations which is often easier to work with.

Suppose $(x(t), y(t), p(t))$ is any union of elements. Then we want conditions on X, Y, and P, such that

$$X(t) = X(x(t), y(t), p(t))$$
$$Y(t) = Y(x(t), y(t), p(t))$$
$$P(t) = P(x(t), y(t), p(t))$$

is also a union of elements. From the first two equations we have

$$\dot{X} = X_x\dot{x} + X_y\dot{y} + X_p\dot{p},$$
$$\dot{Y} = Y_x\dot{x} + Y_y\dot{y} + Y_p\dot{p},$$

and consequently

$$P\dot{X} - \dot{Y} = (PX_x - Y_x)\dot{x} + (PX_y - Y_y)\dot{y} + (PX_p - Y_p)\dot{p}. \tag{3.4}$$

Since $(x(t), y(t), p(t))$ is a curve element, equation (3.2) holds, and

$$P\dot{X} - \dot{Y} = \Big\{(PX_x - Y_x) + p(PX_y - Y_y)\Big\}\dot{x} + (PX_p - Y_p)\dot{p} \tag{3.5}$$

must vanish if (X, Y, P) is to be a union of elements. Equation (3.5) must hold for all curve elements, so that $\dot{x}$ and $\dot{p}$ may be regarded as arbitrary. Consequently if $P\dot{X} - \dot{Y} = 0$, (3.5) yields the two equations

$$(PX_x - Y_x) + p(PX_y - Y_y) = 0, \qquad \text{and} \qquad (PX_p - Y_p) = 0.$$

This leads to:

Theorem 3.1. *An element transformation is a contact transformation if and only if the functions X, Y, and P, satisfy the equations,*

$$\begin{cases} (X_x + pX_y)P = Y_x + pY_y, \\ PX_p = Y_p. \end{cases} \tag{3.6}$$

Proof. We have already shown the necessity of (3.6). To show the sufficiency, let us suppose an element transformation is given which satisfies (3.6). Again (3.5) holds, and since $PX_p = Y_p$, then

$$\begin{aligned} P\dot{X} - \dot{Y} &= (PX_x - Y_x)\dot{x} + (PX_y - Y_y)\dot{y} \\ &= -p(PX_y - Y_y)\dot{x} + (PX_y - Y_y)\dot{y} \\ &= (Y_y - PX_y)(p\dot{x} - \dot{y}), \end{aligned}$$

by the first equation in (3.6). Now $p\dot{x} - \dot{y} = 0$ implies that $P\dot{X} - \dot{Y} = 0$, which proves the theorem. □

Corollary 3.1. *If (3.1) is a contact transformation, the quantity*

$$[X, Y]_{xyp} = (X_x + pX_y)Y_p - (Y_x + pY_y)X_p = 0$$

vanishes on its domain of definition.

Theorem 3.2. *An element transformation is a contact transformation, if and only if, there exists a function* $\rho(x, y, p)$, *such that:*

i) $\rho(x, y, p) \neq 0$ *on the domain of definition of the transformation*
ii) *the functions* ρ, X, Y, *and* P *satisfy,*

$$\begin{cases} PX_x - Y_x - \rho p = 0, \\ PX_y - Y_y + \rho = 0, \\ PX_p - Y_p = 0. \end{cases} \tag{3.7}$$

We will see that

$$P\dot{X} - \dot{Y} = \rho(p\dot{x} - \dot{y}), \tag{3.8}$$

in light of which, ρ *is called the* **multiplier** *for the transformation.*

PROOF. Suppose a function ρ exists so that the element transformation satisfies the equations (3.7). Then by equation (3.5),

$$\begin{aligned} P\dot{X} - \dot{Y} &= (PX_x - Y_x)\dot{x} + (PX_y - Y_y)\dot{y} + (PX_p - Y_p)\dot{p} \\ &= p\rho\dot{x} - \rho\dot{y} \\ &= \rho(p\dot{x} - \dot{y}), \end{aligned}$$

and $p\dot{x} - \dot{y} = 0$, implies that $P\dot{X} - \dot{Y} = 0$. Now suppose the element transformation is a contact transformation. In the proof of Theorem 3.1, we had found that

$$P\dot{X} - \dot{Y} = (Y_y - PX_y)(p\dot{x} - \dot{y}).$$

Define $\rho = Y_y - PX_y$. Thus,

$$PX_x\dot{x} + PX_y\dot{y} + PX_p\dot{p} - Y_x\dot{x} - Y_y\dot{y} - Y_p\dot{p} = \rho p\dot{x} - \rho\dot{y},$$

or

$$(PX_x - Y_x - \rho p)\dot{x} + (PX_y - Y_y + \rho)\dot{y} + (PX_p - Y_p)\dot{p} = 0.$$

Again, the coefficients of $\dot{x}$, $\dot{y}$, $\dot{p}$, must vanish, and this yields (3.7). Finally, $\rho \neq 0$, for if $\rho = 0$ at some point, we find from (3.7) that

$$0 \neq \frac{\partial(X, Y, P)}{\partial(x, y, p)} = \begin{vmatrix} X_x & X_y & X_p \\ Y_x & Y_y & Y_p \\ P_x & P_y & P_p \end{vmatrix} = P \begin{vmatrix} X_x & X_y & X_p \\ X_x & X_y & X_p \\ P_x & P_y & P_p \end{vmatrix} = 0,$$

a contradiction. □

Corollary 3.2. *If* $X = X(x, y, p)$, $Y = Y(x, y, p)$, $P = P(x, y, p)$, *is a contact transformation, then the inverse transformation* $x = x(X, Y, P)$, $y = y(X, Y, P)$, $p = p(X, Y, P)$, *is also a contact transformation.*

An alternative useful formula for ρ, is obtained by multiplying the first formula of equation (3.7) by X_y, the second by X_x, and subtracting. We find

$$\rho = \big(X_x Y_y - X_y Y_x\big)/\big(pX_y + X_x\big) = Y_y - PX_y. \tag{3.9}$$

Example 3.2. The element transformation

$$X = \frac{x}{p} + y, \qquad Y = \frac{y}{p} + x, \qquad P = xy,$$

is not a contact transformation. For if it were, we would need to have for all x, y, and p,

$$xy = P = Y_p/X_p = -(y/p^2)/(-x/p^2) = y/x,$$

which is clearly impossible.

Example 3.3. The *Legendre* transformation

$$X = p, \qquad Y = px - y, \qquad P = x,$$

is a contact transformation. For it is defined for all x, y, p, and it is invertible, the inverse being

$$x = P, \qquad y = PX - Y, \qquad p = X.$$

Moreover,

$$P\dot{X} - \dot{Y} = x\dot{p} - p\dot{x} - \dot{p}x + \dot{y} = -(p\dot{x} - \dot{y}),$$

and $\rho = -1$.

In the next section, we shall give a derivation of the Legendre transformation, as well as a number of related transformations.

Example 3.4. Solve the following initial value problem for the *Clairaut* type equation

$$y = xy' + {y'}^2 + x - 1,$$
$$y(0) = 3.$$

Rewrite this equation as

$$y = px + p^2 + x - 1.$$

Use the Legendre transformation to convert it to the linear equation

$$Y + X^2 + P - 1 = 0,$$

or with $P = dY/dX$,

$$\frac{dY}{dX} + Y = -X^2 + 1.$$

The general solution is

$$Y(X) = -X^2 + 2X - 1 + Ce^{-X}.$$

To determine the constant C, the initial conditions must be transformed. This is done with the help of the differential equation. When $x = 0$, $y = 3$, we have $p^2 = 4$, so there are two possible solutions, one with $p = 2$, and one with $p = -2$. If we take $p = 2$, the point $(0, 3, 2)$ transforms to the point $(2, -3, 0)$, and we must determine C so that $Y(2) = -3$. Thus $C = -2e^2$, and we have,

$$Y(X) = -X^2 + 2X - 1 - 2e^{-(X-2)},$$

which implies

$$P(X) = -2X + 2 + 2e^{-(X-2)}.$$

The solution to the original problem is obtained by applying the inverse transformation. For this it is simplest to think of transforming a union of elements. Parameterize the solution by,

$$\begin{aligned} X &= t, \\ Y(t) &= -t^2 + 2t - 1 - 2e^{-(t-2)}, \\ P(t) &= -2t + 2 + 2e^{-(t-2)}. \end{aligned}$$

The inverse transformation gives the solution in parametric form,

$$\begin{aligned} x(t) &= -2t + 2 + 2e^{-(t-2)}, \\ y(t) &= -t^2 + 1 + 2(t+1)e^{-(t-2)}, \\ p(t) &= t. \end{aligned}$$

Similarly the second solution is found to be

$$\begin{aligned} x(t) &= -2t + 2 - 6e^{-(t+2)}, \\ y(t) &= -t^2 + 1 - 6(t+1)e^{-(t+2)}, \\ p(t) &= t. \end{aligned}$$

Figure 3.2 illustrates the two solutions in the xy–plane.

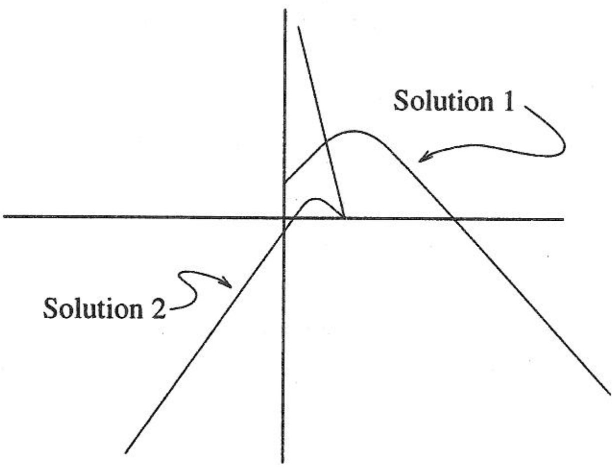

Figure 3.2

Example 3.5 . The extended point transformation is a contact transformation with

$$\rho = (X_x Y_y - X_y Y_x)/(X_x + pX_y).$$

When $\big(X(x,y),\, Y(x,y)\big)$ represents a point transformation, we have seen that the corresponding contact transformation is given by the extended point transformation. We close this section with an answer to the following question.

Suppose $X(x,y,p)$ and $Y(x,y,p)$ have been chosen. Under what conditions is it possible to find a function $P(x,y,p)$, such that the triple (X,Y,P), is a contact transformation?

First, if such a transformation were to exist, the functional determinant $\partial(X,Y,P)/\partial(x,y,p)$ must be nonzero, that is, the row vectors in the matrix representation of the Jacobian would be linearly independent. In particular the rank of the matrix

$$\begin{bmatrix} X_x & X_y & X_p \\ Y_x & Y_y & Y_p \end{bmatrix}$$

must be two. In this case we say that the pair X, Y, is independent.

Next from the Corollary to Theorem 3.1, it is clear that X, Y must satisfy $[X,Y]_{xyp} = 0$.

Finally, by Theorem 3.1, X and Y must be such that

$$X_x + pX_y = 0 \quad \text{implies} \quad Y_x + pY_y = 0, \tag{3.10}$$

and

$$X_p = 0 \quad \text{implies} \quad Y_p = 0, \tag{3.11}$$

because the left hand expressions of (3.10), and (3.11), occur in the denominator of the formula for P, which is a smooth function.

It turns out that these necessary conditions are also sufficient. Thus, once X and Y have been chosen, there is only one way to choose P so that the resulting triple (X,Y,P) is a contact transformation.

Theorem 3.3. *Let $X(x,y,p)$, and $Y(x,y,p)$, be given functions. These functions can be extended to form a contact transformation,*

$$\big(X(x,y,p), Y(x,y,p), P(x,y,p)\big)$$

if and only if the following conditions hold :

i) *$X(x,y,p)$ and $Y(x,y,p)$ are independent,*
ii) *$X(x,y,p)$ and $Y(x,y,p)$ satisfy (3.10) and (3.11),*
iii) $[X,Y]_{xyp} = 0$.

PROOF. The discussion preceeding the formulation of the theorem implies the necessity of (i.)—(iii.) We must show the sufficiency. If (i.)—(iii.) hold, we define

$$P(x,y,p) = \frac{Y_x + pY_y}{X_x + pX_y} = \frac{Y_p}{X_p}.$$

The assumption (iii.) guarantees that both of these definitions for P are the same, provided the denominators do not vanish. However, if they both vanished at a single point, the numerators would also vanish there by (ii.), and so

$$\begin{bmatrix} X_x & X_y & X_p \\ Y_x & Y_y & Y_p \end{bmatrix} = \begin{bmatrix} -pX_y & X_y & 0 \\ -pY_y & Y_y & 0 \end{bmatrix},$$

and the rank would be less than two. Thus, the denominators cannot simultaneously vanish, P is well defined, and we can use either definition.

To show that the resulting transformation is a contact transformation, it is simplest to show there exists a nonzero ρ, for which the equations of Theorem 3.2 are satisfied. Obviously $PX_p - Y_p = 0$.

We set

$$\rho = Y_y - PX_y.$$

Then

$$\begin{aligned} PX_x - Y_x - \rho p &= PX_x - Y_x - p(Y_y - PX_y) \\ &= P(X_x + pX_y) - (Y_x + pY_y) = 0, \end{aligned}$$

and

$$PX_y - Y_y + \rho = (PX_y - Y_y) + (Y_y - PX_y) = 0.$$

Finally; if $\rho = 0$ at some point, then $Y_y = PX_y$ and $Y_x = PX_x$, so that

$$\begin{bmatrix} X_x & X_y & X_p \\ Y_x & Y_y & Y_p \end{bmatrix} = P \begin{bmatrix} X_x & X_y & X_p \\ X_x & X_y & X_p \end{bmatrix},$$

which has rank less than two. □

As a final remark, we note that it is possible to develop a theory for higher order contacts of curves; however, the formulas become unwieldy, and they are difficult to apply in concrete situations.

§ 1.4 The Directrix

Suppose $X(x,y,p)$, $Y(x,y,p)$, $P(x,y,p)$, represents a contact transformation. Furthermore, suppose that the functions

$$X = X(x,y,p), \quad \text{and} \quad Y = Y(x,y,p), \tag{4.1}$$

depend upon p, so that the contact transformation does not arise from the extension of a point transformation. As long as there exists a point (x_0, y_0, p_0), where $X(x_0, y_0, p_0) = 0$, then at points where $X_p(x, y, p) \neq 0$, we can solve for p in terms of X, x, and y, to obtain $p = p(X, x, y)$. Insert this expression into the second equation in (4.1) to find

$$Y = Y\big(x, y, p(X, x, y)\big).$$

This equation can be written more generally in the form $\Omega(X, Y; x, y) = 0$, where in the case just discussed, $\Omega = Y\big(x, y, p(X, x, y)\big) - Y$, but more general possibilities are allowed.

A nontrivial function, $\Omega = \Omega(X, Y; x, y)$, i.e. one for which at least one of its partial derivatives is not identically zero, is called the **directrix** for a contact transformation if

$$\Omega\big(X(x, y, p), Y(x, y, p); x, y\big) = 0 \tag{4.2}$$

for all values, (x, y, p) in the domain of definition of the contact transformation. An equation of the form

$$\Omega(X, Y; x, y) = 0 \tag{4.3}$$

is called a **directrix equation**.

Example 4.1. The first two equations of the Legendre transformation are

$$X = p, \quad \text{and} \quad Y = px - y.$$

The third equation is $P = Y_p/X_p = x$. The directrix equation is obtained by substituting $p = X$ into the equation for Y, so that $Y = xX - y$, or

$$\Omega(X, Y; x, y) = xX - (Y + y).$$

In this section we shall determine sufficient conditions enabling us to derive contact transformations from a given directrix equation. We will use the directrix in the next section, to obtain a wide class of useful contact transformations. At the end of this section we shall return to the case where X and Y are independent of p.

Let $\Omega(X, Y; x, y)$ be the directrix for a contact transformation with $X_p \neq 0$. Differentiate (4.2) with respect to the variables x, y, and p, to obtain the system

$$\Omega_X X_x + \Omega_Y Y_x + \Omega_x = 0, \tag{4.4}$$

$$\Omega_X X_y + \Omega_Y Y_y + \Omega_y = 0, \tag{4.5}$$

$$\Omega_X X_p + \Omega_Y Y_p = 0. \tag{4.6}$$

Since $Y_p = PX_p$, and $X_p \neq 0$, equation (4.6) yields

$$\Omega_X + P\Omega_Y = 0. \tag{4.7}$$

Now multiply equation (4.5) by p, and add the result to (4.4), to find

$$(X_x + pX_y)\Omega_X + (Y_x + pY_y)\Omega_Y + \Omega_x + p\Omega_y = 0.$$

Recall from (3.6) that $(X_x + pX_y)P = Y_x + pY_y$, so

$$(X_x + pX_y)(\Omega_X + P\Omega_Y) + \Omega_x + p\Omega_y = 0,$$

and by (4.7),

$$\Omega_x + p\Omega_y = 0.$$

We summarize these calculations in the following theorem.

Theorem 4.1. *Let $X = X(x, y, p)$, $Y = Y(x, y, p)$, $P = P(x, y, p)$, be a contact transformation, and let $\Omega = \Omega(X, Y; x, y)$ be a directrix for the transformation. Then the equations*

$$\Omega = 0, \tag{4.8}$$

$$\Omega_X + P\Omega_Y = 0, \tag{4.9}$$

$$\Omega_x + p\Omega_y = 0, \tag{4.10}$$

are satisfied.

Observe that as an immediate consequence of (4.9) we see that if at some point $\Omega_Y = 0$, then $\Omega_X = 0$ as well. Similarly by (4.10) $\Omega_y = 0$, implies $\Omega_x = 0$.

Corollary 4.1. *None of the equations*

$$\Omega = \Omega(X; x, y), \quad \Omega = \Omega(Y; x, y), \quad \Omega = \Omega(X, Y; x), \quad \Omega = \Omega(X, Y; y),$$

can be the directrix of a contact transformation with $X_p \neq 0$.

PROOF. Suppose $\Omega = \Omega(X; x, y)$. Then $\Omega_Y = 0$, which implies $\Omega_X = 0$, so $\Omega = \Omega(x, y)$. Differentiate (4.10) with respect to p to conclude that $\Omega_y = 0$, hence $\Omega_x = 0$, i.e. Ω is identically zero. The other cases are treated similarly. □

Example 4.2. The system of equations (4.8), (4.9), and (4.10), for the Legendre transformation is

$$\Omega = xX - (Y + y) = 0,$$
$$\Omega_X + P\Omega_Y = x - P = 0,$$
$$\Omega_x + p\Omega_y = X - p = 0,$$

and it is obvious that we can solve these equations for X, Y, and P, in terms of x, y, and p, to reconstruct the original transformation.

The equations (4.8), (4.9), and (4.10), represent a system of three equations in three unknowns, X, Y, and P. We would expect that under certain circumstances, we could use the system to construct contact transformations. While that is true, the following examples demonstrate that additional hypotheses are required.

The first example shows that not every function of the four variables X, Y; x, y, is the directrix of a contact transformation.

Example 4.3. Let

$$\Omega(X, Y; x, y) = xX + yY = 0.$$

Equations (4.9) and (4.10) are

$$x + Py = 0, \quad \text{and} \quad X + pY = 0.$$

But the first and third of these equations constitute a linear system

$$\begin{bmatrix} x & y \\ 1 & p \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

whose determinant, $px - y$, is in general nonzero, so X and Y vanish identically.

Example 4.4. Now suppose

$$\Omega(X, Y; x, y) = xX + yY - 1 = 0.$$

Then equations (4.9) and (4.10) are

$$x + Py = 0, \quad \text{and} \quad X + pY = 0,$$

and this system of three equations has the unique solution

$$X = \frac{p}{(px - y)}, \qquad Y = \frac{-1}{(px - y)}, \qquad P = \frac{-x}{y}.$$

It is easily checked that this is a contact transformation. Its inverse is

$$x = \frac{P}{(PX - Y)}, \qquad y = \frac{-1}{(PX - Y)}, \qquad p = \frac{-X}{Y},$$

and the function $\rho = 1/\big(y(px - y)\big)$.

It is advantageous to have a number of contact transformations at one's disposal. Often, one can significantly simplify a differential equation through the use of an appropriate contact transformation.

Example 4.5. Solve the initial value problem

$$y'(2y+x^2)-xy=0,$$
$$y(0)=1.$$

Let $y'=p$ and rearrange the differential equation to

$$\frac{2p}{px-y}=\frac{-x}{y}.$$

Apply the contact transformation of Example 4.4 to obtain the differential equation

$$P=2X.$$

The initial conditions $x=0$, $y=1$, applied to the original differential equation give $p=0$. The initial curve element $(0,1,0)$, in the xyp-domain transforms into the curve element $(0,1,0)$ in the XYP-domain. Thus the initial condition becomes $Y(0)=1$. The solution therefore is,

$$Y=X^2+1.$$

The representation of the solution in the xyp-element space is determined most simply by letting $X=t$, so $Y=t^2+1$, and $P=2t$. The parametric form of the solution is then

$$x=\frac{2t}{(t^2-1)},\qquad y=\frac{-1}{(t^2-1)}.$$

In this case, we can explicitly eliminate the parameter t. After some rearranging, we find

$$y=\frac{1}{2}\left(1+\sqrt{1+x^2}\right).$$

See Figure 4.1.

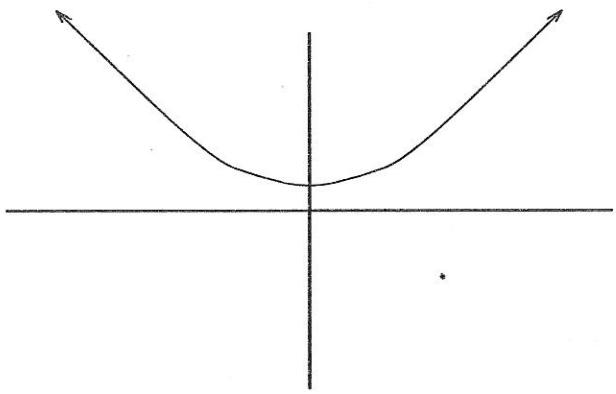

Figure 4.1

Example 4.6. For a, b, c, d, and e, constants, let

$$\Omega = aX + bY + cx + dy + e.$$

The other two equations of the system are then, $a + Pb = 0$, and $c + pd = 0$, leading to the conclusion that P and p are constant. A function $\Omega(X, Y; x, y)$ which depends linearly on its arguments cannot, therefore, be the directrix for a contact transformation.

Example 4.7. Let $\Omega(X, Y; x, y) = Y^2 + xX - y$. The system dictated by (4.8), (4.9), and (4.10), is

$$\begin{aligned} Y^2 + xX - y &= 0, \\ x + 2YP &= 0, \\ X - p &= 0. \end{aligned}$$

The solution to this system is

$$\begin{aligned} X &= p, \\ Y &= \pm\sqrt{y - px}, \\ P &= \mp x / \left(2\sqrt{y - px}\right). \end{aligned}$$

It is easily checked that both of these transformations are contact transformations. In the case of the plus sign for Y, and the minus sign for P, $\rho = 1/\left(2\sqrt{y - px}\right)$. This example demonstrates that one directrix can generate several contact transformations.

The general theorem telling when a single directrix will generate a contact transformation is based on the Implicit Function Theorem. We really only need to be able to solve (4.8) and (4.10) for X, and Y, in terms of x, y, and p, because P is determined immediately from (4.9). Moreover, the transformation must be invertible, which means that we must also be able to solve for x, y, and p, in terms of X, Y, and P. In other words analogous assumptions will be made concerning the solvability of (4.8) and (4.9).

Let us suppose that there is a point $(X_0, Y_0, P_0; x_0, y_0, p_0)$ satisfying (4.8)—(4.10), and suppose that the Jacobian

$$\frac{\partial(\Omega, \Omega_X + P\Omega_Y, \Omega_x + p\Omega_y)}{\partial(X, Y, P)} = -\Omega_Y \frac{\partial(\Omega, \Omega_x + p\Omega_y)}{\partial(X, Y)} \neq 0.$$

Then in a neighborhood of $(X_0, Y_0, P_0; x_0, y_0, p_0)$, (4.8)—(4.10) can be solved for X, Y, and P, in terms of x, y, and p, to obtain the element transformation

$$X = X(x, y, p), \qquad Y = Y(x, y, p), \qquad P = P(x, y, p). \tag{4.11}$$

To show that (4.11) is a contact transformation, we must show that it is one-to-one, and that $p\dot{x} - \dot{y} = 0$, implies that $P\dot{X} - \dot{Y} = 0$.

Let $(X(t), Y(t), P(t))$, and $(x(t), y(t), p(t))$, denote unions of elements satisfying (4.8)—(4.10). Differentiate (4.8) with respect to t to find

$$\Omega_X \dot{X} + \Omega_Y \dot{Y} + \Omega_x \dot{x} + \Omega_y \dot{y} = 0.$$

From (4.9) and (4.10) we have $\Omega_X = -P\Omega_Y$, and $\Omega_x = -p\Omega_y$, which gives

$$-\Omega_Y(P\dot{X} - \dot{Y}) - \Omega_y(p\dot{x} - \dot{y}) = 0,$$

so that

$$P\dot{X} - \dot{Y} = (-\Omega_y/\Omega_Y)\,(p\dot{x} - \dot{y}).$$

If we require that $\Omega_y \neq 0$, and $\Omega_Y \neq 0$, then $p\dot{x} - \dot{y} = 0$ if and only if $P\dot{X} - \dot{Y} = 0$. Differentiating (4.8), (4.9), and (4.10), successively, with respect to x, y, and p, yields,

$$\frac{\partial(\Omega, \Omega_X + P\Omega_Y, \Omega_x + p\Omega_y)}{\partial(X, Y, P)} \frac{\partial(X, Y, P)}{\partial(x, y, p)} = \frac{\partial(\Omega, \Omega_X + P\Omega_Y, \Omega_x + p\Omega_y)}{\partial(x, y, p)}.$$

If we now demand that the Jacobian on the right hand side be nonzero, and note that it is equal to

$$\Omega_y \frac{\partial(\Omega, \Omega_X + P\Omega_Y)}{\partial(x, y)},$$

we arrive at the following theorem.

Theorem 4.2. *Suppose $\Omega(X, Y; x, y)$ satisfies:*

i) $\Omega_Y(X, Y; x, y) \neq 0, \qquad \Omega_y(X, Y; x, y) \neq 0,$

ii) $\dfrac{\partial(\Omega, \Omega_x + p\Omega_y)}{\partial(X, Y)} \neq 0, \qquad \dfrac{\partial(\Omega, \Omega_X + P\Omega_Y)}{\partial(x, y)} \neq 0.$

Then the solution to the (4.8)—(4.10) system :

$$\boxed{\begin{aligned} \Omega(X, Y; x, y) &= 0, \\ \Omega_X + P\Omega_Y &= 0, \\ \Omega_x + p\Omega_y &= 0, \end{aligned}}$$

in the neighborhood of a point $(X_0, Y_0, P_0; x_0, y_0, p_0)$, which satisfies the (i.)—(ii.) system, is a contact transformation.

We have seen that if $X(x,y,p)$, $Y(x,y,p)$, are the first two functions of a contact transformation with $X_p \neq 0$, a single directrix equation can be constructed for it. Conversely, given a single function, $\Omega(X,Y;x,y)$, it is sometimes possible to obtain a contact transformation depending upon x, y, and p.

In the case where $X = X(x,y)$, $Y = Y(x,y)$, i.e. where X, and Y, are independent of p, the above construction of a directrix fails; however, the two functions may still be used to obtain the extended point transformation, which we know to be a contact transformation, thus we can think of X and Y as two directrixes.

Conversely, suppose we are given two equations

$$\begin{cases} F(X,Y;x,y) = 0 \\ G(X,Y;x,y) = 0, \end{cases} \tag{4.12}$$

and suppose they are satisfied at some point $(X_0, Y_0; x_0, y_0)$. If the Jacobian

$$\frac{\partial(F,G)}{\partial(X,Y)} \neq 0$$

there, then we can solve for X and Y in terms of x and y. Additionally, if we proceed as in the proof of Theorem 4.2, and require that

$$\frac{\partial(F,G)}{\partial(x,y)} \neq 0,$$

then we can conclude that

$$\frac{\partial(X,Y)}{\partial(x,y)} \neq 0,$$

and hence the pair $X(x,y)$, $Y(x,y)$, can be extended to a contact transformation.

To summarize, a single function Ω, of the four variables $(X,Y;x,y)$, will, under certain conditions, give rise to a contact transformation, where the functions X and Y depend upon x, y, and p. Similarly two functions F and G of the four variables $(X,Y;x,y)$, will, under certain conditions, give rise to functions, X and Y of (x,y), which can then be extended to a contact transformation.

§ 1.5 The Contact Transformations of the Polarity

In this section we study contact transformations which arise from directrix functions of the form

$$\Omega(X,Y;x,y) = axX + b(yX + xY) + cyY + \alpha(x+X) + \beta(y+Y) + \kappa = 0, \tag{5.1}$$

where we assume throughout that $a^2 + b^2 + c^2 \neq 0$. From the definition, it is obvious that Ω is symmetric in (X, Y) and (x, y), that is

$$\Omega(X, Y; x, y) = \Omega(x, y; X, Y). \tag{5.2}$$

For $x = X$, $y = Y$, the equation,

$$\Omega(X, Y; x, y) = aX^2 + 2bXY + cY^2 + 2\alpha X + 2\beta Y + \kappa = 0, \tag{5.3}$$

describes the conic sections in the plane. It is possible to introduce new coordinates and reduce the complexity of (5.1); however, for applications it is more convenient to work with (5.1) directly.

Let us set

$$\begin{cases} A(x,y) := ax + by + \alpha, \\ B(x,y) := bx + cy + \beta, \\ C(x,y) := \alpha x + \beta y + \kappa, \end{cases} \tag{5.4}$$

and rewrite (5.1) in the form,

$$\Omega(X, Y; x, y) = A(x,y)X + B(x,y)Y + C(x,y), \tag{5.5}$$

which, because of the symmetry relationship (5.2), is also given by

$$\Omega(x, y; X, Y) = A(X,Y)x + B(X,Y)y + C(X,Y). \tag{5.6}$$

From the representation (5.5), we observe that for a fixed point $(\overline{x}, \overline{y})$, $\Omega(X, Y; \overline{x}, \overline{y}) = 0$, represents the equation of a straight line in the XY-plane (assuming of course that $A^2(\overline{x}, \overline{y}) + B^2(\overline{x}, \overline{y}) \neq 0$). The slope P, of the straight line is

$$P = -\frac{A(\overline{x}, \overline{y})}{B(\overline{x}, \overline{y})}.$$

If $(\overline{x}, \overline{y})$ lies on the conic section, $\Omega(x, y; x, y) = 0$, then P is the slope of the tangent line to the conic section, at the point $(\overline{x}, \overline{y})$.

Now let us suppose that a straight line

$$\ell x + my + n = 0, \qquad \ell^2 + m^2 \neq 0,$$

is given in the xy-plane.

Use the representation (5.6) for Ω to find a point $(\overline{X}, \overline{Y})$ such that the straight line,

$$\Omega(x, y; \overline{X}, \overline{Y}) = 0,$$

coincides with the given line. This will be the case when there exists a constant λ such that ,

$$(5.7)\qquad \begin{cases} A(\overline{X},\overline{Y}) = \lambda\ell, \\ B(\overline{X},\overline{Y}) = \lambda m, \\ C(\overline{X},\overline{Y}) = \lambda n. \end{cases}$$

Equation (5.7) represents a system of three equations in the three unknowns $\overline{X}$, $\overline{Y}$, and λ. This system is uniquely solvable when the coefficient determinant is nonzero, i.e. when

$$\begin{vmatrix} a & b & -\ell \\ b & c & -m \\ \alpha & \beta & -n \end{vmatrix} \neq 0.$$

In this construction, the point $(\overline{X},\overline{Y})$ is called the **pole** and the line $\Omega(x,y;\overline{X},\overline{Y})$ the **polar**. The one-to-one relationship between poles and lines is called the **polarity**.

Example 5.1. Let

$$\Omega(X,Y;X,Y) = (X/2)^2 + Y^2 - 1 = 0 \quad \text{be an ellipse.}$$

Then

$$\Omega(X,Y;x,y) = xX/4 + yY - 1 = 0.$$

If $x = 4$, $y = 1$, the straight line, or polar, is $X + Y - 1 = 0$.

On the other hand, if, for example, $y = x - 4$, the corresponding point, or pole, is obtained by solving the (5.7) system,

$$1/4\overline{X} = -\lambda, \qquad \overline{Y} = \lambda, \qquad 1 = -4\lambda$$

so that $\lambda = -1/4$, $\overline{X} = 1$, $\overline{Y} = -1/4$. See Figure 5.1.

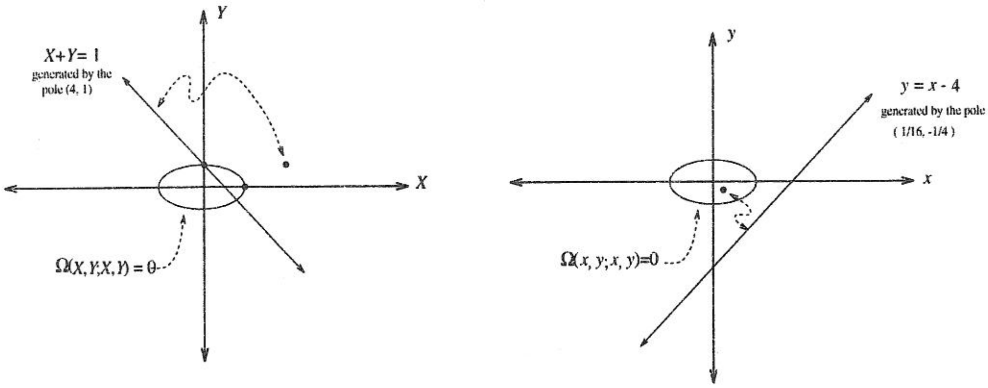

Figure 5.1

The polar equation, $\Omega = 0$, of a conic section, under certain conditions, is the directrix of a contact transformation. By Theorem 4.1, we must solve the system

$$\left\{\begin{aligned} A(x,y)X + B(x,y)Y + C(x,y) &= 0, \\ A(x,y) + PB(x,y) &= 0, \\ A(X,Y) + pB(X,Y) &= 0, \end{aligned}\right. \tag{5.8}$$

for (X, Y, P) in terms of (x, y, p). Observe that the first equation can be written as

$$A(X,Y)x + B(X,Y)y + C(X,Y) = 0,$$

so that the system is symmetric in the variables (X, Y, P) and (x, y, p). This means that the solution of (5.8) for (x, y, p) in terms of (X, Y, P) can be obtained by simply interchanging the variables. Symbolically, if we represent the solution to (5.8) for (X, Y, P) in terms of (x, y, p) by

$$(X, Y, P) = T(x, y, p), \tag{5.9}$$

then the inverse transformation

$$(x, y, p) = T^{-1}(X, Y, P), \tag{5.10}$$

obeys

$$T = T^{-1}. \tag{5.11}$$

Transformations which satisfy equation (5.11) are incidentally called **involutions**.

Example 5.2. The polar equation for the ellipse

$$2X^2 + Y^2 + 4X = 1,$$

is $\Omega(X, Y; x, y) = 2xX + yY + 2(X + x) - 1 = 0$. The system (5.8) is then

$$\begin{aligned} (2x+2)X + yY &= 1 - 2x, \\ (2x+2) + Py &= 0, \\ (2X+2) + pY &= 0. \end{aligned}$$

The contact transformation is

$$\begin{aligned} X &= \big(p(1-2x) + 2y\big)/(2px + 2p - 2y), \\ Y &= 3/(y - px - p), \\ P &= -(2x+2)/y, \end{aligned}$$

and because of (5.11), we immediately find the inverse

$$\begin{aligned} x &= \big(P(1-2X) + 2Y\big)/(2PX + 2P - 2Y), \\ y &= 3/(Y - PX - P), \\ p &= -(2X+2)/Y. \end{aligned}$$

Example 5.3. The polar equation for an imaginary conic section can also be used as the directrix for generating a contact transformation. The imaginary circle

$$X^2 + Y^2 + 1 = 0,$$

has the polar equation

$$xX + yY + 1 = 0,$$

and the resulting contact transformation is

$$\begin{aligned} X &= p/(px - y), \\ Y &= -1/(px - y), \\ P &= -x/y. \end{aligned}$$

Once again in view of (5.11), the inverse is

$$\begin{aligned} x &= P/(PX - Y), \\ y &= -1/(PX - Y), \\ p &= -X/Y. \end{aligned}$$

Example 5.4. Not every quadratic form has a polar equation which is the directrix for a contact transformation. We already have seen (cf. Example 4.3) that the polar equation for the circle with radius zero

$$\Omega(X, Y; X, Y) = X^2 + Y^2$$

does not lead to a contact transformation. Nor do the polar equations for

$$\Omega(X, Y; X, Y) = (X \pm Y)^2 + 2\alpha(X \pm Y) + \kappa,$$

which is essentially a function of the single variable $\xi = X \pm Y$, lead to contact transformations.

Theorem 4.1 can be applied to the polar equation $\Omega = 0$, and this will allow us to conclude that in the regions where the denominators do not vanish, the resulting transformation is a contact transformation, which is an involution.

Theorem 5.1. *Let*

$$\begin{aligned} \Delta(x, y, p) &= (ac - b^2)(px - y) + (\alpha c - \beta b)p + (\alpha b - a\beta) \\ \delta(x, y) &= bx + cy + \beta. \end{aligned}$$

Suppose that $\Delta(x, y, p) \neq 0$, $\delta(x, y) \neq 0$, *and that* $\Delta(X, Y, P) \neq 0$, $\delta(X, Y) \neq 0$. *Then the solution to the (5.8) system,*

$$(5.12) \qquad \begin{cases} X = \dfrac{(\beta b - ac)(px - y) + (\beta^2 - c\kappa)p + (\alpha\beta - b\kappa)}{\Delta(x, y, p)} \\ Y = \dfrac{(\alpha b - a\beta)(px - y) + (b\kappa - \alpha\beta)p + (a\kappa - \alpha^2)}{\Delta(x, y, p)} \\ P = \dfrac{ax + by + \alpha}{\delta(x, y)}, \end{cases}$$

is a contact transformation. Furthermore, the inverse transformation is obtained by interchanging (X, Y, P) *with* (x, y, p), *respectively in (5.12).*

§ 1.6 Contact Transformations and Differential Equations

The main application we have in mind for contact transformations is the integration of differential equations. We have seen that in order for a solution of a transformed differential equation to be a solution of the original differential equation, the transformation must necessarily be a contact transformation. However, an additional condition is also required, as is illustrated by the following example.

Example 6.1. Consider the Clairaut Equation

$$y = px + f(p), \qquad p = y'(x).$$

Under the Legendre transformation

$$\begin{aligned} X &= p, \\ Y &= px - y, \\ P &= x, \end{aligned}$$

every solution transforms into a solution of

$$Y = -f(X),$$

which is not a differential equation.

If we set $X = t$, $Y = -f(t)$, $P = -f'(t)$, we find that it transforms back to a curve whose curve elements are $\left(-f(t),\, -tf'(t) + f(t),\, t\right)$. This solution is a singular solution for the Clairaut equation. It is the envelope of the solutions with $p = t$ giving the parameterization.

Although singular solutions are occasionally of interest, our primary focus is on initial value problems of the form

$$p = y' = f(x, y). \tag{6.1}$$

Under a contact transformation, (6.1) transforms into a nonlinear (in general) equation involving X, Y, P, and we would then like to solve for P to obtain

$$P = \frac{dY}{dX} = F(X, Y). \tag{6.2}$$

It is this equation which is integrated, subject to the transformed initial conditions. To guarantee that equation (6.2) can be set up, we need to introduce an additional concept relating the contact transformation and the differential equation.

Consider the initial value problem

$$\begin{cases} p = y' = f(x, y) \\ y(x_0) = y_0 \end{cases} \tag{6.3}$$

and determine p_0 by

$$p_0 = f(x_0, y_0). \tag{6.4}$$

We wish to find a union of elements, $\big(x(t), y(t), p(t)\big)$, which satisfies (6.3), and has the element (x_0, y_0, p_0), as one of its members.

Let

$$X = X(x, y, p), \qquad Y = Y(x, y, p), \qquad P = P(x, y, p) \tag{6.5}$$

be a contact transformation defined on a region of xyp-space containing the point (x_0, y_0, p_0). Let

$$X_0 = X(x_0, y_0, p_0), \qquad Y_0 = Y(x_0, y_0, p_0), \qquad P_0 = P(x_0, y_0, p_0) \tag{6.6}$$

and let

$$x = x(X, Y, P), \qquad y = y(X, Y, P), \qquad p = p(X, Y, P) \tag{6.7}$$

denote the inverse of (6.5). The differential equation in (6.3) transforms into

$$p(X, Y, P) = f\big(x(X, Y, P), y(X, Y, P)\big). \tag{6.8}$$

The next step is to solve (6.8) for P in terms of (X, Y).

We say that the contact transformation (6.5) is **admissible** for (6.3) if it transforms solutions to

$$p = f(x, y) \quad \text{with} \quad p = dy/\,dx \quad \text{and} \quad y(x_0) = y_0$$

into solutions to

$$P = F(X, Y) \quad \text{with} \quad P = dY/\,dX \quad \text{and} \quad Y(X_0) = Y_0.$$

We summarize the above discussion as the following theorem.

Theorem 6.1. *The contact transformation (6.5) is admissible for the problem (6.3) in a neighborhood of (x_0, y_0, p_0) if*

$$\frac{\partial}{\partial P}\Big\{p(X, Y, P) - f\big(x(X, Y, P), y(X, Y, P)\big)\Big\}\Bigg|_{(X_0, Y_0, P_0)} \neq 0.$$

If the transformation is admissible, then the initial value problem (6.3) transforms into an initial value problem of the form

$$\begin{cases} P = F(X, Y) \\ Y(X_0) = Y_0, \end{cases} \tag{6.9}$$

and solutions to (6.3) transform to solutions to (6.9), and conversely.

Example 6.2. Consider the general Clairaut Equation

$$f(p, px - y) = 0, \qquad f(p_0, p_0x_0 - y_0) = 0, \qquad \frac{\partial f(p, px - y)}{\partial p} \neq 0.$$

Under a Legendre transformation: $X = p,\ Y = px - y,\ P = x$, this equation becomes

$$f(X, Y) = 0.$$

Thus the Legendre transformation is not admissible for this equation.

Example 6.3. Let $f(x, p, px - y) = 0$ be such that the equation can be solved for p. Under the Legendre transformation this equation becomes

$$f(P, X, Y) = 0$$

and if P can be solved for, the transformation is admissible.

Most of the earlier examples of transformations applied to differential equations were admissible.

Example 6.4. A differential equation may be inadmissible with respect to some transformation but still be integrable. For example, the Legendre transformation applied to the Clairaut equation

$$y = px + p \quad \text{leads to} \quad Y = -X$$

and so is inadmissible. We saw that the reverse transformation leads to the envelope of solutions, in this case described by the element $x = -1$, $y = 0$, p arbitrary. The general solution to this equation is

$$y = C(x + 1), \quad C \text{ a constant,}$$

and every solution passes through $(-1, 0)$ with some slope C.

References

1. E. Kamke, *Differentialgleichungen, Lösungsmethoden und Lösungen*; I, *Gewöhnliche Differentialgleichungen*, Akademische Verlagsgesellschaft, Leipzig, 1951; II, *Partielle Differentialgleichungen erster Ordnung für eine gesuchte Funktion*, Akademische Verlagsgesellschaft, Leipzig, 1948.

2. E.A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, Robert E. Krieger, Malabar, Florida, 1984.

II

Contact Transformations in Space

§ 2.1 Extended Point Transformations

A point transformation in the three dimensional xyz-space can be represented by the functions

$$(1.1) \qquad X = X(x,y,z) \qquad Y = Y(x,y,z) \qquad Z = Z(x,y,z),$$

which associate with each point, (x,y,z), in the domain of definition, a point, (X,Y,Z), in the range or imagespace. We shall assume that these functions are as continuously differentiable as necessary, and that the mapping is one-to-one, from a domain in the xyz-space, onto a domain in the XYZ-space. Moreover, we require that the inverse transformation is continuously differentiable as well, so that the Jacobian

$$(1.2) \qquad \frac{\partial(X,Y,Z)}{\partial(x,y,z)} = \begin{vmatrix} X_x & X_y & X_z \\ Y_x & Y_y & Y_z \\ Z_x & Z_y & Z_z \end{vmatrix} \neq 0.$$

Now suppose that σ is a surface described by the equation

$$(1.3) \qquad \sigma: \quad z = f(x,y).$$

Its tangent plane at an arbitrary point (a,b,c), where $c = f(a,b)$ is described by

$$z - c = p(x-a) + q(y-b)$$

where $p = f_x(a,b) = z_x(a,b)$, and $q = f_y(a,b) = z_y(a,b)$. The vector $(-p,-q,1)$ is the direction of the normal to the tangent plane of the surface

σ. The components are proportional to the standard direction cosines of the normal (α, β, γ) i.e.

$$-p : -q : 1 = \alpha : \beta : \gamma,$$

whence p and q are referred to as the direction coefficients of the tangent plane.

Under (1.1), surface σ is transformed into a surface

$$\Sigma: \quad Z = F(X, Y), \tag{1.4}$$

and the point $(a, b, c) \in \sigma$ is transformed into the point $(A, B, C) \in \Sigma$, where $A = X(a, b, c)$, $B = Y(a, b, c)$, $C = Z(a, b, c)$.

The corresponding tangent plane to Σ, at (A, B, C), is given by

$$Z - C = P(X - A) + Q(Y - B), \tag{1.5}$$

where $P = Z_X$, and $Q = Z_Y$. To determine P and Q in terms of p and q, we observe that by equation (1.4),

$$Z(x, y, f(x, y)) = F\left(X\left(x, y, f(x, y)\right), Y\left(x, y, f(x, y)\right)\right).$$

Differentiate first with respect to x, and then with respect to y, and use the definitions of p, q, P, Q, to obtain

$$\begin{cases} Z_x + pZ_z = (X_x + pX_z)P + (Y_x + pY_z)Q, \\ Z_y + qZ_z = (X_y + qX_z)P + (Y_y + qY_z)Q. \end{cases} \tag{1.6}$$

(1.6) is a system of two equations in two unknowns, P and Q. If we denote by $\Delta = \Delta(x, y, z, p, q)$ the determinant of the coefficients,

$$\Delta = (X_xY_y - X_yY_x) + p(X_zY_y - Y_zX_y) + q(X_xY_z - X_zY_x), \tag{1.7}$$

the solution to the system is given by

$$\begin{cases} P = \dfrac{(Y_yZ_x - Y_xZ_y) + p(Y_yZ_z - Y_zZ_y) + q(Y_zZ_x - Y_xZ_z)}{\Delta} \\ Q = \dfrac{(X_xZ_y - X_yZ_x) + p(X_zZ_y - X_yZ_z) + q(X_xZ_z - X_zZ_x)}{\Delta}. \end{cases} \tag{1.8}$$

The equations (1.1), and (1.8), represent the **extended point transformation** in 3-D. It associates with each element of a plane (x, y, z, p, q), another planar element (X, Y, Z, P, Q). The mapping is one-to-one since the Jacobian

$$\frac{\partial(X, Y, Z, P, Q)}{\partial(x, y, z, p, q)} = \frac{\partial(X, Y, Z)}{\partial(x, y, z)} \left\{ \frac{\partial P}{\partial p}\frac{\partial Q}{\partial q} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q} \right\}$$

can be shown, after a clumsy calculation, to be nonzero. By its derivation, the transformation (1.1), (1.8) preserves first order contact, and we again call such a transformation, a contact transformation.

Example 1.1. The point transformation

$$X = yz, \qquad Y = xz, \qquad Z = xy$$

maps the domain $\{ x > 0, y > 0, z > 0 \}$ one-to-one, onto the domain $\{ X > 0, Y > 0, Z > 0 \}$. The Jacobian is $2xyz$.

The inverse transformation is

$$x = \sqrt{YZ/X}, \qquad y = \sqrt{ZX/Y}, \qquad z = \sqrt{XY/Z}.$$

The extended point transformation is obtained by appending the two additional equations

$$P = \frac{xyq - xz - px^2}{\Delta},$$
$$Q = \frac{xyp - yz - qy^2}{\Delta},$$

with $\Delta = -z(z + px + qy)$.

§ 2.2 Definition of a Contact Transformation in Space

A point, (x, y, z), in three dimensional Cartesian space, $\mathbb{R}^3$, together with a plane passing through the point, is called an **element**. Since the plane is determined by the point through which it passes, along with the direction coefficients, p, and q, the element is denoted by five coordinates, that is, a point (x, y, z, p, q) in a five dimensional space. A continuously differentiable, one-to-one transformation defined on a domain in $xyzpq$-space with range in $XYZPQ$-space, which may or may not coincide with the original space, given by the functions

$$(2.1) \quad \begin{cases} X = X(x, y, z, p, q), \quad Y = Y(x, y, z, p, q), \quad Z = Z(x, y, z, p, q) \\ P = P(x, y, z, p, q), \quad Q = Q(x, y, z, p, q) \end{cases}$$

is called an **element transformation**. We shall assume that both it, and its inverse, are sufficiently differentiable so that the computations below make sense, and that the Jacobian of the transformation

$$(2.2) \qquad \frac{\partial(X, Y, Z, P, Q)}{\partial(x, y, z, p, q)} \neq 0.$$

The definition of a contact transformation in $\mathbb{R}^3$ is more complicated than it is in $\mathbb{R}^2$, and there are several distinct cases which must be considered.

Let σ be an arbitrary surface in $\mathbb{R}^3$, which can be represented by three, twice continuously differentiable functions

$$(2.3) \qquad \sigma: \qquad x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v),$$

where u, and v, range over some set in the uv-parameter space.

Case (i). We consider first the case that (2.3) represents a two dimensional surface in $\mathbb{R}^3$. For this to be true, it is necessary and sufficient that at least one of the determinants

$$\frac{\partial(x,y)}{\partial(u,v)}, \qquad \frac{\partial(y,z)}{\partial(u,v)}, \qquad \frac{\partial(z,x)}{\partial(u,v)}, \tag{2.4}$$

be nonzero, that is, the rank of the matrix

$$\begin{bmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{bmatrix}$$

must be two.

Example 2.1. Suppose the first determinant in (2.4) does not vanish. Then we can solve for u, and v, in terms of x, y, in (2.3) and insert the result into the expression for z to obtain

$$z = z\big(u(x,y), v(x,y)\big) := f(x,y).$$

the standard expression for a surface.

Now let (x,y,z) be a point on σ, and $(-p,-q,1)$ be the components of the normal for the tangent plane to σ at this point. If σ is given by the equation $z = f(x,y)$, we have $p = \partial f/\partial x$, $q = \partial f/\partial y$; in which case we call (x,y,z,p,q) a **surface element**. Each surface element is characterized by the fact that

$$p\,dx + q\,dy - dz = 0 \tag{2.5}$$

at every point on the surface. In terms of the variables (u,v), equation (2.5) may be written equivalently as

$$px_u + qy_u - z_u = 0, \tag{2.6}$$

$$px_v + qy_v - z_v = 0, \tag{2.7}$$

obtained by considering (2.5) along the curves: $v =$ constant, and $u =$ constant, respectively.

A **strip** , on the surface σ, is determined by a curve γ on σ, where to each point (x,y,z) on γ is associated the tangent plane to σ at this point. See Figure 2.1.

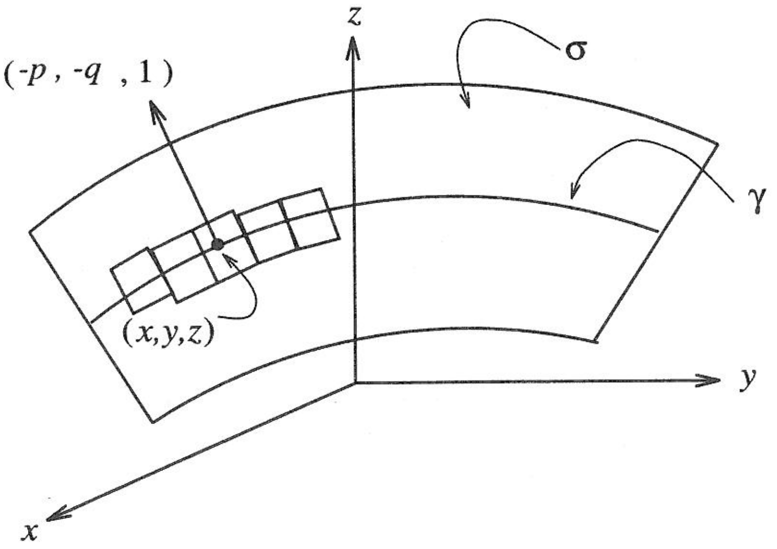

Figure 2.1

If γ is the image on a surface σ of the curve $u = u(t)$, $v = v(t)$, in the uv-plane, and

$$\dot{x} = \frac{\partial x}{\partial u}\frac{du}{dt} + \frac{\partial x}{\partial v}\frac{dv}{dt},$$

we obtain from (2.5) the so-called **strip equation**

$$p\dot{x} + q\dot{y} - \dot{z} = 0 \tag{2.8}$$

Clearly (2.6) and (2.7) represent the strip condition for the curves $u =$ constant, and $v =$ constant, respectively.

Case (ii). We now consider the case where the three determinants, (2.4), vanish identically in a domain D of the uv-plane, which is mapped by (2.3) into the xyz-space, but where not all of the six derivatives $x_u, \ldots, z_v$ vanish. In this case the three functions (2.3) map the domain D onto a curve in the xyz-space.

In order to prove this assertion, we may assume, without loss of generality, that $x_u \neq 0$. There are then two cases to consider.

First, let $x_v = 0$. Then since the determinants in (2.4) all vanish, we may conclude that $y_v = 0$, and $z_v = 0$. Thus $x(u,v)$, $y(u,v)$, $z(u,v)$, are independent of v, and we may write

$$x = x(u), \qquad y = y(u), \qquad z = z(u),$$

which are the parametric equations for a curve in $\mathbb{R}^3$.

Secondly, suppose that $x_v \neq 0$. Then since all the determinants in (2.4) vanish, the vectors (x_u, y_u, z_u), and (x_v, y_v, z_v), are linearly dependent. Thus there exists a function $\alpha = \alpha(u, v) \neq 0$ such that

$$x_u = \alpha x_v, \qquad y_u = \alpha y_v, \qquad z_u = \alpha z_v. \tag{2.9}$$

The idea now is to seek a new parameterization so that x, y, and z, depend not upon the two variables u, and v, but on simply one variable. Let us set

$$u = \phi(s, \tau), \qquad v = \psi(s, \tau) \tag{2.10}$$

and determine the functions ϕ, and ψ. In term of s, and τ,

$$x = x\big(\phi(s, \tau), \psi(s, \tau)\big). \tag{2.11}$$

We will show that under the assumption that $x_u \neq 0$, $x_v \neq 0$, it is possible to choose ϕ, and ψ, so that x is independent of τ, and simultaneously so are y, and z. To guarantee that, we must have

$$\frac{\partial x}{\partial \tau} = x_u \phi_\tau + x_v \psi_\tau = 0$$

whence by (2.9), and the fact that $x_v \neq 0$, we get

$$\alpha(\phi, \psi)\phi_\tau + \psi_\tau = 0, \tag{2.12}$$

where $\alpha = x_u / x_v$. Choose ϕ and ψ so that

$$\begin{aligned} \phi_\tau &= 1, \\ \psi_\tau &= -\alpha(\phi, \psi), \end{aligned} \tag{2.13}$$

and for initial conditions choose

$$\phi(s, 0) = s, \qquad \psi(s, 0) = \lambda(s),$$

where $\lambda(s)$ is any function chosen so that the Jacobian

$$\begin{aligned} \left.\frac{\partial(\phi, \psi)}{\partial(s, \tau)}\right|_{\tau=0} &= \begin{vmatrix} \phi_s & \phi_\tau \\ \psi_s & \psi_\tau \end{vmatrix}_{\tau=0} = \begin{vmatrix} 1 & 1 \\ \lambda' & -\alpha(\phi, \psi) \end{vmatrix}_{\tau=0} \\ &= \lambda' + \alpha\big(\phi(s, 0), \psi(s, 0)\big) \neq 0. \end{aligned} \tag{2.14}$$

This condition guarantees the independence of ϕ, and ψ. Such a λ can always be found. For example: λ equal to a constant will do since $\alpha \neq 0$. If, therefore, one sets

$$\phi(s, \tau) = s + \tau,$$

and determines $\psi(s,\tau)$ such that for each s

$$\psi_\tau(s,\tau) = -\alpha\big(s+\tau, \psi(s,\tau)\big), \qquad \psi(s,0) = \lambda(s),$$

where $\lambda(s)$ is chosen to satisfy (2.14), then

$$\begin{aligned} x &= x\big(\phi(s,\tau), \psi(s,\tau)\big) = \tilde{x}(s),\\ y &= y\big(\phi(s,\tau), \psi(s,\tau)\big) = \tilde{y}(s),\\ z &= z\big(\phi(s,\tau), \psi(s,\tau)\big) = \tilde{z}(s), \end{aligned}$$

which was the assertion.

REMARK: The assertion is intuitively obvious. For if $\big(u(t), v(t)\big)$, is any curve in the uv-plane passing through a point on the surface, say,

$$x\big(u(0), v(0)\big) = \xi, \qquad y\big(u(0), v(0)\big) = \eta, \qquad z\big(u(0), v(0)\big) = \zeta,$$

then by (2.9) the tangent vector at (ξ, η, ζ),

$$\frac{d}{dt}\begin{bmatrix} x\big(u(t), v(t)\big) \\ y\big(u(t), v(t)\big) \\ z\big(u(t), v(t)\big) \end{bmatrix}_{t=0} = (\alpha+1)\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}$$

always points in the same direction, so it would seem reasonable to expect that we could introduce coordinates such that one of them would move along the line in the direction of the tangent, and so describe the curve.

Case (iii). If all six derivatives $x_u, \ldots, z_v$ are identically zero in (u,v) then $(x,y,z) = (x_0, y_0, z_0)$ is a fixed point. Thus p and q can be given arbitrarily. The strip then consists of a bundle of planes passing through (x_0, y_0, z_0).

DEFINITION. A contact transformation in space is an element transformation which maps every strip of surface elements, one-to-one, onto a strip of surface elements. Alternatively, the element transformation (2.1) is a contact transformation if and only if it maps some domain D in $xyzpq$-space, one-to-one, onto a domain D' in $XYZPQ$-space, (2.2) holds, and

(2.15) $$p\,dx + q\,dy - dz = 0 \quad \text{implies} \quad P\,dX + Q\,dY - dZ = 0.$$

Theorem 2.1. *Equation (2.1) represents a contact transformation, if and only if, there is a function* $\rho = \rho(x,y,z,p,q) \neq 0$ *such that*

(2.16) $$P\,dX + Q\,dY - dZ = \rho(p\,dx + q\,dy - dz)$$

PROOF. If a function $\rho \neq 0$ exists so that (2.16) is satisfied, then the form $P\,dX + Q\,dY - dZ$ vanishes if and only if $p\,dx + q\,dy - dz$ does. (2.1) is therefore a contact transformation.

Assume now that (2.1) is a contact transformation and that $p\,dx+q\,dy-dz = 0$. Then since

$$\begin{aligned} dX &= X_x\,dx + X_y\,dy + X_z\,dz + X_p\,dp + X_q\,dq, \\ dY &= Y_x\,dx + Y_y\,dy + Y_z\,dz + Y_p\,dp + Y_q\,dq, \\ dZ &= Z_x\,dx + Z_y\,dy + Z_z\,dz + Z_p\,dp + Z_q\,dq, \end{aligned}$$

we see that

(2.17)

$$\begin{aligned} P\,dX + Q\,dY - dZ =& (PX_x + QY_x - Z_x)\,dx + (PX_y + QY_y - Z_y)\,dy \\ &+ (PX_z + QY_z - Z_z)\,dz + (PX_p + QY_p - Z_p)\,dp \\ &+ (PX_q + QY_q - Z_q)\,dq. \end{aligned}$$

When $dz = p\,dx + q\,dy$, the left hand side must vanish, that is we must have

$$\begin{aligned} &\{(PX_x + QY_x - Z_x) + p(PX_z + QY_z - Z_z)\}\,dx \\ &+ \{(PX_y + QY_y - Z_y) + q(PX_z + QY_z - Z_z)\}\,dy \\ &+ (PX_p + QY_p - Z_p)\,dp + (PX_q + QY_q - Z_q)\,dq = 0. \end{aligned}$$

This identity must hold for all (dx, dy, dz, dp, dq) so that if we set

$$\rho = -(PX_z + QY_z - Z_z), \tag{2.18}$$

we obtain the four additional relations

$$\begin{cases} PX_x + QY_x - Z_x = \rho p, \\ PX_y + QY_y - Z_y = \rho q, \end{cases} \tag{2.19}$$

$$\begin{cases} PX_p + QY_p - Z_p = 0, \\ PX_q + QY_q - Z_q = 0. \end{cases} \tag{2.20}$$

If we insert expressions (2.18), (2.19), and (2.20), into (2.17), we find that

$$(P\,dX + Q\,dY - dZ) = \rho(p\,dx + q\,dy - dz).$$

It remains only to show that $\rho \neq 0$. By way of contradiction, suppose $\rho = 0$. Then we could write (2.18), (2.19), and (2.20), in the form

$$P\begin{bmatrix} X_x \\ X_y \\ X_z \\ X_p \\ X_q \end{bmatrix} + Q\begin{bmatrix} Y_x \\ Y_y \\ Y_z \\ Y_p \\ Y_q \end{bmatrix} - \begin{bmatrix} Z_x \\ Z_y \\ Z_z \\ Z_p \\ Z_q \end{bmatrix} = 0,$$

and obviously $P^2 + Q^2 + 1 \neq 0$. Consequently, the row vectors

$$\begin{bmatrix} X_x \\ X_y \\ X_z \\ X_p \\ X_q \end{bmatrix} \qquad \begin{bmatrix} Y_x \\ Y_y \\ Y_z \\ Y_p \\ Y_q \end{bmatrix} \qquad \begin{bmatrix} Z_x \\ Z_y \\ Z_z \\ Z_p \\ Z_q \end{bmatrix}$$

are linearly dependent, so that the Jacobian

$$\frac{\partial(X, Y, Z, P, Q)}{\partial(x, y, z, p, q)} = 0$$

everywhere, contradicting our initial assumption that (2.1) is a contact transformation. □

Example 2.2. The Legendre transformation in space

$$X = p, \qquad Y = q, \qquad Z = px + qy - z, \qquad P = x, \qquad Q = y,$$

is a contact transformation, with $\rho = -1$.

We have seen that in the plane, if two independent functions are given, they can be extended to form a contact transformation. The situation in three space is similar. Let us suppose that

$$X = X(x, y, z), \qquad Y = Y(x, y, z), \qquad Z = Z(x, y, z), \tag{2.21}$$

are given, and are independent. The equations (2.18), (2.19), and (2.20), can be used to extend this transformation to a contact transformation. To see how this is done, note first that (2.20) is trivially satisfied. Since the functions X, Y, Z, are independent, (2.18) and (2.19) can be solved for P/ρ, Q/ρ, $-1/\rho$, and from that, P and Q can be determined explicitly. Alternatively, eliminate ρ, and solve for P and Q. One obtains compatibility conditions which the functions in (2.20) must satisfy, which are analogous to those we found in Chapter I. In Chapter III, we shall treat these questions in more generality. At the moment, we remark only that while the program outlined above could be carried out, at this stage it is very tedious.

§ 2.3 The Directrix Equations

We have seen that a contact transformation

(3.1)
$$\begin{cases} X = X(x, y, z, p, q), \qquad Y = Y(x, y, z, p, q), \qquad Z = Z(x, y, z, p, q) \\ P = P(x, y, z, p, q), \qquad Q = Q(x, y, z, p, q) \end{cases}$$

is a one-to-one transformation from some domain of the $xyzpq$-space, onto a domain of the $XYZPQ$-space, which is characterized by the fact that there is a function $\rho = \rho(x, y, z, p, q) \neq 0$, such that

$$P\,dX + Q\,dY - dZ = \rho(p\,dx + q\,dy - dz). \tag{3.2}$$

By expanding the differentials dX, dY, dZ, and comparing coefficients, we have seen that (3.2) is equivalent to the system

$$\begin{cases} PX_x + QY_x - Z_x = \rho p \\ PX_y + QY_y - Z_y = \rho q \\ PX_z + QY_z - Z_z = -\rho \\ PX_p + QY_p - Z_p = 0 \\ PX_q + QY_q - Z_q = 0. \end{cases} \tag{3.3}$$

The system (3.3) can be regarded as an over determined, linear system for the quantities P/ρ, Q/ρ, $-1/\rho$, so that (3.3) implicitly contains compatibility conditions which the functions $X(x, y, z, p, q)$, $Y(x, y, z, p, q)$, and $Z(x, y, z, p, q)$, must satisfy if they are to be the first three functions of a contact transformation. Furthermore, we know from the remarks at the end of §2.2 that once these functions are given, then $P(x, y, z, p, q)$ and $Q(x, y, z, p, q)$ are determined. In fact we shall see that contact transformations can be constructed from equations which are independent of p, and q, the so called directrix equations, just as was done in §1.4.

To begin with, suppose the functions X, Y, Z, in equation (3.1) are independent of p and q. Then X, Y, Z, satisfy a system of the form

$$\begin{cases} f(X, Y, Z; x, y, z) = 0 \\ g(X, Y, Z; x, y, z) = 0 \\ h(X, Y, Z; x, y, z) = 0. \end{cases} \tag{3.4}$$

Conversely, let us suppose that we have been given three functions of the form (3.4). We ask when we can use these functions to determine a contact transformation. In the event that they determine one (or more) contact transformations, they are called the directrix equations for the contact transformation(s).

Guided by the development in §1.4, we shall assume that

$$\frac{\partial(f, g, h)}{\partial(x, y, z)} \neq 0, \qquad \frac{\partial(f, g, h)}{\partial(X, Y, Z)} \neq 0, \tag{3.5}$$

in some domain. Consequently, (3.4) can be solved for X, Y, Z in terms of x, y, z, at least locally, and

$$\frac{\partial(X,Y,Z)}{\partial(x,y,z)} \neq 0. \tag{3.6}$$

Once these functions are known, we set up the system

$$\begin{cases} PX_x + QY_x - Z_x = \rho p, \\ PX_y + QY_y - Z_y = \rho q, \\ PX_z + QY_z - Z_z = -\rho, \end{cases} \tag{3.7}$$

for the unknowns P, Q, and ρ. Now $\rho \neq 0$, for if $\rho = 0$, we could conclude by (3.6) that $(P, Q, -1) = (0, 0, 0)$, which is impossible. Thus, (3.7) can be regarded as a linear system for the unknowns P/ρ, Q/ρ, and $-1/\rho$, and solved. Thus, we can determine functions $X(x,y,z)$, $Y(x,y,z)$, $Z(x,y,z)$, $P(x,y,z,p,q)$, $Q(x,y,z,p,q)$, and $\rho(x,y,z,p,q)$ such that (3.2) is satisfied. We have to show that the resulting transformation is one-to-one. To see that, note first that

$$\frac{\partial(X,Y,Z,P,Q)}{\partial(x,y,z,p,q)} = \left(\frac{\partial(X,Y,Z)}{\partial(x,y,z)}\right)\left(\frac{\partial(P,Q)}{\partial(p,q)}\right). \tag{3.8}$$

We must show that

$$\frac{\partial(X,Y,Z,P,Q)}{\partial(x,y,z,p,q)} \neq 0.$$

By (3.6) and (3.8), it suffices to show that

$$\frac{\partial(P,Q)}{\partial(p,q)} \neq 0.$$

To investigate that, let us first differentiate (3.7) with respect to p, and q, to obtain

$$\begin{cases} P_pX_x + Q_pY_x = \rho + p\rho_p, \\ P_pX_y + Q_pY_y = q\rho_p, \\ P_pX_z + Q_pY_z = -\rho_p, \end{cases} \quad \text{and} \quad \begin{cases} P_qX_x + Q_qY_x = p\rho_q, \\ P_qX_y + Q_qY_y = \rho + q\rho_q, \\ P_qX_z + Q_qY_z = -\rho_q. \end{cases} \tag{3.9}$$

Now suppose

$$\frac{\partial(P,Q)}{\partial(p,q)} = 0.$$

Then the vectors

$$\begin{bmatrix} P_p \\ Q_p \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_q \\ Q_q \\ 0 \end{bmatrix}$$

are linearly dependent, and we can say

$$(3.10)\qquad \begin{bmatrix} P_p \\ Q_p \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} P_q \\ Q_q \\ 0 \end{bmatrix}, \qquad \text{for some } \lambda \neq 0.$$

Multiply the second set of equations in (3.9) by λ, and use (3.10) to conclude that

$$\rho + p\rho_p = \lambda q\rho_q, \qquad q\rho_p = \lambda(\rho + q\rho_q), \qquad \rho_p = \lambda\rho_q,$$

and consequently

$$q\rho_p = \lambda\rho + \lambda q\rho_q = \lambda\rho + q\rho_p.$$

That is

$$\lambda\rho = 0, \quad \text{or } \rho = 0, \quad \text{since } \lambda \neq 0,$$

which is a contradiction since $\rho \neq 0$. Therefore,

$$\frac{\partial(P,Q)}{\partial(p,q)} \neq 0,$$

and the determinant in (3.8) is nonzero. We have proved the following theorem:

Theorem 3.1. *Suppose the functions f, g, and h, satisfy (3.5). Then they are the directrix equations for a contact transformation. More specifically, the solutions $X = X(x,y,z)$, $Y = Y(x,y,z)$, $Z = Z(x,y,z)$, to the system in (3.4):*

$$\begin{cases} f(X,Y,Z;x,y,z) = 0 \\ g(X,Y,Z;x,y,z) = 0 \\ h(X,Y,Z;x,y,z) = 0. \end{cases}$$

are independent functions, which can be extended to a contact transformation.

Next let us suppose that the first three functions X, Y, Z, of a contact transformation are given. Suppose further that the rank of the matrix

$$\begin{bmatrix} X_p & Y_p & Z_p \\ X_q & Y_q & Z_q \end{bmatrix}$$

is one. We have seen in §2.2, that in this case a new parameter, u can be introduced so that X, Y, Z, can be regarded as functions of the variables

(x, y, z, u). If u is eliminated from the system, we obtain two equations of the form

$$\text{(3.11)} \qquad \begin{cases} f(X, Y, Z; x, y, z) = 0 \\ g(X, Y, Z; x, y, z) = 0 \end{cases}$$

which the contact transformation has to satisfy. In this case, the contact transformation satisfies two directrix equations.

Now let us turn to the converse question of determining when two functions, f, and g, of the variables $(X, Y, Z; x, y, z)$, represent the directrix functions for a contact transformation. To see when this is possible, we calculate the differentials

$$\text{(3.12)} \qquad \begin{cases} df = f_X\, dX + f_Y\, dY + f_Z\, dZ + f_x\, dx + f_y\, dy + f_z\, dz = 0 \\ dg = g_X\, dX + g_Y\, dY + g_Z\, dZ + g_x\, dx + g_y\, dy + g_z\, dz = 0. \end{cases}$$

If X, Y, Z, P, Q are to represent a contact transformation, the differentials must also satisfy

$$\text{(3.13)} \qquad P\, dX + Q\, dY - dZ - \rho(p\, dx + q\, dy - dz) = 0.$$

There must exist parameters λ, μ, such that

$$\text{(3.14)} \qquad \begin{cases} \lambda f_X + \mu g_X = P & \lambda f_x + \mu g_x = -\rho p \\ \lambda f_Y + \mu g_Y = Q & \lambda f_y + \mu g_y = -\rho q \\ \lambda f_Z + \mu g_Z = -1 & \lambda f_z + \mu g_z = \rho \\ f(X, Y, Z; x, y, z) = 0 & g(X, Y, Z; x, y, z) = 0 \end{cases}$$

The parameter ρ can be eliminated from this system, and we obtain the homogeneous linear system

$$\text{(3.15)} \qquad \begin{cases} \lambda(f_x + pf_z) + \mu(g_x + pg_z) = 0, \\ \lambda(f_y + qf_z) + \mu(g_y + qg_z) = 0. \end{cases}$$

Obviously, not both λ, and μ, can vanish. Consequently, the determinant of the coefficients must vanish, that is,

(3.16)

$$h(X, Y, Z; x, y, z, p, q) \equiv (f_x + pf_z)(g_y + qg_z) - (g_x + pg_z)(f_y + qf_z) = 0$$

We arrive at the following system for the construction of X, Y, Z,

$$\text{(3.17)} \qquad \begin{aligned} f(X, Y, Z; x, y, z) &= 0, \\ g(X, Y, Z; x, y, z) &= 0, \\ h(X, Y, Z; x, y, z, p, q) &= 0. \end{aligned}$$

We must now give sufficient conditions which will allow us to solve (3.17) for X, Y, and Z, in terms of (x,y,z,p,q), and at the same time allow us to conclude that these functions, together with the P and Q obtained from (3.14), form a contact transformation.

First, in order to be able to solve for X, Y, and Z we demand that

$$\frac{\partial(f,g,h)}{\partial(x,y,z)} \neq 0. \tag{3.18}$$

(3.18) allows us to construct the functions

(3.19)

$$X = X(x,y,z,p,q), \qquad Y = Y(x,y,z,p,q), \qquad Z = Z(x,y,z,p,q).$$

Now recall that the case of the two directrix equations occurred when X, Y, and Z, were not independent of both p and q. Since p and q occur only in the equation for h, we first demand that at least one of the derivatives, $h_p \neq 0$ or $h_q \neq 0$. In other words,

$$\text{either } \frac{\partial(f,g)}{\partial(x,z)} \neq 0, \quad \text{or } \frac{\partial(f,g)}{\partial(y,z)} \neq 0, \quad \text{or both are nonzero.} \tag{3.20}$$

Equation (3.20) implies that $\rho \neq 0$. For if ρ were zero and say

$$\frac{\partial(f,g)}{\partial(x,z)} \neq 0,$$

we would conclude from (3.14), that both λ and μ would be zero, which we have already precluded. (3.20) also implies that not all of the functions, $f_x + pf_z$, $g_x + pg_z$, $f_y + qf_z$, $g_y + qg_z$, can vanish identically in the region where X, Y, Z, are to be constructed. For if, say

$$\frac{\partial(f,g)}{\partial(x,z)} \neq 0 \quad \text{there, and if} \quad \begin{aligned} f_x + pf_z &= 0, \\ g_x + pg_z &= 0, \end{aligned}$$

then we could infer that $(1,p) = (0,0)$, which is impossible. Hence, we can construct from (3.15), the ratio λ/μ, if both $f_x + pf_z$, and $g_x + pg_z$, are nonzero; or if perhaps $f_x + pf_z = 0$, and $g_x + pg_z \neq 0$, then λ can be arbitrary and $\mu = 0$, and so on. In any case, we obtain from (3.14) the expressions

$$\begin{cases} \rho = -\dfrac{\lambda f_z + \mu g_z}{\lambda f_Z + \mu g_Z} \\ P = -\dfrac{\lambda f_X + \mu g_X}{\lambda f_Z + \mu g_Z} \\ Q = -\dfrac{\lambda f_Y + \mu g_Y}{\lambda f_Z + \mu g_Z}. \end{cases} \tag{3.21}$$

The functions (3.19), and (3.21), satisfy (3.2). To prove this assertion, note that since X, Y, Z, satisfy $f(X,Y,Z) = 0$, and $g(X,Y,Z) = 0$, and

$$\begin{aligned} 0 = \lambda\, df + \mu\, dg = (\lambda f_X + \mu g_X)\, dX + (\lambda f_Y + \mu g_Y)\, dY + (\lambda f_Z + \mu g_Z)\, dZ \\ +(\lambda f_x + \mu g_x)\, dx + (\lambda f_y + \mu g_y)\, dy + (\lambda f_z + \mu g_z)\, dz. \end{aligned}$$

Now insert the expressions from (3.14), to conclude that the identity holds.

Finally, if we can show that

$$\frac{\partial(X,Y,Z,P,Q)}{\partial(x,y,z,p,q)} \neq 0, \tag{3.22}$$

then we shall be able to conclude that X, Y, Z, P, Q, is a contact transformation.

In order to make this conclusion, we need additional hypotheses. To motivate these assumptions, let us note that if (3.1) is a contact transformation, then so is its inverse,

$$\begin{cases} x = x(X,Y,Z,P,Q), \qquad y = y(X,Y,Z,P,Q), \qquad z = z(X,Y,Z,P,Q) \\ p = p(X,Y,Z,P,Q), \qquad q = q(X,Y,Z,P,Q). \end{cases}$$

We would expect, therefore, that the hypotheses for the X, Y, Z, P, Q, variables should be the same as those for the x, y, z, p, q, variables.

We construct the function $H = H(X,Y,Z,P,Q;x,y,z)$ as we did the function h, to find

$$H(X,Y,Z,P,Q;x,y,z) = (f_X + Pf_Z)(g_Y + Qg_Z) - (g_X + Pg_Z)(f_Y + Qf_Z). \tag{3.23}$$

Now assume that

$$\frac{\partial(f,g,H)}{\partial(x,y,z)} \neq 0, \tag{3.24}$$

and at least one of the determinants

$$\frac{\partial(f,g)}{\partial(X,Z)}, \qquad \frac{\partial(f,g)}{\partial(Y,Z)} \tag{3.25}$$

is nonzero. We now verify immediately that

$$\frac{\partial\left(f,g,h,P + \dfrac{\lambda f_X + \mu g_X}{\lambda f_Z + \mu g_Z}, Q + \dfrac{\lambda f_Y + \mu g_Y}{\lambda f_Z + \mu g_Z}\right)}{\partial(X,Y,Z,P,Q)} = \frac{\partial(f,g,h)}{\partial(X,Y,Z)} \neq 0,$$

and similarly that,

$$\frac{\partial\left(f, g, H, p+\dfrac{\lambda f_x+\mu g_x}{\lambda f_z+\mu g_z}, q+\dfrac{\lambda f_y+\mu g_y}{\lambda f_z+\mu g_z}\right)}{\partial(x, y, z, p, q)}=\frac{\partial(f, g, H)}{\partial(x, y, z)} \neq 0,$$

so that the transformation (3.1) has an inverse, and thus the Jacobian is nonzero. We summarize this discussion as a theorem.

Theorem 3.2. *Suppose f and g are functions of the variables $(X, Y, Z; x, y, z)$. Let h be defined by (3.16), and H by (3.23), and suppose that (3.18), (3.20), (3.24), and (3.25) hold. Then the solution X, Y, Z, to the (3.17) system, together with the functions P, Q, ρ, given by (3.21), define a contact transformation.*

Example 3.1. Suppose both f and g are independent of at least one of the variables X, Y, or Z. Suppose for definiteness that $f_X = g_X = 0$. Then the solution X, Y, Z, to the (3.17) system, is not a contact transformation.

Example 3.2. Suppose

$$\begin{aligned} f(X, Y, Z; x, y, z) &= z + Z + xX, \\ g(X, Y, Z; x, y, z) &= y - Y. \end{aligned}$$

Note that

$$\frac{\partial(f,g)}{\partial(x,z)} = \begin{vmatrix} X & 1 \\ 0 & 0 \end{vmatrix} = 0, \qquad \frac{\partial(f,g)}{\partial(y,z)} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1,$$

$$\frac{\partial(f,g)}{\partial(X,Z)} = \begin{vmatrix} x & 1 \\ 0 & 0 \end{vmatrix} = 0, \qquad \frac{\partial(f,g)}{\partial(Y,Z)} = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = 1.$$

Construct the function

$$h = (X + p)(1) - (0)(0 + q1) = X + p,$$

and note that

$$\frac{\partial(f,g,h)}{\partial(X,Y,Z)} = \begin{vmatrix} x & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{vmatrix} = 1.$$

The solution to the three equations

$$\begin{aligned} z + Z + xX &= 0 \\ y - Y &= 0 \\ X + p &= 0 \end{aligned}$$

is

$$X = -p, \qquad Y = y, \qquad Z = px - z.$$

The (3.15) equations are

$$\lambda(X + p) + \mu \cdot 0 = 0$$
$$\lambda \cdot q + \mu \cdot 1 = 0.$$

Since $X + p = 0$, we find that $\mu/\lambda = -q$. Thus, we see from (3.21) that

$$\rho = -1, \qquad P = -x, \qquad Q = -q$$

so that the transformation is

$$\begin{cases} X = -p, \qquad Y = y, \qquad Z = px - z, \qquad P = -x, \qquad Q = -q \\ (PdX + QdY - dZ) = -(pdx + qdy - dz) \end{cases}$$

and

$$\frac{\partial(X, Y, Z, P, Q)}{\partial(x, y, z, p, q)} = -1.$$

The inverse transformation is given by

$$x = -P, \qquad y = Y, \qquad z = PX - Z, \qquad p = -X, \qquad q = -Q.$$

This transformation is sometimes called the *Euler - Ampère* transformation.

Example 3.3. Similarly, one can find the contact transformation generated by the two directrix equations

$$f(X, Y, Z; x, y, z) = z + Z + xX + yY,$$
$$g(X, Y, Z; x, y, z) = x + X + y + Y.$$

It is

$$\begin{aligned} X &= \frac{1}{2}[(q - p) - (x + y)] \\ Y &= -\frac{1}{2}[(q - p) + (x + y)] \\ Z &= -\frac{x}{2}[(q - p) - (x + y)] + \frac{y}{2}[(q - p) + (x + y)] - z \\ P &= p - x + \frac{1}{2}[(q - p) - (x + y)] \\ Q &= q - y + \frac{1}{2}[(q - p) + (x + y)] \\ \rho &= -1 \end{aligned}$$

Let us again return to the contact transformation (3.1) and consider the first three functions $X(x, y, z, p, q)$, $Y(x, y, z, p, q)$, $Z(x, y, z, p, q)$. Suppose the rank of the matrix

$$\begin{bmatrix} X_p & Y_p & Z_p \\ X_q & Y_q & Z_q \end{bmatrix} \tag{3.26}$$

is two. Then two of the equations can be solved for p and q, in terms of the other variables, and the result inserted into the third equation. This leads to the single relationship

$$\Omega(X, Y, Z; x, y, z) = 0, \tag{3.27}$$

among the variables X, Y, Z, x, y, z. In other words, if (3.1) is a contact transformation, and the matrix (3.26) has rank two, the variables X, Y, Z, x, y, z, satisfy a single directrix equation.

We now ask the reverse question. Suppose one has a function Ω which is dependent on the variables (X, Y, Z, x, y, z). When is it the directrix for a contact transformation? The derivation is nearly identical to the previously encountered, two dimensional case, so we can be brief. The differential of Ω is given by

$$d\Omega = \Omega_X\,dX + \Omega_Y\,dY + \Omega_Z\,dZ + \Omega_x\,dx + \Omega_y\,dy + \Omega_z\,dz \tag{3.28}$$

and since it must be proportional to

$$0 = P\,dX + Q\,dY - dZ - \rho p\,dx - \rho p\,dy + \rho\,dz,$$

we see that

$$\begin{cases} P = \lambda\Omega_X, & Q = \lambda\Omega_Y, & -1 = \lambda\Omega_Z, \\ \rho p = -\lambda\Omega_x, & \rho q = -\lambda\Omega_y, & \rho = \lambda\Omega_z. \end{cases} \tag{3.29}$$

From (3.29), we see that $\lambda \neq 0$, and $\Omega_Z \neq 0$, so that

$$\lambda = -1/\Omega_Z. \tag{3.30}$$

Since ρ cannot be zero, we conclude that $\Omega_z \neq 0$, so that

$$\rho = -\Omega_z/\Omega_Z. \tag{3.31}$$

We are led to the system of equations

$$\begin{cases} \Omega(X, Y, Z; x, y, z) = 0 \\ \Omega_X + P\Omega_Z = 0, \quad \Omega_Y + Q\Omega_Z = 0, \\ \Omega_x + p\Omega_z = 0, \quad \Omega_y + q\Omega_z = 0. \end{cases} \tag{3.32}$$

The idea is to use the first and third equations of (3.32) to solve for X, Y, Z, and use the second equations to construct P, and Q. We require, therefore, that

(3.33)
$$\frac{\partial(\Omega, \Omega_x + p\Omega_z, \Omega_y + q\Omega_z)}{\partial(X, Y, Z)} \neq 0, \qquad \frac{\partial(\Omega, \Omega_X + P\Omega_Z, \Omega_Y + Q\Omega_Z)}{\partial(x, y, z)} \neq 0.$$

By arguing as in the case of Theorem 3.2, we can prove immediately

Theorem 3.3. *Suppose the function* $\Omega = \Omega(X, Y, Z; x, y, z)$ *satisfies* $\Omega_z \neq 0$, *and* $\Omega_Z \neq 0$. *Suppose further that equation (3.33) holds. Then the equation* $\Omega = 0$, *represents the directrix equation for a contact transformation. That is, the equations (3.32) can be used to construct functions* X, Y, Z, P, Q, *and these functions represent a contact transformation, with* ρ *given by (3.31).*

Example 3.4. Suppose

$$\Omega(X, Y, Z; x, y, z) = xX + yY - (z + Z).$$

Then obviously $\Omega_z = \Omega_Z = -1$, and are thus both nonzero, as is ρ, i.e. $\rho = -1$. The system (3.32) takes the form

$$\begin{cases} xX + yY - (z + Z) = 0 \\ x - P = 0, \qquad y - Q = 0, \\ X - p = 0, \qquad Y - q = 0. \end{cases}$$

The contact transformation is

$$X = p, \qquad Y = q, \qquad Z = px + qy - z, \qquad P = x, \qquad Q = y,$$

with the inverse transformation given by

$$x = P, \qquad y = Q, \qquad z = PX + QY - Z, \qquad p = X, \qquad q = Y.$$

This is the *Legendre* transformation in space.

Example 3.5. The contact transformation whose directrix equation is

$$\Omega(X, Y, Z; x, y, z) = xX + yY + zZ - 1 = 0$$

is given by

$$\begin{cases} X = -p/(z - px - qy) \\ Y = -q/(z - px - qy) \\ Z = 1/(z - px - qy) \\ P = -x/z \\ Q = -y/z \end{cases} \quad \text{with inverse} \quad \begin{cases} x = -P/(Z - PX - QY) \\ y = -Q/(Z - PX - QY) \\ z = 1/(Z - PX - QY) \\ p = -X/Z \\ q = -Y/Z \end{cases}$$

§ 2.4 Geometric Theory of Partial Differential Equations

Let $z = f(x,y)$ denote an unknown function having continuous partial derivatives

$$z_x = \frac{\partial f}{\partial x}, \qquad z_y = \frac{\partial f}{\partial y}$$

in some domain of the xy-plane. A partial differential equation of the first order is a functional relationship between the function z, its first order partial derivatives z_x, and z_y, and the independent variables x, and y,

$$\phi(x, y, z, z_x, z_y) = 0. \tag{4.1}$$

If $z = f(x,y)$ is a surface, σ, then $p = z_x$ and $q = z_y$ are the direction coefficients of the normal to σ at the point (x, y, z), so that (4.1) can be regarded at the same time as a condition on the elements (x, y, z, p, q)

$$\phi(x, y, z, p, q) = 0. \tag{4.2}$$

We shall assume that $\phi_p^2 + \phi_q^2 \neq 0$. An element $(x_0, y_0, z_0, p_0, q_0)$ which satisfies (4.2) is called an **integral element** for the differential equation. Let us now solve (4.2) for q to obtain

$$q = \psi(x, y, z, p). \tag{4.3}$$

Let (x_0, y_0, z_0) be a fixed point on σ. Then we obtain a one parameter family of elements

$$\big(x_0, y_0, z_0, p, \psi(x_0, y_0, z_0, p)\big).$$

As p varies, we obtain a set of elements which generate a cone, the *Monge cone*, with vertex (x_0, y_0, z_0). See Figure 4.1.

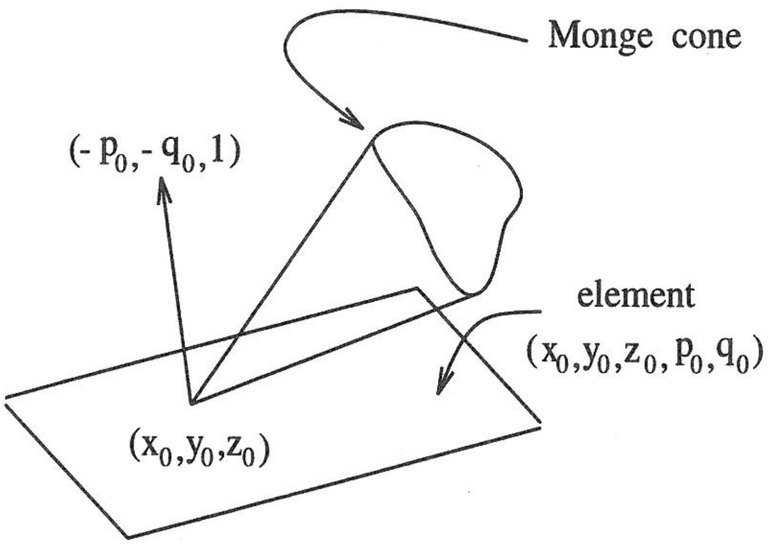

Figure 4.1

The surface $z = f(x, y)$ is a solution to (4.1) if at each point $(x_0, y_0, f(x_0, y_0))$ the tangent plane determines an integral element

$$\big(x_0, y_0, f(x_0, y_0), f_x(x_0, y_0), f_y(x_0, y_0)\big).$$

Geometrically, it is clear from this discussion that a surface will be an integral surface if and only if at each point, its tangent element should touch the Monge cone of the equation.

Now let us take a curve in the surface which is tangent to the generators of the cones. If

$$x = x(\tau), \qquad y = y(\tau), \qquad z(\tau) = z\big(x(\tau), y(\tau)\big), \tag{4.4}$$

is a parametric description of the curve, then

$$\dot z = z_x \dot x + z_y \dot y = p\dot x + q\dot y \tag{4.5}$$

Moreover, the five functions $\big(x(\tau), y(\tau), z(\tau), p(\tau), q(\tau)\big)$ must form an integral strip, that is

$$F\big(x(\tau), y(\tau), z(\tau), p(\tau), q(\tau)\big) = 0. \tag{4.6}$$

The envelope condition is obtained from (4.5) by differentiating with respect to p and is

$$\dot x + \frac{\partial q}{\partial p}\dot y = 0, \tag{4.7}$$

or from (4.2)

$$\frac{\partial \phi}{\partial p} + \frac{\partial \phi}{\partial q}\frac{\partial q}{\partial p} = 0,$$

whence from (4.7)

$$\frac{\partial \phi}{\partial p}\dot y - \frac{\partial \phi}{\partial q}\dot x = 0. \tag{4.8}$$

(4.8) together with (4.5) implies the validity of the ratios

$$\dot x : \dot y : \dot z = \frac{\partial \phi}{\partial p} : \frac{\partial \phi}{\partial q} : \left(p\frac{\partial \phi}{\partial p} + q\frac{\partial \phi}{\partial q}\right),$$

whence by adjusting, if necessary, the parameter τ,

$$\dot x = \frac{\partial \phi}{\partial p}, \qquad \dot y = \frac{\partial \phi}{\partial q}, \qquad \dot z = p\,\frac{\partial \phi}{\partial p} + q\,\frac{\partial \phi}{\partial q}. \tag{4.9}$$

Differentiating equation (4.2) with respect to x and y, we find

$$\phi_x + \phi_z p + \phi_p p_x + \phi_q q_x = 0,$$
$$\phi_y + \phi_z q + \phi_p p_y + \phi_q q_y = 0.$$

By (4.9) and the fact that $q_x = p_y$, we obtain

$$\begin{aligned} \frac{d}{d\tau} p &= \frac{d}{d\tau} z_x(x,y) = p_x \dot{x} + p_y \dot{y} = p_x \frac{\partial \phi}{\partial p} + q_x \frac{\partial \phi}{\partial q} = -(\phi_x + p\phi_z), \\ \frac{d}{d\tau} q &= \frac{d}{d\tau} z_y(x,y) = q_x \dot{x} + q_y \dot{y} = p_y \frac{\partial \phi}{\partial p} + q_y \frac{\partial \phi}{\partial q} = -(\phi_y + q\phi_z). \end{aligned} \tag{4.10}$$

The formulas (4.9) and (4.10) combined give

$$\begin{aligned} &\dot{x} = \phi_p, \qquad \dot{y} = \phi_q, \qquad \dot{z} = p\phi_p + q\phi_q, \\ &\dot{p} = -(\phi_x + p\phi_z), \qquad \dot{q} = -(\phi_y + q\phi_z), \end{aligned} \tag{4.11}$$

where the dot represents differentiation with respect to τ.

The directions of the generators of the Monge cone are called **characteristic directions** and the equations (4.11) are called the **characteristic equations**.

The solution to this set of ordinary differential equations gives the surface in the parametric form

$$x = x(s,\tau),$$
$$y = y(s,\tau),$$
$$z = z(s,\tau),$$

where s is the curve parameter for an initial curve $\big(x_0(s), y_0(s), z_0(s)\big)$. The initial values for p and q are obtained by solving

$$z_0'(s) = p_0(s)\, x_0'(s) + q_0(s)\, y_0'(s)$$

and

$$\phi\big(x_0(s), y_0(s), z_0(s), p_0(s), q_0(s)\big) = 0$$

for p and q. From the geometric derivation, it is obvious that the characteristic equations (4.11) can be solved, and a nontrivial solution surface constructed if the initial strip $\big(x_0(s), y_0(s), z_0(s), p_0(s), q_0(s)\big)$ is not itself a solution to (4.11).

The characteristic equations themselves can be difficult to treat, and just as in the case of ordinary differential equations, it is sometimes possible to transform the original, first order partial differential equation using a contact transformation.

Example 4.1. The characteristic equations for the differential equation

$$\phi(x, y, z, z_x, z_y) = xz_x + yz_y + x + y - z = 0$$

are

$$\begin{aligned} dx/d\tau &= x \\ dy/d\tau &= y \\ dz/d\tau &= px + qy \\ dp/d\tau &= -1 \\ dq/d\tau &= -1, \end{aligned}$$

which is a non-linear system.

Using the Legendre transformation, we find that the partial differential equation transforms to the linear equation

$$\phi(X, Y, Z, P, Q) = P + Q + Z = 0.$$

The characteristic equations are

$$\begin{aligned} dX/d\tau &= 1 \\ dY/d\tau &= 1 \\ dZ/d\tau &= P + Q \\ dP/d\tau &= -P \\ dQ/d\tau &= -Q. \end{aligned}$$

Note that the last two equations for P and Q can be integrated immediately once X and Y are known. The three equations are essentially linear. They can be solved subject to appropriate initial conditions, then using the inverse transformation the solution to the original problem can be obtained. As a rule, such a contact transformation is difficult to find.

§ 2.5 Further Simple Examples of Contact Transformations

In the plane, we saw that the polar equation for a conic section led to a directrix equation from which simple contact transformations could be obtained. We can also follow such a procedure in constructing contact transformations in space. However, in order to avoid the introduction of ten (more or less) arbitrary constant coefficients, we restrict our discussion to the polarity with respect to the unit sphere

$$x^2 + y^2 + z^2 - 1 = 0,$$

so that the directrix equation is

$$\Omega(X, Y, Z; x, y, z) = xX + yY + zZ - 1 = 0. \tag{5.1}$$

If we regard (X, Y, Z) as the running coordinates, then (5.1) is the equation of the **polar plane** Σ_{xyz}, associated with the **pole** (x, y, z), by the polarity with respect to the sphere. If (x, y, z) is a point on the sphere, the plane Σ_{xyz} is the tangent plane to the sphere at this point. In no other case is the pole a point of the polar plane. For an arbitrary pole on the sphere, the polar plane can be constructed geometrically as it has been described for the polar line with respect to a conic in the plane, (cf. §1.5).

By means of the equations (3.32), the contact transformation is obtained by solving (5.1) together with

$$\begin{cases} x + Pz = 0, & y + Qz = 0, \\ X + pZ = 0, & Y + qZ = 0, \end{cases} \tag{5.2}$$

and we find as in Example 3.5

$$X = \frac{p}{px + qy - z}, \qquad Y = \frac{q}{px + qy - z}, \qquad Z = \frac{-1}{px + qy - z}, \tag{5.3}$$

$$P = \frac{-x}{z}, \qquad Q = \frac{-y}{z}. \tag{5.4}$$

According to Theorem 3.1, these equations give an analytic representation of the envelope of the polar planes $\Sigma_{xyz} \equiv S_{xy}$, if $z = f(x, y)$ is the equation of the surface σ on which the point (x, y, z) moves.

The representation of the contact transformation (5.3), (5.4), can be formally simplified by the introduction of **plane coordinates**. To see how this is done, let us note that the equation of the plane Σ_{xyz} has the form

$$p(\xi - x) + q(\eta - y) - (\zeta - z) = 0,$$

so that

$$p\xi + q\eta - \zeta = px + qy - z,$$

or

$$u\xi + v\eta + w\zeta = 1,$$

where

$$u = \frac{p}{(px + qy - z)}, \qquad v = \frac{q}{(px + qy - z)}, \qquad w = \frac{-1}{(px + qy - z)}. \tag{5.5}$$

The quantities (u, v, w) are the plane coordinates for Σ_{xyz}. In terms of the plane coordinates, the contact transformations (5.3), (5.4) take the simple form

$$X = u, \qquad Y = v, \qquad Z = w, \tag{5.6}$$

(5.7) $$U = x, \qquad V = y, \qquad W = z,$$

where $P = -U/W = -x/z$, and $Q = -V/W = -y/z$. From the forms of (5.6), and (5.7), it is obvious that the transformation coincides with its own inverse. This fact is not apparent from the form of (5.3), (5.4), but from equations (5.1), (5.2) one would expect it.

Note that by "adding an additional dimension", equations (5.3), and (5.4), simplify considerably. This turns out to be a general phenomenon and we shall exploit it more extensively in Chapter III.

The contact transformation known as a **dilatation** is derived from the directrix equation

(5.8) $$\Omega(X, Y, Z; x, y, z) = (X - x)^2 + (Y - y)^2 + (Z - z)^2 - a^2 = 0,$$

where a is a positive constant.

The surface Σ_{xyz} is now the sphere centered at (x, y, z), with radius $a > 0$. If the point (x, y, z) moves on a surface σ, then the point (X, Y, Z) moves on the envelope of the corresponding spheres Σ_{xyz}; this is a surface parallel to σ which is a constant distance a from σ, measured along the normal of σ at the point (x, y, z). The equations required for the determination of the contact transformation are given by

(5.9) $$\begin{cases} (X - x) + p(Z - z) = 0, & (Y - y) + q(Z - z) = 0, \\ (X - x) + P(Z - z) = 0, & (Y - y) + Q(Z - z) = 0, \end{cases}$$

whence

(5.10) $$P = p, \qquad Q = q,$$

and

(5.11) $$X = x - ahp, \qquad Y = y - ahq, \qquad Z = z + ah,$$

where

(5.12) $$h = \pm(p^2 + q^2 + 1)^{-1/2}.$$

The square root in (5.12) can be chosen to be either positive or negative, but it must be the same in all of the equations in (5.11). Hence, the envelope consists of two separate shells.

If we consider the real parameter a to be a variable, then (5.8) represents a family of dilatations. It is readily seen that if a dilation with the parameter value a is followed by a dilatation with the parameter value b, then we have carried out the dilatation with parameter value $a + b$. The set of all dilatations then, can be shown to form a **semigroup** which is isomorphic to the semigroup of non-negative real numbers.

§ 2.6 The Apsidal Transformation

We now take up a certain transformation which, because of its applications, is of importance in its own right. This is the so-called Apsidal transformation, it enables us to understand geometrically the Fresnel wave front.

The geometric basis is the following. Let $m = (x, y, z)$ be any point in space and let O be the origin of the coordinates. Let $r = (x^2 + y^2 + z^2)^{1/2}$ denote the length of the radius vector $\overrightarrow{Om}$. Now construct the plane Σ perpendicular to $\overrightarrow{Om}$, which passes through the origin. Next, let $\mathcal{K}$ be the circle of radius r, lying in Σ. Thus if m is taken to be the North (or South) pole of the sphere σ with radius r, centered at the origin, then $\mathcal{K}$ is the equator. Let $M = (X, Y, Z)$ be a point on this circle and let R denote the length of the vector $\overrightarrow{OM}$. See Figure 6.1.

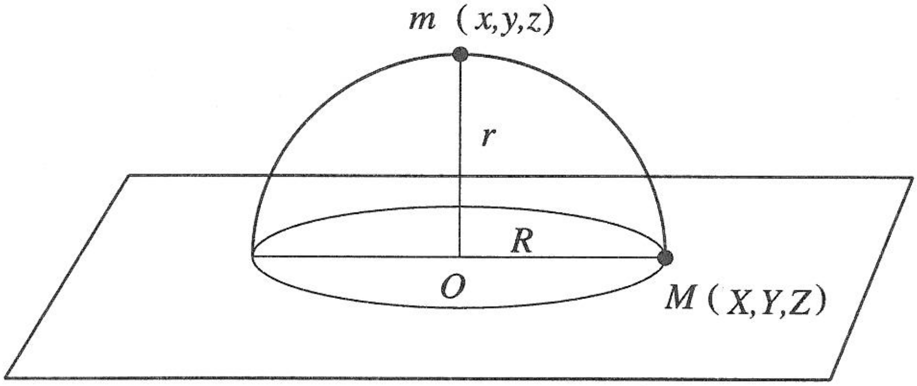

Figure 6.1

From the construction we must have

$$\overrightarrow{Om} \perp \overrightarrow{OM}, \quad \text{and} \quad R = r. \tag{6.1}$$

In order to obtain a relationship between m and M, we regard (6.1) as the directrix equations of a contact transformation, that is

$$\begin{cases} f(X, Y, Z, x, y, z) = (\overrightarrow{OM})^2 - (\overrightarrow{Om})^2 \\ \qquad\qquad = X^2 + Y^2 + Z^2 - x^2 - y^2 - z^2 = 0 \\ \\ g(X, Y, Z, x, y, z) = \overrightarrow{Om} \cdot \overrightarrow{OM} = xX + yY + zZ = 0. \end{cases} \tag{6.2}$$

Using the conditions (3.14), we obtain the additional equations

$$\tag{6.3} \begin{cases} \lambda X + \mu x = P, & \lambda x - \mu X = \rho p, \\ \lambda Y + \mu y = Q, & \lambda y - \mu Y = \rho q, \\ \lambda Z + \mu z = -1, & \lambda z - \mu Z = -\rho, \end{cases}$$

together with

$$\tag{6.4} \begin{cases} (X + PZ)(y + Qz) - (Y + QZ)(x + Pz) = 0, \\ \quad (x + pz)(Y + qZ) - (y + qz)(X + pZ) = 0. \end{cases}$$

If we multiply the corresponding equations in (6.3) together, we see that

$$\rho(Pp + Qq + 1) = (\lambda^2 - \mu^2)\,\overrightarrow{Om} \cdot \overrightarrow{OM} - \lambda\mu(R^2 - r^2) = 0,$$

which confirms the geometrically obvious fact that the tangent plane to σ at m is perpendicular to the tangent plane to σ at M.

Now multiply the equation for P by X, the equation for Q by Y, and the equation for -1 by Z, then use (6.2) to conclude

$$PX + QY - Z = \lambda R^2 \quad \text{and similarly} \quad \rho(px + qy - z) = \lambda r^2.$$

In view of the occurence of the combinations $PX+QY-Z$, and $px+qy-z$, it is natural to introduce the plane coordinates of §2.5, whence

$$P = U(PX + QY - Z) = \lambda U R^2.$$

Proceeding in a similar manner with the other coordinates leads to the system

$$\tag{6.5} \begin{cases} P = \lambda U R^2, & \rho p = \lambda u r^2, \\ Q = \lambda V R^2, & \rho q = \lambda v r^2, \\ -1 = \lambda W R^2, & -\rho = \lambda w r^2. \end{cases}$$

Now, using (6.3) we find

$$\tag{6.6} \lambda U R^2 = P = \lambda X + \mu x.$$

Obviously λ cannot be zero, so set $\tau = \mu/\lambda$. We thus conclude as in the case of (6.6) that

$$\tag{6.7} \begin{cases} UR^2 = X + \tau x, & ur^2 = x - \tau X, \\ VR^2 = Y + \tau y, & vr^2 = y - \tau Y, \\ WR^2 = Z + \tau z, & wr^2 = z - \tau Z. \end{cases}$$

Since $R^2 = r^2$, we find upon squaring and adding the three equations in each set that

$$(U^2 + V^2 + W^2)R^2 = 1 + \tau^2, \qquad (u^2 + v^2 + w^2)r^2 = 1 + \tau^2. \tag{6.8}$$

From (6.8), we see that $U^2+V^2+W^2 = u^2+v^2+w^2$ and τ is determined, up to its sign. Having determined τ, the Apsidal transformation, denoted by $\mathcal{U}$, is determined by

$$\begin{cases} \tau X = x - ur^2, \\ \tau Y = y - vr^2, \\ \tau Z = z - wr^2 \end{cases} \tag{6.9}$$

together with the equations for the plane coordinates. From (6.7), (6.8), and (6.9),

$$\begin{aligned} \tau R^2 U &= \tau X + \tau^2 x = x - ur^2 + \tau^2 x = (1 + \tau^2)x - ur^2 \\ &= (u^2 + v^2 + w^2)R^2 x - R^2 u \end{aligned}$$

and division by R^2 then yields the desired equations,

$$\begin{cases} \tau U = -u + (u^2 + v^2 + w^2)x, \\ \tau V = -v + (u^2 + v^2 + w^2)y, \\ \tau W = -w + (u^2 + v^2 + w^2)z. \end{cases} \tag{6.10}$$

The inverse transformation $\mathcal{U}^{-1}$ is given by

$$\begin{cases} \tau x = -X + UR^2, \\ \tau y = -Y + VR^2, \\ \tau z = -Z + WR^2 \end{cases} \tag{6.11}$$

and

$$\begin{cases} \tau u = U + (U^2 + V^2 + W^2)X, \\ \tau v = V + (U^2 + V^2 + W^2)Y, \\ \tau w = W + (U^2 + V^2 + W^2)Z. \end{cases} \tag{6.12}$$

Because of the ambiguity of the sign of τ, the Apsidal transformation is itself ambiguous. For each element (x, y, z, u, v, w), there are two elements (X, Y, Z, U, V, W) represented by diametrically opposite points of the sphere, (cf. Fig. 6.1).

Theorem 6.1. *The Apsidal transformation $\mathcal{U}$ commutes with the polarity $\mathcal{R}$ represented by equations (5.6)–(5.7).*

PROOF. We consider $\mathcal{U}$ and $\mathcal{R}$ as operating in the system of elements written as sextuplets. Then we have by definition

$$\mathcal{R}(x, y, z, u, v, w) = (u, v, w, x, y, z)$$

and so by direct computation

$$\begin{aligned}&\mathcal{U}R(x, y, z, u, v, w)\\ &\quad = \frac{1}{\tau}\big(-u + \alpha^2 x,\ -v + \alpha^2 y,\ -w + \alpha^2 z,\ x - r^2 u,\ y - r^2 v,\ z - r^2 w\big)\end{aligned}$$

where $\alpha^2 = u^2 + v^2 + w^2$.

Next we have

$$\begin{aligned}&\mathcal{U}(x, y, z, u, v, w)\\ &\quad = \frac{1}{\tau}\big(x + r^2 u,\ y + r^2 v,\ z + r^2 w,\ -u - \alpha^2 x,\ -v - \alpha^2 y,\ -w - \alpha^2 z\big)\end{aligned}$$

and an application of $\mathcal{R}$ to this equation yields the assertion. □

In the theory of wave motion, the Apsidal transformation is used to obtain the Fresnel wave surface as the image of the ellipsoid given by

$$ax^2 + by^2 + cz^2 = 1, \qquad \text{where } a,\ b, \text{ and } c \text{ are positive constants.}$$

The tangent plane to this ellipsoid σ, at the point (x, y, z), has the equation

$$ax\xi + by\eta + cz\zeta = 1;$$

thus it has the plane coordinates $u = ax$, $v = by$, $w = cz$, which satisfy the condition

$$ux + vy + wz = ax^2 + by^2 + cz^2 = 1.$$

Hence in view of (6.2),

$$\begin{aligned}\tau X &= x - r^2 ax &&= x - R^2 ax,\\ \tau Y &= y - r^2 by &&= y - R^2 by,\\ \tau Z &= z - r^2 cz &&= z - R^2 cz.\end{aligned}$$

and

$$x = \frac{\tau X}{(1 - aR^2)}, \qquad y = \frac{\tau Y}{(1 - bR^2)}, \qquad z = \frac{\tau Z}{(1 - cR^2)}.$$

Now $xX + yY + zZ = 0$, so that if we multiply these equations successively by x, y, z, add, and then divide out the τ, we obtain

$$\frac{x^2}{1 - aR^2} + \frac{y^2}{1 - bR^2} + \frac{z^2}{1 - cR^2} = 0. \tag{6.13}$$

This is the equation for the **Fresnel wave surface** .

By a polarity $\mathcal{R}$, the Fresnel wave surface is turned into a Fresnel wave surface with reciprocal axes. This fact is established by Theorem 6.1; for it is only necessary to show that if σ is the ellipsoid $ax^2 + by^2 + cz^2 = 1$, then $\mathcal{R}(\sigma)$ is the ellipsoid $x^2/a + y^2/b + z^2/c = 1$. Indeed, by the polarity (5.6)–(5.7),

$$ax = u = X, \qquad by = v = Y, \qquad cz = w = Z;$$

hence, we have merely to substitute $x = X/a$, $y = Y/b$, $z = Z/c$, into the equation of the ellipsoid and we find $X^2/a + Y^2/b + Z^2/c = 1$. This ellipsoid is turned by $\mathcal{U}$ into the Fresnel wave surface

$$\frac{X^2}{1 - R^2/a} + \frac{Y^2}{1 - R^2/b} + \frac{Z^2}{1 - R^2/c} = 1, \tag{6.14}$$

which was the assertion.

III

Special Contact Transformations

§ 3.1 Special Contact Transformations

Let S represent an element transformation taking $(x, z, p) \in \mathbb{R}_n \times \mathbb{R}_1 \times \mathbb{R}_n$, to the element $(X, Z, P) \in \mathbb{R}_n \times \mathbb{R}_1 \times \mathbb{R}_n$, which we write in the form

$$S(x, z, p) = \big(X(x, z, p), Z(x, z, p), P(x, z, p)\big). \tag{1.1}$$

Let h be an arbitrary constant. We denote by T_h the translation operator

$$T_h(x, z, p) = (x, z + h, p). \tag{1.2}$$

T_h is obviously a contact transformation and the set $\{T_h : -\infty < h < \infty\}$ forms the group of all translations in the z direction.

Lemma 1.1. *The element transformation, S, commutes with the operators, T_h for all real h, if and only if S has the form*

$$X = \widetilde{X}(x, p), \qquad Z = \widetilde{Z}(x, p) + z, \qquad P = \widetilde{P}(x, p). \tag{1.3}$$

PROOF. Suppose S is defined by (1.1). We calculate

$$ST_h(x, z, p) = S(x, z + h, p) = \big(X(x, z + h, p), Z(x, z + h, p), P(x, z + h, p)\big)$$

and

$$\begin{aligned} T_h S(x, z, p) &= T_h\big(X(x, z, p), Z(x, z, p), P(x, z, p)\big) \\ &= \big(X(x, z, p), Z(x, z, p) + h, P(x, z, p)\big). \end{aligned}$$

If $ST_h = T_h S$, then

$$X(x, z+h, p) = X(x, z, p), \qquad P(x, z+h, p) = P(x, z, p),$$

so that X and P are independent of z, i.e. $X = \widetilde{X}(x, p)$, $P = \widetilde{P}(x, p)$. On the other hand,

$$Z(x, z+h, p) = Z(x, z, p) + h.$$

We conclude that $Z_z(x, z, p) = 1$, so that upon integration

$$Z(x, z, p) = z + \widetilde{Z}(x, p).$$

Conversely, it is readily verified that every transformation of the form (1.3) commutes with T_h. □

A contact transformation of the form (1.3) is called a **special contact transformation**. Special contact transformations appear to be somewhat restrictive, but they are easier to analyze than the general contact transformations. Furthermore, some of the most important applications of special contact transformations are to Hamiltonian systems. Moreover, it turns out that a contact transformation in $\mathbb{R}_{n+1}$ can be extended to a special contact transformation in $\mathbb{R}_{n+2}$, in such a way that the extension, when restricted to $\mathbb{R}_{n+1}$, coincides with the original contact transformation. To see how this construction is carried out, let

$$U\colon\; X = X(x, z, p), \qquad Z = Z(x, z, p), \qquad P = P(x, z, p), \tag{1.4}$$

be a general contact transformation in $\mathbb{R}_{n+1}$. By Theorem 3.2 of §I.3, there is a function $\rho = \rho(x, z, p) \neq 0$ such that

$$P \cdot dX - dZ = \rho(p \cdot dx - dz). \tag{1.5}$$

Let $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_n, \overline{x}_{n+1})$ be a point in $\mathbb{R}_{n+1}$ where $\overline{x}_i = x_i$ for $i = 1, \ldots, n$ and $\overline{x}_{n+1} = -z$. Then $(\overline{x}, \overline{z}) \in \mathbb{R}_{n+2}$. In the image space, we adjoin an additional coordinate, $\overline{Z}$, so that $(\overline{X}, \overline{Z})$ is in a domain in $\mathbb{R}_{n+2}$, where $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_n, \overline{X}_{n+1})$, $\overline{X}_i = X_i$ for $i = 1, \ldots, n$ and $\overline{X}_{n+1} = -Z$. Next, let $\overline{p}_{n+1}, \overline{P}_{n+1}$ be direction coefficients. We shall choose $\overline{P}_{n+1}$ appropriately below. Equation (1.5) becomes

$$\sum_{i=1}^{n} P_i \, d\overline{X}_i + d\overline{X}_{n+1} = \rho \left(\sum_{i=1}^{n} p_i \, d\overline{x}_i + d\overline{x}_{n+1} \right). \tag{1.6}$$

Let $\overline{p}_1, \ldots, \overline{p}_{n+1}$ be direction coefficients, where $\overline{p}_1, \ldots, \overline{p}_n$ are related to $p_1, \ldots, p_n$ by

$$\overline{p}_i = p_i \, \overline{p}_{n+1}, \qquad i = 1, \ldots, n. \tag{1.7}$$

Also, define $\overline{P}_1, \ldots, \overline{P}_n$ by

$$\overline{P}_i = P_i \overline{P}_{n+1}, \qquad i = 1, \ldots, n, \tag{1.8}$$

and $\overline{P}_{n+1}$ is chosen as follows. From (1.6) we have

$$\sum_{i=1}^{n} \frac{\overline{P}_i}{\overline{P}_{n+1}} d\overline{X}_i + d\overline{X}_{n+1} = \rho \left(\sum_{i=1}^{n} \frac{\overline{p}_i}{\overline{p}_{n+1}} d\overline{x}_i + d\overline{x}_{n+1} \right)$$

or

$$\sum_{i=1}^{n+1} \overline{P}_i \, d\overline{X}_i = \frac{\rho \overline{P}_{n+1}}{\overline{p}_{n+1}} \sum_{i=1}^{n+1} \overline{p}_i \, d\overline{x}_i. \tag{1.9}$$

The transformation $U\colon (x, z, p) \longrightarrow (X, Z, P)$ is extended to the transformation $\overline{U}\colon (\overline{x}, \overline{z}, \overline{p}) \longrightarrow (\overline{X}, \overline{Z}, \overline{P})$ by adjoining to the $2n+1$ equations defining U, the two additional equations

$$\overline{Z} = \overline{z}, \qquad \overline{P}_{n+1} = \frac{1}{\rho} \overline{p}_{n+1}. \tag{1.10}$$

The system of equations

$$\left\{ \begin{array}{l} \overline{X}_j = X_j(x, -x_{n+1}, p) \equiv \overline{X}_j(\overline{x}, \overline{p}), \quad j = 1, \ldots, n \\ \overline{X}_{n+1} = -Z(x, -x_{n+1}, p) \equiv \overline{X}_{n+1}(\overline{x}, \overline{p}) \\ \overline{Z} = 0 + \overline{z} \quad (\text{i.e. } \widetilde{Z}(\overline{x}, \overline{p}) = 0) \\ \overline{P}_j = (\overline{p}_{n+1}/\rho) \, P_j(x, -x_{n+1}, p) \equiv \overline{P}_j(\overline{x}, \overline{p}), \quad j = 1, \ldots, n \\ \overline{P}_{n+1} = (1/\rho) \, \overline{p}_{n+1} \equiv \overline{P}_{n+1}(\overline{x}, \overline{p}). \end{array} \right. \tag{1.11}$$

is a special contact transformation in $\mathbb{R}_{n+2}$ which satisfies

$$\overline{P} \cdot d\overline{X} = \overline{p} \cdot d\overline{x} \tag{1.12}$$

Conversely, when restricted to $\mathbb{R}_{n+1}$, (1.11) defines a contact transformation which coincides with (1.4). We have proven the following theorem.

Theorem 1.1. *A (general) contact transformation U in the $n+1$-dimensional xz–space, $\mathbb{R}_{n+1}$, can be extended to a special contact transformation $\overline{U}$ in the $n+2$-dimensional $\overline{x}\overline{z}$–space, $\mathbb{R}_{n+2}$, which when restricted to the subspace $\mathbb{R}_{n+1}$ of $\mathbb{R}_{n+2}$ has the same effect as U.*

Example 1.1. The Legendre transformation in $\mathbb{R}_{n+1}$ is

$$U\colon\ X = p, \quad Z = p \cdot x - z, \quad P = x.$$

We have seen that

$$P \cdot dX - dZ = -(p \cdot dx - dz)$$

so that $\rho = -1$.

The extension, $\overline{U}$ is given by

$$\begin{aligned}
&\overline{X}_i = X_i = p_i = \frac{\overline{p}_i}{\overline{p}_{n+1}}, \quad i = 1, \ldots, n \\
&\overline{X}_{n+1} = -Z = -\left(\sum_{i=1}^{n} \frac{\overline{p}_i}{\overline{p}_{n+1}} \overline{x}_i - \overline{x}_{n+1}\right) \\
&\overline{Z} = \overline{z} \\
&\overline{P}_i = \frac{P_i}{\overline{P}_{n+1}} = \frac{x_i}{\overline{P}_{n+1}} = \frac{\overline{x}_i}{\overline{P}_{n+1}} = -\frac{\overline{x}_i}{\overline{p}_{n+1}} \\
&\overline{P}_{n+1} = -\overline{p}_{n+1}.
\end{aligned}$$

Returning now to transformations in $\mathbb{R}_{n+1}$, we drop the tilde notation and have the following result.

Theorem 1.2. *An element transformation of the form (1.3)*

$$X = X(x,p), \qquad Z = Z(x,p) + z, \qquad P = P(x,p)$$

is a special contact transformation if and only if the equation

$$P \cdot dX - p \cdot dx = d(Z - z) = dZ \tag{1.13}$$

holds, where dZ is the total differential of a function Z of (x,p).

The condition (1.13) yields

$$\begin{aligned}
\sum_{j=1}^{n}\left(\sum_{i=1}^{n} P_i \frac{\partial X_i}{\partial x_j} - p_j\right) dx_j &+ \sum_{j=1}^{n}\left(\sum_{i=1}^{n} P_i \frac{\partial X_i}{\partial p_j}\right) dp_j \\
&= \sum_{j=1}^{n} \frac{\partial Z}{\partial x_j} dx_j + \sum_{j=1}^{n} \frac{\partial Z}{\partial p_j} dp_j,
\end{aligned}$$

or comparing coefficients,

$$\begin{cases}
\dfrac{\partial Z}{\partial x_j} = \displaystyle\sum_{i=1}^{n} P_i \frac{\partial X_i}{\partial x_j} - p_j, & j = 1, \ldots, n \\[2ex]
\dfrac{\partial Z}{\partial p_j} = \displaystyle\sum_{i=1}^{n} P_i \frac{\partial X_i}{\partial p_j}, & j = 1, \ldots, n.
\end{cases} \tag{1.14}$$

These conditions characterize contact transformations of the form $X = X(x,p)$, $P = P(x,p)$ in the $2n$ dimensional xp–space. We shall in the sequel refer to such contact transformations as **canonical transformations**.

By using the equivalence of the mixed partial derivatives equation (1.14) can be used to obtain

$$\frac{\partial^2 Z}{\partial x_k\,\partial x_j} = \frac{\partial^2 Z}{\partial x_j\,\partial x_k}, \qquad \frac{\partial^2 Z}{\partial p_k\,\partial x_j} = \frac{\partial^2 Z}{\partial x_j\,\partial p_k}, \qquad \frac{\partial^2 Z}{\partial p_k\,\partial p_j} = \frac{\partial^2 Z}{\partial p_j\,\partial p_k},$$

we obtain the following conditions on $\big(X(x,p), P(x,p)\big)$ which are independent of Z:

$$(1.15)\qquad \begin{cases} \displaystyle\sum_{i=1}^{n}\left[\frac{\partial P_i}{\partial x_k}\frac{\partial X_i}{\partial x_j} - \frac{\partial P_i}{\partial x_j}\frac{\partial X_i}{\partial x_k}\right] = 0, & j,k = 1,\ldots,n, \\[2ex] \displaystyle\sum_{i=1}^{n}\left[\frac{\partial P_i}{\partial p_k}\frac{\partial X_i}{\partial x_j} - \frac{\partial P_i}{\partial x_j}\frac{\partial X_i}{\partial p_k}\right] = \delta_{jk}, & j,k = 1,\ldots,n, \\[2ex] \displaystyle\sum_{i=1}^{n}\left[\frac{\partial P_i}{\partial p_k}\frac{\partial X_i}{\partial p_j} - \frac{\partial P_i}{\partial p_j}\frac{\partial X_i}{\partial p_k}\right] = 0, & j,k = 1,\ldots,n, \end{cases}$$

where $\delta_{jk} = 0$ if $j \neq k$ and $\delta_{kk} = 1$, is the Kronecker delta.

The conditions (1.15) can be written more succinctly in terms of matrices. Let

$$X_x = \left[\frac{\partial X_i}{\partial x_j}\right], \qquad P_x = \left[\frac{\partial P_i}{\partial x_j}\right],$$

$$X_p = \left[\frac{\partial X_i}{\partial p_j}\right], \qquad P_p = \left[\frac{\partial P_i}{\partial p_j}\right].$$

Let ε denote the $n \times n$ identity matrix, and let a prime denote the matrix transpose. Then (1.15) takes the form:

$$(1.16)\qquad \begin{bmatrix} X'_x & P'_x \\ X'_p & P'_p \end{bmatrix}\begin{bmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{bmatrix}\begin{bmatrix} X_x & X_p \\ P_x & P_p \end{bmatrix} = \begin{bmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{bmatrix}.$$

It is traditional to set

$$J = \begin{bmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{bmatrix}.$$

The following simple properties of J are summarized in

Lemma 1.2.

i) $J^2 = -E$, *where E is the $2n \times 2n$ identity matrix*
ii) *The determinant of J, $det(J) = 1$*
iii) $J^{-1} = J' = -J$

A $2n \times 2n$ matrix M which satisfies

$$MJM' = J \tag{1.17}$$

is called a **symplectic** *matrix. Its most important properties are contained in Lemma 1.3*

Lemma 1.3. *Suppose M is a symplectic matrix. Then*
i) *$det(M) = \pm 1$.*
ii) *M' and M^{-1} are also symplectic.*

REMARK. One can actually prove that $\det(M) = +1$, but the proof is more difficult and the statement in the lemma suffices for our purposes.
PROOF.
(i) By (1.17) and Lemma 1.2,

$$1 = \det(J) = \det(MJM') = \det(M)\det(M') = \left[\det(M)\right]^2$$

whence, $\det(M) = \pm 1$.

(ii). Next, $(MJM')^{-1} = J^{-1} = -J$ and so

$$-J = M'^{-1}\,J^{-1}\,M^{-1} = M'^{-1}\,(-J)\,M^{-1}$$

or

$$M'^{-1}\,J\,M^{-1} = J.$$

Multiply on the left by M', and on the right by M to find

$$J = M'\,J\,M = M'\,J\,M''$$

i.e. M' is symplectic.
Finally,

$$-J = J^{-1} = (M'\,J\,M)^{-1} = M^{-1}\,(J^{-1})\,M'^{-1} = M^{-1}\,(-J)\,M^{-1'}$$

so

$$J = M^{-1}\,J\,(M^{-1})'$$

that is, M^{-1} is symplectic, which proves the Lemma. □

On the basis of Lemma 1.3 and equation (1.16) we know that the Jacobi matrix

$$S = \begin{bmatrix} X_x & X_p \\ P_x & P_p \end{bmatrix} \tag{1.18}$$

and its transpose, S', are symplectic. We combine this result, (1.16), and Lemma 1.3 to obtain

Theorem 1.3. *The element transformation $\big(X(x,p), P(x,p)\big)$ is canonical if and only if its Jacobi matrix, (1.18), is symplectic.*

If we make use of the fact that S is symplectic and write out the equation

$$S\,J\,S' = J, \tag{1.19}$$

we find

$$\left\{\begin{aligned} &\sum_{k=1}^{n}\left[\frac{\partial X_i}{\partial x_k}\frac{\partial X_j}{\partial p_k} - \frac{\partial X_i}{\partial p_k}\frac{\partial X_j}{\partial x_k}\right] = 0, \qquad j,k = 1,\ldots,n \\ &\sum_{k=1}^{n}\left[\frac{\partial X_i}{\partial x_k}\frac{\partial P_j}{\partial p_k} - \frac{\partial X_i}{\partial p_k}\frac{\partial P_j}{\partial x_k}\right] = \delta_{jk}, \qquad j,k = 1,\ldots,n \\ &\sum_{k=1}^{n}\left[\frac{\partial P_i}{\partial x_k}\frac{\partial P_i}{\partial p_k} - \frac{\partial P_i}{\partial p_k}\frac{\partial P_i}{\partial x_k}\right] = 0, \qquad j,k = 1,\ldots,n \end{aligned}\right. \tag{1.20}$$

It is convenient to introduce the **Poisson bracket** symbols at this point to simplify the notation. Let $f = f(x,p)$, and $g = g(x,p)$ be two differentiable functions. The Poisson bracket of f and g is defined to be

$$[\![f,g]\!] = [\![f,g]\!]_{xp} = \sum_{k=1}^{n}\left[\frac{\partial f}{\partial x_k}\frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k}\frac{\partial g}{\partial x_k}\right]$$

The principle properties of Poisson brackets are summarized in

Theorem 1.4.

i) $[\![f,g]\!] = -[\![g,f]\!]$ *with* $[\![f,f]\!] = 0$.

ii) $[\![f+g,h]\!] = [\![f,h]\!] + [\![g,h]\!]$.

iii) $[\![fg,h]\!] = f[\![g,h]\!] + g[\![f,h]\!], \quad [\![\alpha f,g]\!] = \alpha[\![f,g]\!]$

iv) $[\![f,[\![g,h]\!]]\!] + [\![g,[\![h,f]\!]]\!] + [\![h,[\![f,g]\!]]\!] = 0$ *(Jacobi identity)*

v) $[\![x_i,x_j]\!] = 0,\ [\![x_i,p_j]\!] = \delta_{ij},\ [\![p_i,p_j]\!] = 0, \quad i,j = 1,\ldots,n.$

The conditions (1.20) can be rewritten in terms of the Poisson brackets. Theorem 1.3 then yields immediately

Theorem 1.5. *The equations $X = X(x,p)$, $P = P(x,p)$ represent a canonical transformation if and only if*

$$\begin{bmatrix} ([\![X_i,X_j]\!]) & ([\![X_i,P_j]\!]) \\ ([\![P_i,X_j]\!]) & ([\![P_i,P_j]\!]) \end{bmatrix} = \begin{bmatrix} ([\![x_i,x_j]\!]) & ([\![x_i,p_j]\!]) \\ ([\![p_i,x_j]\!]) & ([\![p_i,p_j]\!]) \end{bmatrix}.$$

These conditions represent a system of $(2n)^2 - n^2 = 3n^2$ partial differential equations which characterize the canonical transformations.

Example 1.2. Let $n = 1$. The transformation

$$X = \sqrt{x}\cos 2p \qquad P = \sqrt{x}\sin 2p$$

which Poincaré introduced in his studies in celestial mechanics is canonical. For obviously $[\![X, X]\!] = [\![P, P]\!] = 0$, and

$$\begin{aligned}[\![X, P]\!] &= \frac{\partial X}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial X}{\partial p}\frac{\partial P}{\partial x} \\ &= (\tfrac{1}{2}\, x^{-1/2}\cos 2p)(2\, x^{1/2}\cos 2p) + (2\, x^{1/2}\sin 2p)(\tfrac{1}{2}\, x^{-1/2}\sin 2p) = 1\end{aligned}$$

The function Z can be constructed from the equation

$$\begin{aligned}d(Z - z) &= P\, dX - p\, dx \\ &= x^{1/2}\sin 2p\,\{\tfrac{1}{2}\, x^{-1/2}\cos 2p\, dx - 2\, x^{1/2}\sin 2p\, dp\} - p\, dx \\ &= \big(\tfrac{1}{2}\sin 2p\cos 2p - p\big)\, dx - 2x\sin^2 2p\, dp.\end{aligned}$$

We integrate

$$\begin{aligned}\frac{\partial}{\partial x}(Z - z) &= \tfrac{1}{2}\sin 2p\cos 2p - p = \tfrac{1}{4}\sin 4p - p \\ \frac{\partial}{\partial p}(Z - z) &= -2x\sin^2 2p = -x + x\cos 4p\end{aligned}$$

to find

$$Z = z + \tfrac{1}{4}\,(x\sin 4p - 4xy) + \text{const.}$$

An important property of the Poisson brackets is contained in the next theorem.

Theorem 1.6. *The Poisson bracket is invariant under a canonical transformation. More precisely, if $f = f(x, p)$ and $g = g(x, p)$ are given functions, and $X = X(x, p)$, $P = P(x, p)$ with $x = x(X, P)$, $p = p(X, P)$, are canonical transformations, and if we define*

$$F(X, P) = f\big(x(X, P), p(X, P)\big) \quad \textit{and} \quad G(X, P) = g\big(x(X, P), p(X, P)\big),$$

then

$$[\![f, g]\!]_{xp} = [\![F, G]\!]_{XP}.$$

Proof. The proof is simply an unpleasant calculation. Note first that

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{n}\left[\frac{\partial F}{\partial X_i}\frac{\partial X_i}{\partial x_k} + \frac{\partial F}{\partial P_i}\frac{\partial P_i}{\partial x_k}\right],$$

and similar formulas hold for $\partial f/\partial P_k$, $\partial g/\partial x_k$, and $\partial g/\partial p_k$. Thus, by Theorem 1.5

$$
\begin{aligned}
[\![f,g]\!]_{xp} &= \sum_{k=1}^{n}\left[\frac{\partial f}{\partial x_k}\frac{\partial g}{\partial p_k}-\frac{\partial f}{\partial p_k}\frac{\partial g}{\partial x_k}\right]\\
&=\sum_{k=1}^{n}\left\{\sum_{i,j=1}^{n}\frac{\partial F}{\partial X_i}\frac{\partial G}{\partial X_j}\left(\frac{\partial X_i}{\partial x_k}\frac{\partial X_j}{\partial x_k}-\frac{\partial X_j}{\partial x_k}\frac{\partial X_i}{\partial x_k}\right)\right\}\\
&\quad+\sum_{k=1}^{n}\left\{\sum_{i,j=1}^{n}\frac{\partial F}{\partial X_i}\frac{\partial G}{\partial P_j}\left(\frac{\partial X_i}{\partial x_k}\frac{\partial P_j}{\partial p_k}-\frac{\partial X_j}{\partial p_k}\frac{\partial P_i}{\partial x_k}\right)\right\}\\
&\quad+\sum_{k=1}^{n}\left\{\sum_{i,j=1}^{n}\frac{\partial F}{\partial P_j}\frac{\partial G}{\partial X_i}\left(\frac{\partial P_j}{\partial x_k}\frac{\partial X_i}{\partial p_k}-\frac{\partial P_i}{\partial x_k}\frac{\partial X_j}{\partial p_k}\right)\right\}\\
&\quad+\sum_{k=1}^{n}\left\{\sum_{i,j=1}^{n}\frac{\partial F}{\partial P_i}\frac{\partial G}{\partial P_j}\left(\frac{\partial P_i}{\partial x_k}\frac{\partial P_j}{\partial p_k}-\frac{\partial P_j}{\partial x_k}\frac{\partial P_i}{\partial x_k}\right)\right\}\\
&=\sum_{i,j=1}^{n}\frac{\partial F}{\partial X_i}\frac{\partial G}{\partial P_j}[\![X_i,P_j]\!]+\sum_{i,j=1}^{n}\frac{\partial F}{\partial P_i}\frac{\partial G}{\partial X_j}[\![P_j,X_i]\!]\\
&=\sum_{j=1}^{n}\left[\frac{\partial F}{\partial X_j}\frac{\partial G}{\partial P_j}-\frac{\partial F}{\partial P_j}\frac{\partial G}{\partial X_i}\right]=[\![F,G]\!]_{XP},
\end{aligned}
$$

which was the assertion. □

We have seen that the matrix S defined by (1.18), and its transpose are symplectic. The equation

$$S'\,J\,S=J$$

is a restatement of (1.15). We introduce the so-called **Lagrange brackets**

$$[\![X,P]\!]^{x_jp_k}=\sum_{i=1}^{n}\left[\frac{\partial X_i}{\partial x_j}\frac{\partial P_i}{\partial p_k}-\frac{\partial X_i}{\partial p_k}\frac{\partial P_i}{\partial x_j}\right],\qquad i,j=1,\dots,n$$

and obtain

$$
(1.21)\qquad
\begin{cases}
[\![X,P]\!]^{x_jx_k}=0, & [\![X,P]\!]^{x_jp_k}=\delta_{jk},\\[2ex]
[\![X,P]\!]^{p_jx_k}=-\delta_{jk}, & [\![P,P]\!]^{p_jp_k}=0.
\end{cases}
$$

Now let $x=x(X,P)$, $p=p(X,P)$ be the inverse of the transformation $X=X(x,p)$, $P=P(x,p)$. By Lemma 1.3, the matrix

$$S'=\begin{bmatrix}x_X & x_P\\ p_X & p_P\end{bmatrix}$$

is symplectic and therefore

$$[\![x_i, x_j]\!]_{XP} = 0, \qquad [\![x_i, p_j]\!]_{XP} = \delta_{ij},$$
$$[\![p_i, x_j]\!]_{XP} = -\delta_{ij}, \qquad [\![p_i, p_j]\!]_{XP} = 0.$$

Since

$$(S' J S)(S^{-1} J S'^{-1}) = S' J^2 S'^{-1} = -E,$$

the product of the $2n \times 2n$ matrices

$$\begin{bmatrix} ([\![X,P]\!]^{x_i x_j}) & ([\![X,P]\!]^{x_i p_j}) \\ ([\![X,P]\!]^{p_i x_j}) & ([\![X,P]\!]^{p_i p_j}) \end{bmatrix} \begin{bmatrix} ([\![x_i, x_j]\!]_{XP}) & ([\![x_i, p_j]\!]_{XP}) \\ ([\![p_i, x_j]\!]_{XP}) & ([\![p_i, p_j]\!]_{XP}) \end{bmatrix} = -E$$

We summarize these results as

Theorem 1.7. *If $X = X(x,p)$, $P = P(x,p)$ and $x = x(X,P)$, $p = p(X,P)$ are two mutually inverse canonical transformations, then*

$$\sum_{j=1}^{n} [\![X,P]\!]^{x_i p_j} [\![x_j, p_k]\!]_{XP} = \delta_{ik}$$
$$\sum_{j=1}^{n} [\![x_i, p_j]\!]_{XP} [\![X,P]\!]^{x_j p_k} = \delta_{ik}.$$

§ 3.2 Determination of $P(x,p)$ from $X(x,p)$

In applying the theory of contact or canonical transformations to the solution of concrete, mechanical problems, one usually chooses the n functions

$$X_i = X_i(x,p) \tag{2.1}$$

to achieve certain, desired simplifications. The question then arises how much freedom one has in choosing the n additional functions

$$P_i = P_i(x,p) \tag{2.2}$$

so that the $(2n+1)$ equations

(2.3)

$$X_i = X_i(x,p), \qquad Z = z + \Omega(x,p), \qquad P_i = P_i(x,p), \qquad i = 1, \ldots, n$$

define a contact transformation, or equivalently, so that the $2n$ equations

$$X_i = X_i(x,p), \qquad P_i = P_i(x,p), \qquad i = 1, \ldots, n \tag{2.4}$$

define a contact transformation. We shall show that the P_i are essentially uniquely determined from the X_i.

Let us suppose that the $X_i(x,p)$ have been chosen. In view of the results of the previous section, we may assume these X_i to be independent and to satisfy

$$[\![X_i, X_j]\!]_{xp} = 0, \qquad i,j = 1, \ldots, n \tag{2.5}$$

The P_i can be constructed as follows. Choose n additional, arbitrary differentiable functions, $Q_i(x,p)$ such that the Jacobian

$$\frac{\partial(X,Q)}{\partial(x,p)} \neq 0. \tag{2.6}$$

Let

$$\mathcal{A} = \begin{bmatrix} X_x & X_p \\ Q_x & Q_p \end{bmatrix}$$

denote the Jacobi matrix and calculate

$$\mathcal{A}\,J\,\mathcal{A}' = \begin{bmatrix} (X_x\,X_p' - X_p\,X_x) & (X_x\,Q_p' - X_p\,Q_x') \\ (Q_x\,X_p' - Q_p\,X_x') & (Q_x\,Q_p' - Q_p\,Q_x') \end{bmatrix}$$

$$= \begin{bmatrix} ([\![X_i, X_j]\!]_{xp}) & ([\![X_i, Q_j]\!]_{xp}) \\ ([\![Q_i, X_j]\!]_{xp}) & ([\![Q_i, Q_j]\!]_{xp}) \end{bmatrix} \tag{2.7}$$

$$= \begin{bmatrix} (0) & ([\![X_i, Q_j]\!]_{xp}) \\ ([\![Q_i, X_j]\!]_{xp}) & ([\![Q_i, Q_j]\!]_{xp}) \end{bmatrix},$$

where the last entry follows from (2.5).

In general, $\mathcal{A}\,J\,\mathcal{A}' \neq J$; however, should $\mathcal{A}\,J\,\mathcal{A}' = J$, take $P_i = Q_i$.

By (2.6) we can solve for (x,p) in terms of (X,Q) to obtain

$$x = x(X,Q), \qquad p = p(X,Q). \tag{2.8}$$

The inverse matrix for $\mathcal{A}$ is given by

$$\mathcal{A}^{-1} = \begin{bmatrix} x_X & x_Q \\ p_X & p_Q \end{bmatrix},$$

so that

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

$$= -\begin{bmatrix} (x'_X\,p_X - p'_X\,x_X) & (x'_X\,p_Q - p'_X\,x_Q) \\ (x'_Q\,p_X - p'_Q\,x_X) & (x'_Q\,p_Q - p'_Q\,x_q) \end{bmatrix}$$

(2.9)
$$= \begin{bmatrix} (p'_X\,x_X - x'_X\,p_X) & (p'_X\,x_Q - x'_X\,p_Q) \\ (p'_Q\,x_X - x'_Q\,p_X) & (p'_Q\,x_Q - x'_Q\,p_Q) \end{bmatrix}$$

$$= \begin{bmatrix} (\llbracket p,x \rrbracket^{X_i X_j}) & (\llbracket p,x \rrbracket^{X_i Q_j}) \\ (\llbracket p,x \rrbracket^{Q_i X_i}) & (\llbracket p,x \rrbracket^{Q_i Q_j}) \end{bmatrix}.$$

Now the matrix $\mathcal{A}\,J\mathcal{A}'$ has the structure

$$\mathcal{A}\,J\mathcal{A}' = \begin{bmatrix} 0 & \alpha \\ -\alpha' & \beta \end{bmatrix},$$

so that the matrix $(\mathcal{A}\,J\,\mathcal{A}')^{-1}$ must have the structure

$$(\mathcal{A}\,J\,\mathcal{A}')^{-1} = \begin{bmatrix} \alpha'^{-1}\,\beta\,\alpha^{-1} & -\alpha^{-1} \\ \alpha^{-1} & 0 \end{bmatrix},$$

which means that the $\llbracket p,x \rrbracket^{Q_i Q_j}$ in (2.9) must vanish, that is

(2.10)
$$\sum_{k=1}^{n} \frac{\partial p_k}{\partial Q_i}\,\frac{\partial x_k}{\partial Q_j} = \sum_{k=1}^{n} \frac{\partial p_k}{\partial Q_j}\,\frac{\partial x_k}{\partial Q_i},$$

or

(2.11)
$$\frac{\partial}{\partial Q_i}\left(\sum_{k=1}^{n} p_k\,\frac{\partial x_k}{\partial Q_j}\right) = \frac{\partial}{\partial Q_j}\left(\sum_{k=1}^{n} p_k\,\frac{\partial x_k}{\partial Q_i}\right), \qquad i,j = 1,\ldots,n.$$

If we think of the $X = (X_1, \ldots, X_n)$ as parameters, i.e. as having fixed but arbitrary values, then (2.11) implies that the quantities

(2.12)
$$\sum_{k=1}^{n} p_k\,\frac{\partial x_k}{\partial Q_i}, \qquad i = 1,\ldots,n$$

are the components of the differential of a function of Q, that is, there is a function $\widetilde{\Omega} = \widetilde{\Omega}(X, Q)$ such that

$$d_Q\widetilde{\Omega} = \sum_{i,k=1}^{n} p_k \frac{\partial x_k}{\partial Q_i}\, dQ_i, \tag{2.13}$$

where the subscript means that the differential is calculated only with respect to Q. $\widetilde{\Omega}$ is then obtained by the quadrature

$$\widetilde{\Omega} = \int \sum_{i,j=1}^{n} p_k \frac{\partial x_k}{\partial Q_i}\, dQ_i.$$

By the previous section we must demand

$$\sum_{k=1}^{n} \big(P_k\, dX_k - p_k\, dx_k\big) = d\widetilde{\Omega}$$

or, after expanding the differential dx_k

$$\sum_{k=1}^{n} P_k\, dX_k - \sum_{k,j=1}^{n} p_k \frac{\partial x_k}{\partial X_j}\, dX_j - \sum_{k,j=1}^{n} p_k \frac{\partial x_k}{\partial Q_j}\, dQ_j = \sum_{k=1}^{n} \frac{\partial \widetilde{\Omega}}{\partial X_k}\, dX_k - \sum_{k=1}^{n} \frac{\partial \widetilde{\Omega}}{\partial Q_k}\, dQ_k,$$

whence we conclude that

$$P_j = P_j(x,p) = \sum_{k=1}^{n} p_k \frac{\partial x_k}{\partial X_j} + \frac{\partial \widetilde{\Omega}}{\partial X_j}, \qquad j = 1, \ldots, n, \tag{2.14}$$

where the right hand side is evaluated at $X = X(x,p)$, $Q = Q(x,p)$.

Example 2.1. Let $n = 1$ and suppose

$$X(x,p) = xe^p.$$

$Q(x,p)$ is chosen arbitrarily, say

$$Q = e^{-p}.$$

The Jacobian has the value -1 so it never vanishes. The inverse transformation is

$$x = XQ$$
$$p = -\ln Q.$$

Next

$$d_Q\,\widetilde{\Omega} = p\,\frac{\partial x}{\partial Q}\,dQ = -(\ln Q)\,X\,dQ$$

so that

$$\widetilde{\Omega} = [Q - Q\ln Q]X.$$

Finally,

$$P = \Big(pQ - [Q - Q\ln Q]\Big)\Big|_{Q=e^{-p}} = pe^{-p} - e^{-p} - pe^{-p}$$

so that

$$P = -e^{-p},$$

and the canonical transformation is

$$X = xe^{p}$$
$$P = -e^{-p}.$$

Example 2.2. Let

$$X = p$$

Take, for example

$$Q = e^{-x}$$

The Jacobian is e^{-x} so it never vanishes. The inverse transformation is

$$x = -\ln Q, \qquad p = X$$

and so

$$d_Q\,\widetilde{\Omega} = -X\,\tfrac{1}{Q}\,dQ$$

or

$$\widetilde{\Omega} = -X\,\ln Q.$$

Consequently

$$P = \ln Q\Big|_{Q=e^{-x}} = -x$$

and we arrive at the canonical transformation

$$X = p, \qquad P = -x$$

which leads to a Legendre type transformation with $\rho = 1$.

This construction makes use of the fact that the transformation $\big(X(x,p), P(x,p)\big)$ is canonical if and only if there is a function $\widetilde{\Omega}$ such that

$$\sum_{i=1}^{n}\big(P_i\,dX_i - p_i\,dx_i\big) = d\widetilde{\Omega}.$$

If $\det(\partial X_i/\partial x_j) \neq 0$, we can solve $X = X(x,p)$ for p in terms of (X,x) and insert the result into the expression, $\widetilde{\Omega}$, to obtain

$$\Omega(X,x) = \widetilde{\Omega}\big(x, p(X,x)\big). \tag{2.15}$$

Conversely, if a function, $\Omega(X,x)$, is given, it can be used to determine a canonical transformation. For

$$\sum_{i=1}^{n}(P_i\,dX_i - p_i\,dx_i) = d\Omega = \sum_{i=1}^{n}\left(\frac{\partial\Omega}{\partial X_i}\,dX_i + \frac{\partial\Omega}{\partial x_i}\,dx_i\right)$$

set

$$P_i = \frac{\partial\Omega(X,x)}{\partial X_i}, \qquad p_i = -\frac{\partial\Omega(X,x)}{\partial x_i} \tag{2.16}$$

and solve for X in terms of (x,p). This necessitates that we demand

$$\det\left(\frac{\partial^2\Omega}{\partial X_i\,\partial x_i}\right) \neq 0.$$

Insert the resulting expression into the formula for P_i in (2.16) to obtain the canonical transformation

$$X = X(x,p), \qquad P = P(x,p). \tag{2.17}$$

The canonical transformation can be characterized by the function $\Omega(X,x)$. $\Omega(X,x)$ is called the **generating function** for the canonical transformation.

Example 2.3. Suppose

$$\Omega(X,x) = e^x \arctan X.$$

$$P = \frac{\partial\Omega}{\partial X} = \frac{e^x}{1+X^2}$$

$$p = -e^x \arctan X \quad \text{so} \quad X = -\tan(pe^{-x})$$

The canonical transformation is

$$X = -\tan(pe^{-x})$$

$$P = \frac{e^x}{1+\tan^2(pe^{-x})} = \frac{e^x}{\sec^2(pe^{-x})} = e^x\cos^2(pe^{-x}).$$

On occasion it is more convenient to think of a contact transformation as determined by the variables (P,x) rather than (X,x). In that case introduce the generating function $S(P,x)$ by

$$S(P,x) = P\cdot X - \Omega(X,x), \tag{2.18}$$

where the X is thought of as having been obtained by solving the first equation in (2.16) for X in terms of P, and inserting the result into (2.18). In order to carry out those computations, we require that

$$\det\left(\frac{\partial^2}{\partial X_i \partial x_j}\right) \neq 0.$$

When S is defined in that way, it is easily checked that

$$\frac{\partial S}{\partial P_j} = X_j \qquad \text{and} \qquad \frac{\partial S}{\partial x_j} = p_j.$$

These remarks will be of importance in §V.2 when we discuss action-angle variables for mechanical systems.

§ 3.3 Characterization of the General Contact Transformation

In §1 we showed that the study of a single contact transformation could be reduced to the study of a special contact transformation in one higher dimension. This step-up to a higher dimension is usually inconvenient in any concrete situation and we shall rewrite the conditions characterizing a contact transformation in terms of the original coordinates. The conditions characterizing the analytical representation of the functions $X = X(x,p)$, $P = P(x,p)$ describing a special contact transformation were given by a system of partial differential equations, expressed in terms of the Poisson brackets. We shall introduce a similar bracket symbol for the characterization of the general contact transformation in the xzp–space.

Recall that the transformation $(x, z, p) \longrightarrow (\overline{x}, \overline{z}, \overline{p})$, where we again make use of the bar notation, was defined as follows:

$$(3.1) \qquad \overline{x}_i = x_i, \qquad \overline{x}_{n+1} = -z, \qquad \overline{p}_i = \overline{p}_{n+1}\, p_i, \qquad i = 1, \ldots, n$$

Let $f = f(x,z,p)$, $g = g(x,z,p)$ be two differentiable functions, and

$$\begin{aligned} f(x,z,p) &= f(x_1,\ldots,x_n,z,p_1,\ldots,p_n) \\ &= f(\overline{x}_1,\ldots,\overline{x}_n,-\overline{x}_{n+1},\overline{p}_1/\overline{p}_{n+1},\cdots,\overline{p}_n/\overline{p}_{n+1}) \\ &\equiv \overline{f}(\overline{x},\overline{p}) \end{aligned}$$

and similarly

$$g(x,z,p) \equiv \overline{g}(\overline{x},\overline{p}).$$

The Poisson bracket for the pair of functions $\overline{f}$ and $\overline{g}$ is given by

$$(3.2) \qquad [\![\overline{f},\overline{g}]\!]_{\overline{x},\overline{p}} = \sum_{j=1}^{n+1} \left[\frac{\partial \overline{f}}{\partial \overline{x}_j}\frac{\partial \overline{g}}{\partial \overline{p}_j} - \frac{\partial \overline{f}}{\partial \overline{p}_j}\frac{\partial \overline{g}}{\partial \overline{x}_j} \right].$$

We may now rewrite this expression in terms of the original variables

$$\frac{\partial \overline{f}}{\partial \overline{x}_i} = \frac{\partial f}{\partial x_i}, \qquad \frac{\partial \overline{f}}{\partial \overline{p}_i} = \frac{1}{\overline{p}_{n+1}}\frac{\partial f}{\partial p_i} \quad i = 1,\ldots,n$$

$$\frac{\partial \overline{f}}{\partial \overline{x}_{n+1}} = -\frac{\partial f}{\partial z} \qquad \frac{\partial \overline{f}}{\partial \overline{p}_{n+1}} = -\sum_{i=1}^{n} \frac{\partial f}{\partial p_i}\frac{\overline{p}_i}{\overline{p}_{n+1}^2} = -\frac{1}{\overline{p}_{n+1}}\sum_{i=1}^{n} p_i \frac{\partial f}{\partial p_i},$$

and similar formulas for $\overline{g}$ hold. The formula (3.2) takes the form

$$\begin{aligned}
&[\![\overline{f},\overline{g}]\!] \\
&= \frac{1}{\overline{p}_{n+1}} \sum_{j=1}^{n} \left[\frac{\partial f}{\partial x_j} \frac{\partial g}{\partial p_j} - \frac{\partial f}{\partial p_j} \frac{\partial g}{\partial x_j} \right] \\
(3.3) \qquad &+ \frac{1}{\overline{p}_{n+1}} \sum_{j=1}^{n} \left[\frac{\partial f}{\partial z} p_j \frac{\partial g}{\partial p_j} - \frac{\partial g}{\partial z} p_j \frac{\partial f}{\partial p_j} \right] \\
&= \frac{1}{\overline{p}_{n+1}} \sum_{j=1}^{n} \left[\left(\frac{\partial f}{\partial x_j} + p_j \frac{\partial f}{\partial z} \frac{\partial g}{\partial p_j} \right) - \left(\frac{\partial g}{\partial x_j} + p_j \frac{\partial g}{\partial z} \frac{\partial f}{\partial p_j} \right) \right].
\end{aligned}$$

The symbol

$$(3.4) \qquad \{f,g\}_{xzp} = \sum_{j=1}^{n} \left\{ \left(\frac{\partial f}{\partial x_j} + p_j \frac{\partial f}{\partial z} \right) \frac{\partial g}{\partial p_j} - \left(\frac{\partial g}{\partial x_j} + p_j \frac{\partial g}{\partial z} \right) \frac{\partial f}{\partial p_j} \right\}$$

is called the **Mayer bracket** of f and g. The equation (3.3) in terms of the Mayer bracket takes the form

$$(3.5) \qquad [\![\overline{f},\overline{g}]\!]_{\overline{x}\overline{p}} = \frac{1}{\overline{p}_{n+1}} \{f,g\}_{xzp}.$$

The Mayer bracket of two functions satisfies relations analogous to those of the Poisson brackets. The simplest properties are given in the following theorem.

Theorem 3.1. *Let f, g, h, be differentiable functions of the variables (x,y,z), and let α be constant. Then*

i) $\{f,g\} = -\{g,f\}, \qquad \{f,f\} = 0$

ii) $\{\alpha,f\} = 0, \qquad \{\alpha f,g\} = \alpha\{f,g\}$

iii) $\{f+g,h\} = \{f,h\} + \{g,h\}$

iv) $\{fg,h\} = g\{f,h\} + f\{g,h\}$

v) *The Jacobi identity holds in the form*

$$\{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} + f_z\{g,h\} + g_z\{h,f\} + h_z\{f,g\} = 0$$

Note that the indices have been dropped. These properties (i)—(v) can all be checked by direct computation.

The equation (3.5) leads to a formula describing how the Mayer brackets change under a contact transformation, which is analogous to the formula we derived in the previous section for canonical transformations.

To obtain this formula, let

$$X = X(x,z,p), \qquad Z = Z(x,z,p), \qquad P = P(x,z,p) \tag{3.6}$$

be a contact transformation, and let

$$x = x(X,Z,P), \qquad z = z(X,Z,P), \qquad p = p(X,Z,P)$$

be its inverse. Set

$$F(X,Z,P) = F\big(X(x,z,p), Z(x,z,p), P(x,z,p)\big) \equiv f(x,z,p)$$
$$G(X,Z,P) = G\big(X(x,z,p), Z(x,z,p), P(x,z,p)\big) \equiv g(x,z,p).$$

Now lift the variables one dimension and set

$$\overline{X}_i = X_i, \qquad \overline{X}_{n+1} = -Z, \qquad \overline{P}_i = \overline{P}_{n+1}\, P_i, \qquad \overline{P}_{n+1} = \frac{1}{\rho}\overline{p}_{n+1}.$$

We find by (3.5)

$$\frac{1}{\overline{p}_{n+1}}\,\{f,g\}_{xzp} = [\![\overline{f},\overline{g}]\!]_{\overline{x}\overline{p}} = [\![\overline{F},\overline{G}]\!]_{\overline{X}\overline{P}} = \frac{1}{\overline{P}_{n+1}}\,\{F,G\}_{XZP},$$

or, since $\overline{P}_{n+1}/\overline{p}_{n+1} = 1/\rho$,

$$\{F,G\}_{XZP} = \frac{1}{\rho}\{f,g\}_{xzp}. \tag{3.7}$$

Theorem 3.2. *The element transformation*

$$X = X(x,z,p), \qquad Z = Z(x,z,p), \qquad P = P(x,z,p)$$

is a contact transformation if and only if, up to a factor $1/\rho$ it leaves the Mayer bracket of two arbitrary, differentiable functions invariant:

$$\{F,G\}_{XZP} = \frac{1}{\rho}\,\{f,g\}_{xzp}.$$

Theorem 3.2 suggests that the Mayer brackets play the same role for general contact transformations as the Poisson brackets play for the special, or canonical transformations. In order to derive the analogue of Theorem 2.5 of the previous section, we calculate the values of $\{f, g\}_{xzp}$ when f and g are any two of the functions $X_i(x, z, p)$, $Z(x, z, p)$, $P_i(x, z, p)$. By Theorem 3.1, and equation (3.4), we find immediately

$$\{X_i, X_j\}_{xzp} = \rho\{X_i, X_j\}_{XZP} = 0, \qquad i, j = 1, \ldots, n.$$

The other computations are done similarly. This proves the first part of the following theorem.

Theorem 3.3. *In order for the one to one element transformation*

$$X = X(x, z, p), \qquad Z = Z(x, z, p), \qquad P = P(x, z, p),$$

which satisfies the relationship

$$\sum_{j=1}^{n} P_j \, dX_j - dZ = \rho \left(\sum_{k=1}^{n} p_k \, dx_k - dz \right). \tag{3.8}$$

with $\rho(x, z, p) \neq 0$, to be a contact transformation, it is necessary and sufficient that the the following relations are satisfied,

$$\begin{aligned}
\{X_i, X_j\}_{xzp} &= 0, & i, j &= 1, \ldots, n \\
\{X_i, P_j\}_{xzp} &= \rho\delta_{ij}, & i, j &= 1, \ldots, n \\
\{X_i, Z\}_{xzp} &= 0, & i &= 1, \ldots, n \\
\{P_i, P_j\}_{xzp} &= 0, & i, j &= 1, \ldots, n \\
\{P_i, Z\}_{xzp} &= -\rho P_i, & i &= 1, \ldots, n.
\end{aligned}$$

Moreover, the following conditions hold:

$$\begin{aligned}
\{\rho, X_j\} &= \rho \frac{\partial X_j}{\partial z}, \\
\{\rho, Z\} &= \rho \frac{\partial Z}{\partial z} - \rho^2, \\
\{\rho, P_j\} &= \rho \frac{\partial P_j}{\partial z}.
\end{aligned}$$

Again, the indices on the bracket symbols have been dropped, but they are xzp.

Proof. The last three identities are referred to as the **Darboux** formulas. They are most simply verified by using the Jacobi identity of Theorem 3.1 together with the first five formulas of this theorem. Assume that $n \geq 2$

and note that the Jacobi identity when applied to the functions f, X_j, P_j, yields

$$\begin{aligned}(3.9)\qquad &\{f,\{X_j,P_j\}\}+\{X_j,\{P_j,f\}\}+\{P_j,\{f,X_j\}\}\\ &+\frac{\partial f}{\partial z}\{X_j,P_j\}+\frac{\partial X_j}{\partial z}\{P_j,f\}+\frac{\partial P_j}{\partial z}\{f,X_j\}=0,\end{aligned}$$

where the indices have been omitted and f is an arbitrary function. Now take $f = Z$ and use the first identities in the theorem to find

$$\{Z,\rho\}-\{X_j,\rho P_j\}+\frac{\partial Z}{\partial z}\rho-\frac{\partial X_j}{\partial z}\rho P_j=0,$$

or upon expanding,

$$\{X_j,\rho P_j\}=\rho\{X_j,P_j\}+P_j\{X_j,\rho\}=\rho^2+P_j\{X_j,\rho\},$$

$$(3.10)\qquad \{Z,\rho\}-P_j\{X_j,\rho\}-\rho^2+\frac{\partial Z}{\partial z}\rho-\frac{\partial X_j}{\partial z}\rho P_j=0.$$

Next take $f = X_i$, $i = 1, \ldots, n$, where $i \neq j$, in (3.9) to obtain

$$(3.11)\qquad \{X_i\,\rho\}+\frac{\partial X_i}{\partial z}\rho=0,\qquad i=1,\ldots,n.$$

Similarly, take $f = P_i$, $i = 1, \ldots, n$, $i \neq j$, in (3.9) to obtain

$$(3.12)\qquad \{P_i,\rho\}+\frac{\partial P_i}{\partial z}\rho=0.$$

Combining (3.11) and (3.10) yields

$$(3.13)\qquad \{Z,\rho\}-\rho^2+\frac{\partial Z}{\partial z}\rho=0.$$

The equations (3.11), (3.12), (3.13) are the Darboux equations. These equations still hold when $n = 1$, but the proof must be altered somewhat. We begin with the equations

$$(3.14)\qquad \begin{cases} PX_x - Z_x = \rho p, \\ PX_z - Z_z = -\rho, \\ PX_p - Z_p = 0. \end{cases}$$

Differentiate the first equation in (3.14) with respect to z and the second with respect to x and subtract to get

$$(3.15)\qquad \rho_x + p\rho_z = X_xP_z - X_zP_x.$$

Then differentiate the first equation in (3.14) with respect to p and the third with respect to x and subtract to get

$$\rho + p\rho_p = X_x P_p - X_p P_x. \tag{3.16}$$

Finally, differentiate the second equation in (3.14) with respect to p and the third with respect to z and subtract to obtain

$$\rho_p = X_p P_z - X_z P_p. \tag{3.17}$$

Now calculate directly using (3.15), (3.16), (3.17) to find

$$\{\rho, X\} = \rho X_z, \qquad \{\rho, P\} = \rho P_z. \tag{3.18}$$

(3.10) holds when $n = 1$ so that by the first equation in (3.18) and (3.10), we obtain immediately (3.13) for $n = 1$, which completes the proof of the theorem. □

Corollary 3.1.

$$\frac{\partial \rho}{\partial X_j} = \frac{\partial P_j}{\partial z} + \frac{1}{\rho}\frac{\partial \rho}{\partial z} P_j, \qquad j = 1, \ldots, n.$$

$$\frac{\partial \rho}{\partial P_j} = -\frac{\partial X_j}{\partial z}, \qquad j = 1, \ldots, n.$$

$$\frac{\partial \rho}{\partial Z} = -\frac{1}{\rho},\frac{\partial \rho}{\partial z}.$$

Proof. These identities follow by manipulating the formulas already derived. For example, we have on the one hand that

$$\{\rho, X_j\}_{XZP} = \sum_{i=1}^{n}\left[\left(\frac{\partial \rho}{\partial X_i} + P_i\frac{\partial \rho}{\partial Z}\right)\frac{\partial X_j}{\partial P_i} - \left(\frac{\partial X_j}{\partial X_i} + P_i\frac{\partial X_j}{\partial Z}\right)\frac{\partial \rho}{\partial P_i}\right] = -\frac{\partial \rho}{\partial P_j}$$

since $\partial X_j/\partial P_i = \partial X_j/\partial Z = 0$ and $\partial X_i/\partial X_j = \delta_{ij}$. On the other hand

$$\{\rho, X_j\}_{XZP} = \frac{1}{\rho}\{\rho, X_j\}_{xzp} = \frac{\partial X_j}{\partial z}$$

by one of the Darboux equalities. □

Corollary 3.2. *The functions (X, P) of a contact transformation are independent of z if and only if ρ is a constant.*

Corollary 3.1 implies immediately that if ρ is a constant, then X_j and P_j are independent of z. On the other hand, we saw in our treatment of special contact transformations, that if $X = X(x,p)$, $Y = Y(x,p)$, $Z = z+\Omega(x,p)$, then ρ was a constant and we could set things up so that ρ was one.

We close this section with a few remarks.

First, if $\big(X(x,z,p), Z(x,z,p), P(x,z,p)\big)$ define a contact transformation, then the Jacobi determinant is

$$\frac{\partial(X,Z,P)}{\partial(x,z,p)} = \rho^{n+1}.$$

This is most easily seen by going back to the equivalent, special contact transformation and noting by direct calculation that

$$1 = \frac{\partial(\overline{X},\overline{P})}{\partial(\overline{x},\overline{p})} = \frac{\overline{P}_{n+1}}{\partial\overline{p}_{n+1}}\,\frac{\partial(X,Z,P,\overline{P}_{n+1})}{\partial(x,z,p,\overline{p}_{n+1})} = \frac{1}{\rho^{n+1}}\,\frac{\partial(X,Z,P)}{\partial(x,z,p)}.$$

We also note that if the $n+1$ independent functions, Z, $X_1,\ldots,X_n$ of (x,z,p) are pairwise in involution, that is if they satisfy

$$\{Z,X_i\}_{xzp} = 0, \qquad \{X_i,X_j\}_{xzp} = 0,$$

then the functions $P_1,\ldots,P_n,\rho$ can be calculated as follows. Since (X,Z) are independent, the rank of the matrix

$$\begin{bmatrix} X_x & X_z \\ Z_x & Z_z \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial X_i}{\partial x_j}\right) & \left(\frac{\partial X_i}{\partial z}\right) \\ \left(\frac{\partial Z}{\partial X_j}\right) & \left(\frac{\partial Z}{\partial z}\right) \end{bmatrix}$$

is $n+1$. Thus, there is at least one system of $n+1$ independent equations in the set

$$\sum_{i=1}^{n} \frac{P_i}{\rho}\,\frac{\partial X_i}{\partial x_j} - \frac{1}{\rho}\,\frac{\partial Z}{\partial X_j} = p_j, \qquad j = 1,\ldots,n$$

$$\sum_{i=1}^{N} \frac{P_i}{\rho}\,\frac{\partial X_i}{\partial z} - \frac{1}{\rho}\,\frac{\partial Z}{\partial z} = -1$$

$$\sum_{i=1}^{n} \frac{P_i}{\rho}\,\frac{\partial X_i}{\partial p_j} - \frac{1}{\rho}\,\frac{\partial Z}{\partial p_j} = 0, \qquad j = 1,\ldots,n$$

of $2n+1$ equations from which one can calculate the P_i/ρ, $i = 1, \ldots, n$, $1/\rho$, and by the above theory, the remaining equations are identically satisfied. This result can be stated as

Theorem 3.4. *Let $X_1, \ldots, X_n, Z$ be $n+1$ independent functions which are pairwise in involution. Then there is precisely one contact transformation for which these are the first $n+1$ functions and the remaining $n+1$ functions $P_1, \ldots, P_n, \rho$ may be obtained by solving a linear system of equations.*

In Chapters I and II, we found that contact transformations could be generated from directrix equations. The general n-dimensional case is similar to the treatment given in Chapter II §3 for contact transformations in space. We sketch the development.

Consider the q independent functions, $\Omega_1, \ldots, \Omega_q$ of the variables $(X, Z; x, z)$, $X = X(X_1, \ldots, X_n)$, $x = x(x_1, \ldots, x_n)$ with $1 \leq q \leq n+1$. The differential of Ω_j is

$$d\Omega_j = \sum_{k=1}^{n} \frac{\partial \Omega_j}{\partial X_k} dX_k + \frac{\partial \Omega_j}{\partial Z} dZ + \sum_{l=1}^{n} \frac{\partial \Omega_j}{\partial x_l} dx_l + \frac{\partial \Omega_j}{\partial z} dz = 0, \; j = 1, \ldots, q.$$

At the same time, if the (X, Z, P) are to define a contact transformation, they must satisfy

$$\sum_{k=1}^{n} P_k \, dX_k - dZ = \rho \left(\sum_{l=1}^{n} p_l \, dx_l - dz \right).$$

There must be parameters $\lambda_1, \ldots, \lambda_q$ such that

$$\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial X_k} = P_k, \qquad \sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial X_l} = -\rho \, p_l,$$

$$\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial Z} = -1, \qquad \sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial z} = +\rho.$$

This yields the following symmetric set of equations for the determination of p_l and P_k once the $\lambda_1, \ldots, \lambda_q$ have been determined.

$$p_l = -\left(\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial x_l} \right) \Big/ \left(\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial z} \right), \qquad l = 1, \ldots, n$$

$$P_k = -\left(\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial X_k} \right) \Big/ \left(\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial Z} \right), \qquad k = 1, \ldots, n$$

The construction of the contact transformation is carried out by eliminating the λ variables from the system

$$\Omega_1(X, Z; x, z) = 0, \ \ldots, \ \Omega_q(X, Z; x, z) = 0$$

$$\left(\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial z}\right) p_l + \sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial x_l} = 0, \qquad l = 1, \ldots, n$$

$$\left(\sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial Z}\right) P_k + \sum_{j=1}^{q} \lambda_j \frac{\partial \Omega_j}{\partial X_k} = 0, \qquad k = 1, \ldots, n,$$

and then finding the functions $X = X(x, z, p)$, $Z = Z(x, z, p)$. One then obtains values for the λ variables by solving an appropriate homogeneous system, and then obtains the functions $P = P(x, z, p)$.

For $n > 3$, $q > 2$, this procedure can be quite laborious. Conditions must be placed on the directrices, $\Omega_i(X, Z; x, z)$, which allow these computations to be carried out. Because of the limited usefulness in actual construction, we shall limit these general considerations to the above remarks.

IV

Continuous Transformation Groups

§ 4.1 One Parameter Groups of Transformations

In the applications of contact transformations, which we treat in the sequel, we shall often deal with transformations depending on a parameter. In this section we shall discuss some general properties of certain classes of transformations which will prove to be useful to us. To begin with, we shall make no reference to the contact preserving properties of these transformations. These considerations will take place in an m dimensional space, and points in this space will be denoted by $\mathfrak{x}$, $\mathfrak{y}$ etc.

A one parameter family of transformations of a region, $\mathfrak{B}$ of this space onto itself is a set of functions, $\mathfrak{F} = \{f(\cdot, u)\}$, of the form

$$\mathfrak{x} \longrightarrow \mathfrak{y} = f(\mathfrak{x}, u). \tag{1.1}$$

Here, $\mathfrak{x}, \mathfrak{y} \in \mathfrak{B}$, and u is a real parameter, $-c < u < c$. The family is such that to each u there corresponds precisely one transformation from the family.

The components of the transformations, $f(\mathfrak{x}, u)$, are assumed to be continuously differentiable with respect to $\mathfrak{x}$ and u. Moreover, we assume that the family contains the identity, which we take to correspond to $u = 0$:

$$f(\mathfrak{x}, 0) = \mathfrak{x}. \tag{1.2}$$

Finally, we assume there is an open interval $(-\delta, \delta)$, $0 < \delta < c$, such that if $u, v \in (-\delta, \delta)$, then the composition of two functions in the family is again a function in the family, that is, if

$$f(\mathfrak{x}, u), f(\mathfrak{x}, v) \in \mathfrak{F}$$

then there is a value $w \in (-c, c)$ which depends on u, v, but not on $\mathfrak{x}$ such that

$$f\big(f(\mathfrak{x}, u),\, v\big) = f(\mathfrak{x}, w) \in \mathfrak{F}. \tag{1.3}$$

Denote this functional relationship by

$$w = \phi(u, v) \tag{1.4}$$

We assume that $\phi(u, v)$ is continuously differentiable for $u, v \in (-\delta, \delta)$.

Theorem 1.1. *Suppose u, v, w, $\phi(u,v)$, $\phi(v,w) \in (-\delta, \delta)$. Then the function ϕ satisfies*

i) *the associative functional equation: $\phi\big(\phi(u,v), w\big) = \phi\big(u, \phi(v,w)\big)$ and*

ii) *is symmetric: $\phi(u,v) = \phi(v,u)$.*

PROOF. To prove the associativity, let

$$\mathfrak{x} = f(\mathfrak{x}^0, u), \qquad \mathfrak{y} = f(\mathfrak{x}, v).$$

Then

$$\begin{aligned} f(\mathfrak{y}, w) &= f\big(f(\mathfrak{x}, v), w\big) = f\big(\mathfrak{x}, \phi(v, w)\big) \\ &= f\big(f(\mathfrak{x}^0, u), \phi(v, w)\big) = f\big(\mathfrak{x}^0, \phi(u, \phi(v, w))\big). \end{aligned}$$

Next,

$$\begin{aligned} f(\mathfrak{y}, w) &= f\big(f(\mathfrak{x}, v), w\big) = f\big(f(f(\mathfrak{x}^0, u), v), w\big) \\ &= f\big(f(\mathfrak{x}^0, \phi(u, v)), w\big) = f\big(\mathfrak{x}^0, \phi(\phi(u, v), w)\big) \end{aligned}$$

which proves (i).

Before proving the symmetry of ϕ, observe that

$$f\big(f(\mathfrak{x}, u), 0\big) = f\big(\mathfrak{x}, \phi(u, 0)\big) = f(\mathfrak{x}, u)$$

so that $\phi(u, 0) = u$. Similarly $\phi(0, v) = v$.

Now set

$$\phi_1(u, v) = \frac{\partial \phi(u, v)}{\partial u} \quad \text{and} \quad \phi_2(u, v) = \frac{\partial \phi(u, v)}{\partial v}.$$

Since $\phi_1(u, 0) = 1$ and $\phi_2(0, v) = 1$, $\phi_1(u, v)$ and $\phi_2(u, v)$ are nonzero for (u, v) near $(0, 0)$. Differentiate the equality in (i), first with respect to v, then with respect to w to obtain

$$\begin{aligned} \phi_1\big(\phi(u, v), w\big)\, \phi_2(u, v) &= \phi_2\big(u, \phi(v, w)\big)\, \phi_1(v, w) \\ \phi_2\big(\phi(u, v), w\big) &= \phi_2\big(u, \phi(v, w)\big)\, \phi_2(v, w). \end{aligned}$$

Divide the first equation by the second and set $w = 0$ to get

$$\frac{\phi_1\big(\phi(u,v),0\big)}{\phi_2\big(\phi(u,v),0\big)}\,\phi_2(u,v) = \frac{\phi_1(v,0)}{\phi_2(v,0)}.$$

Introduce the function $\omega(v) = \phi_1(v,0)/\phi_2(v,0)$ and rewrite this differential equation as

$$\omega(\phi)\,\frac{\partial\phi}{\partial v} = \omega(v). \tag{1.5}$$

Equation (1.5) can be regarded as an ordinary differential equation for ϕ in terms of v with u entering as a parameter. Both $\phi_1(0,0) = 1$ and $\phi_2(0,0) = 1$ so that $\omega(0) = 1$.

Now let ψ be a function such that $\psi'(v) = \omega(v)$ and let $\psi(0) = 0$. Then for u fixed, (1.5) is

$$\psi'(\phi)\,d\phi = \psi'(v)\,dv. \tag{1.6}$$

Note that when $v = 0$, $\phi(u,0) = u$ so that integrating (1.6) yields

$$\psi(\phi)\Big|_u^{\phi} = \psi(v)\Big|_0^{v}$$

or

$$\psi\big(\phi(u,v)\big) = \psi(v) + \psi(u).$$

$\psi'(0) = \omega(0) = 1 \neq 0$ so that $\psi(v)$ is invertible near zero. Thus,

$$\phi(u,v) = \psi^{-1}\big(\psi(v)+\psi(u)\big) = \psi^{-1}\big(\psi(u)+\psi(v)\big) = \phi(v,u)$$

which proves the assertion. □

As a by product of the proof, we see that $\phi(u,v)$ is given in terms of a monotonically increasing function. It is convenient to introduce a new parameter, $t = \psi(u)$. The new parameter is called a canonical or an additive parameter for the family $\mathfrak{F}$. It contains the additive identity $t = 0$ and one obtains the inverse for a transformation $\mathfrak{y} = f(\mathfrak{x},u) = f\big(\mathfrak{x},\psi^{-1}(t)\big)$ by replacing t with $-t$, for

$$\begin{aligned} f\big(\mathfrak{y},\psi^{-1}(-t)\big) &= f\big(f(\mathfrak{x},\psi^{-1}(t)),\psi^{-1}(-t)\big) = f\big(\mathfrak{x},\phi(\psi^{-1}(t),\psi^{-1}(-t))\big) \\ &= f\big(\mathfrak{x},\psi^{-1}(\psi(\psi^{-1}(t)) + \psi(\psi^{-1}(-t)))\big) = f(\mathfrak{x},0) = \mathfrak{x}. \end{aligned}$$

Because of this relationship, we say that the transformations of our family form a commutative or abelian group.

It will be convenient for the subsequent development to introduce the standard notation for transformation groups. In place of the parameters u, v, etc., originally introduced, we use the canonical parameters t, τ, etc. We think of the $t^{\underline{\text{th}}}$ transformation, S_t, as operating on a point $\mathfrak{x} \in \mathfrak{B}$ and sending it to another point $\mathfrak{y} \in \mathfrak{B}$, defined by

$$\mathfrak{y} = S_t\,\mathfrak{x} = f(\mathfrak{x}, \psi^{-1}(t)). \tag{1.7}$$

The product of two transformations, $S_\tau\, S_t$ is obtained by operating first on a point $\mathfrak{x} \in \mathfrak{B}$ by S_t and then operating on $S_t\mathfrak{x}$ by S_τ, which means that it is ultimately defined in terms of the composition of mappings from the set $\mathfrak{F}$. The family satisfies, by what we have proven above, the group properties:

$$\left\{\begin{array}{l} S_0\mathfrak{x} = \mathfrak{x} \quad \text{so} \quad S_0 = E, \text{ the identity} \\ S_t\, S_\tau = S_{t+\tau} = S_{\tau+t} = S_\tau\, S_t \\ S_t(S_\tau\, S_s) = (S_t\, S_\tau)S_s \\ S_{-t} = S_t^{-1} \quad \text{so} \quad S_t\, S_{-t} = S_{t-t} = E. \end{array}\right. \tag{1.8}$$

The group properties for the transformations, $\{S_t\}$, have the same properties as the additive group of real numbers. In this sense we say that the given family represents an additive group of transformations.

Families of transformations indexed naturally by the parameter t, denoting time, arise in the solution of differential equations

$$\dot{\mathfrak{x}} = F(t, \mathfrak{x}), \qquad \mathfrak{x}(0) = \mathfrak{x}_0.$$

If we make certain minimal assumptions on the problem so that there exists a unique solution near points $(0, \mathfrak{x}_0)$, then we see that the solution operator to each point $\mathfrak{x}_0$, is the point $\mathfrak{x}$, which is the value of the solution at the time t. This family of transformations need not define a one parameter group of transformations; however, if F is independent of t, then it does, and in fact, we can prove a little more. Consider for this discussion the autonomous[1] system

$$\dot{\mathfrak{x}} = F(\mathfrak{x}), \qquad \mathfrak{x}(0) = \mathfrak{x}_0, \tag{1.9}$$

where F is assumed to be at least once continuously differentiable. By the standard existence and uniqueness theorem, we know that a solution, $\mathfrak{x}(t)$, to the problem (1.9) exists for $|t|$ sufficiently small. Thus, for $\mathfrak{x}_0$, the solution $\mathfrak{x}$ at time t is given in terms of a family of operators, $\{S_t\}$, such that

$$\mathfrak{x} = S_t(\mathfrak{x}_0). \tag{1.10}$$

The family of operators $\{S_t\}$ is a one parameter group of operators since, by the existence and uniqueness theory for ordinary differentiable equations[2],

[1] The system is called autonomous if the right hand side does not explicitly depend on t

[2] See Coddington and Levinson or Kamke, E.

S_0 is the identity operator, $S_{-t} = S_t^{-1}$ and $S_t\, S_{t_1} = S_{t+t_1}$, from which the associativity and commutativity of the operators follow. The converse of this result also holds, that is if $\{S_t\}$ is a one parameter group of operators, there is a function, $F(\mathfrak{x})$, such that solutions to $\dot{\mathfrak{x}} = F(\mathfrak{x})$ generate the group. We formulate this assertion as

Theorem 1.2.

i) *Let* $\mathfrak{x} = \mathfrak{x}(t)$, $\mathfrak{x}_0 \in \mathfrak{B}$ *and suppose for some* c,

$$\mathfrak{x} = \mathfrak{x}(t) = S_t\, \mathfrak{x}_0, \qquad |t| < c,\ 0 < c \leq \infty \tag{1.11}$$

where $\{S_t\}$ *represents a one parameter group with the additive parameter* t. *Then there exists a vector valued function* $F(\mathfrak{x})$ *which is continuous in* $\mathfrak{B}$ *such that* $\mathfrak{x}(t)$ *given by (1.11) satisfies*

$$\dot{\mathfrak{x}} = F(\mathfrak{x}), \qquad \mathfrak{x}(0) = \mathfrak{x}_0. \tag{1.12}$$

ii) *The solution,* $\mathfrak{x}(t)$, *to the differential equation problem (1.12) is of the form (1.11), where the* $\{S_t\}$ *form a one parameter group of operators with the additive parameter* t.

PROOF. We need only prove (i.)

Let us set $\mathfrak{x} = S_t\, \mathfrak{x}_0$. By supposition

$$S_{t+s}(\mathfrak{x}_0) = S_s(S_t\mathfrak{x}_0) = S_s\, \mathfrak{x}, \qquad S_0\, \mathfrak{x} = \mathfrak{x},$$

so that

$$\lim_{s\to 0} \frac{S_s(\mathfrak{x}) - S_0(\mathfrak{x})}{s} = \frac{d}{ds}\, S_s(\mathfrak{x}) \equiv F(\mathfrak{x})$$

does not depend explicitly on t. To show that $\dot{\mathfrak{x}} = F(\mathfrak{x})$, we simply observe that

$$\dot{\mathfrak{x}}(t+\Delta t) = \lim_{\Delta t\to 0} \frac{\mathfrak{x}(t+\Delta t) - \mathfrak{x}(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{S_{\Delta t}\, \mathfrak{x} - S_0 \mathfrak{x}}{\Delta t} = F(\mathfrak{x}),$$

which completes the proof. □

The function, $F(\mathfrak{x})$, is sometimes called the generator of the group.

Example 1.1. Let $m = 1$ and $\mathfrak{x} = x$. Then the solution to

$$\dot{x} = x, \qquad x(0) = x_0$$

is

$$x = x_0\, e^t$$

and $S_t = e^t$ defines the transformation group.

Example 1.2. Let $m = 2$. The solution to

$$\dot{\mathfrak{x}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathfrak{x}, \qquad \mathfrak{x}(0) = \mathfrak{x}_0$$

is

$$\mathfrak{x}(t) = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \mathfrak{x}_0$$

and the one parameter transformation group is defined by

$$S_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

Example 1.3. Suppose $m = 1$, $\mathfrak{x} = x$. The solution to

$$\dot{x} = x + t, \qquad x(0) = x_0$$

is given by

$$x = e^t (x_0 + 1) + t + 1$$

and the solution operators

$$x(t) = S_t x_0 = e^t (x_0 + 1) + t + 1$$

do not form a group.

Example 1.4. The family of transformations, S_t defined by

$$\mathfrak{x} = S_t \mathfrak{x}_0 = \begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix} \mathfrak{x}_0$$

is a one parameter group. The associated differential equation is given by calculating

$$\dot{\mathfrak{x}} = \lim_{s \to 0} \left[\frac{S_s \mathfrak{x} - S_0 \mathfrak{x}}{s} \right] = \lim_{s \to 0} \begin{bmatrix} \frac{\cosh s - 1}{s} & \frac{\sinh s}{s} \\ \frac{\sinh s}{s} & \frac{\cosh s - 1}{s} \end{bmatrix} \mathfrak{x} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathfrak{x}.$$

The notation $S_t \mathfrak{x}_0$ brings out the fact that a point $\mathfrak{x}_0$ "flows" or "is carried" to the point $\mathfrak{x}$. The set of points $\{S_t \mathfrak{x}_0 : \mathfrak{x}_0 \in \mathfrak{B}' \subset \mathfrak{B}\}$ is called the **flow** and the curve $\mathfrak{x}(t) = S_t \mathfrak{x}_0$ is called the **trajectory** or **solution curve** (passing through $\mathfrak{x}_0$).

§ 4.2 One Parameter Families of Contact Transformations

We now consider the special system of $2n+1$ differential equations for $2n+1$ unknowns $X = (X_1, \dots, X_n)$, Z, $P = (P_1, \dots, P_n)$

$$\dot{X} = \xi(X, Z, P, t), \qquad \dot{Z} = \zeta(X, Z, P, t), \qquad \dot{P} = \pi(X, Z, P, t) \tag{2.1}$$

which satisfy the initial conditions

$$X = x, \qquad Z = z, \qquad P = p, \quad \text{when } t = 0. \tag{2.2}$$

The functions $\xi = (\xi_1, \dots, \xi_n)$, ζ, $\pi = (\pi_1, \dots, \pi_n)$ are all assumed to be continuously differentiable. The solutions to (2.1), (2.2)

$$X = X(x, z, p, t), \qquad Z = Z(x, z, p, t), \qquad P = P(x, z, p, t) \tag{2.3}$$

determine a family of transformations,

$$S_t \colon (x, z, p) \longrightarrow (X, Z, P). \tag{2.4}$$

In this section we give the necessary and sufficient conditions on ξ, ζ, and π, which imply that the transformations (2.4) are contact transformations uniformly in t.

In Chapter III, we had found that the transformations must satisfy

$$P \cdot dX - dZ = \rho(p \cdot dx - dz), \qquad \rho \neq 0 \tag{2.5}$$

(2.5) is supposed to hold when the differentials are calculated only with respect to the spatial variables. When X, Z, P, also depend upon t, then dZ, for example, is actually given by

$$dZ = \sum_{j=1}^{n} \frac{\partial Z}{\partial x_j}\, dx_j + \frac{\partial Z}{\partial z}\, dz + \sum_{j=1}^{n} \frac{\partial Z}{\partial p_j}\, dp_j + \frac{\partial Z}{\partial t}\, dt.$$

A similar assertion holds for the dX_i. Thus, the condition (2.5) must be replaced by

(2.6)
$$\left(\sum_{i=1}^{n} P_i\, dX_i - dZ\right) - \left(\sum_{i=1}^{n} P_i \frac{\partial X_i}{\partial t} - \frac{\partial Z}{\partial t}\right) dt = \rho \left(\sum_{i=1}^{n} p_i\, dx_i - dz\right).$$

By (2.1), $\partial X_i / \partial t = \xi_i(X, Z, P, t)$, $\partial Z / \partial t = \zeta(X, Z, P, t)$. Let us introduce the function

$$\mathcal{H} = \mathcal{H}(X, Z, P, t) = \sum_{i=1}^{n} P_i\, \xi_i(X, Z, P, t) - \zeta(X, Z, P, t). \tag{2.7}$$

Then the relation (2.6) takes the form

$$P \cdot dX - dZ = \rho\,(p \cdot dx - dz) + \mathcal{H}\,dt \tag{2.8}$$

If $dt = 0$, equation (2.8) reduces to (2.5). (2.8) represents a system of $2n+2$ equations relating the variables (X, Z, P, t) with those of (x, z, p, t), which is obtained by expanding the differentials and comparing coefficients. To obtain the conditions we seek, we shall rewrite these conditions in the (X, Z, P, t) variables. This is most simply done by working directly with (2.8). First differentiate (2.8) with respect to t and note that the differential operator, d, commutes with the differentiation d/dt. This leads to

$$\sum_{j=1}^{n} \pi_j\,dX_j + \sum_{j=1}^{n} P_j\,d\xi_j - d\zeta = \dot{\rho}\left(\sum_{j=1}^{n} p_j\,dx_j - dz\right) + \dot{\mathcal{H}}\,dt \tag{2.9}$$

where the dot, as usual, represents d/dt. From (2.8) and (2.9) we obtain

$$\sum_{j=1}^{n} \pi_j\,dX_j + \sum_{j=1}^{n} P_j\,d\xi_j - d\zeta - \dot{\mathcal{H}}\,dt = \frac{\dot{\rho}}{\rho}\left(\sum_{j=1}^{n} P_j\,dX_j - dZ - \mathcal{H}\,dt\right). \tag{2.10}$$

From (2.7), we find

$$d\mathcal{H} = \sum_{j=1}^{n} \xi_j\,dP_j + \sum_{j=1}^{n} P_j\,d\xi_j - d\zeta$$

so that (2.10) takes the form

$$d\mathcal{H} + \sum_{j=1}^{n} \pi_j\,dX_j - \sum_{j=1}^{n} \xi_j\,dP_j = \frac{\dot{\rho}}{\rho}\left(\sum_{j=1}^{n} P_j\,dX_j - dZ\right) + \left(\dot{\mathcal{H}} - \frac{\dot{\rho}}{\rho}\mathcal{H}\right) dt. \tag{2.11}$$

Expand $d\mathcal{H}$ in the form

$$d\mathcal{H} = \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial X_j}\,dX_j + \frac{\partial \mathcal{H}}{\partial Z}\,dZ + \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial P_j}\,dP_j + \frac{\partial \mathcal{H}}{\partial t}\,dt$$

insert the result into (2.11) and compare coefficients to obtain the following system of equalities

$$\begin{cases} \dfrac{\partial \mathcal{H}}{\partial X_j} = -\pi_j + \dfrac{\dot{\rho}}{\rho} P_j, & \dfrac{\partial \mathcal{H}}{\partial P_j} = \xi_j, \\[2ex] \dfrac{\partial \mathcal{H}}{\partial Z} = -\dfrac{\dot{\rho}}{\rho}, & \dfrac{\partial \mathcal{H}}{\partial t} = \dot{\mathcal{H}} - \dfrac{\dot{\rho}}{\rho}\mathcal{H}. \end{cases} \tag{2.12}$$

We now obtain expressions for ξ_j, ζ, π_j. The ξ_j and π_j are obtained directly from (2.12) by eliminating the quotient, $\dot{\rho}/\rho$, and solving. ζ is obtained by combining (2.7) with (2.12). We find

$$\left\{\begin{aligned} \xi_j &= \frac{\partial \mathcal{H}}{\partial P_j}, \\ \zeta &= \sum_{j=1}^{n} P_j \xi_j - \mathcal{H} = \sum_{j=1}^{n} P_j \frac{\partial \mathcal{H}}{\partial P_j} - \mathcal{H}, \\ \pi_j &= -\frac{\partial \mathcal{H}}{\partial X_j} - P_j \frac{\partial \mathcal{H}}{\partial Z}. \end{aligned}\right. \tag{2.13}$$

These computations prove

Theorem 2.1. *In order for the solution (2.3) of the system (2.1) to represent a one parameter family of contact transformations containing the identity, it is necessary that (2.1) be a canonical system, that is, that there exists a function, $\mathcal{H} = \mathcal{H}(X, Z, P, t)$ called the characteristic function, such that the system (2.1) has the form*

$$\left\{\begin{aligned} \frac{d}{dt} X_j &= \frac{\partial \mathcal{H}}{\partial P_j}, \\ \frac{d}{dt} Z &= \sum_{j=1}^{n} P_j \frac{\partial \mathcal{H}}{\partial P_j} - \mathcal{H}, \\ \frac{d}{dt} P_j &= -\frac{\partial \mathcal{H}}{\partial X_j} - P_j \frac{\partial \mathcal{H}}{\partial Z}. \end{aligned}\right. \tag{2.14}$$

The converse to this theorem is also valid. We state and prove

Theorem 2.2. *The solution to the canonical equations (2.14), which satisfy the initial conditions (2.2), generates a one parameter family of contact transformations, which for $t = 0$ contains the identity.*

PROOF. We must show that every solution (2.14) and (2.2) satisfies the strip condition (2.8). For notational purposes, let us set

$$\Omega = \Omega(t) \equiv \sum_{j=1}^{n} P_j \, dX_j - dZ - \mathcal{H} \, dt$$

and

$$\Omega(0) \equiv \omega = \sum_{j=1}^{n} p_j \, dx_j - dz.$$

Then the strip condition, (2.8), takes the form

$$\Omega(t) = \rho\omega.$$

We set up a differential equation for $\dot{\Omega}$ making use of (2.14). The proof is simply a calculation. We find

$$\dot{\Omega} = \sum_{j=1}^{n} \left(\dot{P}_j\, dX_j + P_j\, d\dot{X}_j\right) - d\dot{Z} - \dot{\mathcal{H}}\, dt.$$

Now

$$\mathcal{H} = \sum_{j=1}^{n} P_j\, \dot{X}_j - \dot{Z}$$

so that

$$\begin{aligned}
\dot{\Omega} &= \sum_{j=1}^{n} \left(\dot{P}_j\, dX_j - \dot{X}_j\, dP_j\right) \\
&= -\sum_{j=1}^{n} \left(\frac{\partial \mathcal{H}}{\partial X_j} + P_j \frac{\partial \mathcal{H}}{\partial Z}\right) dX_j - \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial P_j}\, dP_j \\
&= -\sum_{j=1}^{n} \left(\frac{\partial \mathcal{H}}{\partial X_j}\, dX_j + \frac{\partial \mathcal{H}}{\partial P_j}\, dP_j\right) - \frac{\partial \mathcal{H}}{\partial Z} \sum_{j=1}^{n} P_j\, dX_j \\
&= -\left\{\sum_{j=1}^{n} \left(\frac{\partial \mathcal{H}}{\partial X_j}\, dX_j + \frac{\partial \mathcal{H}}{\partial P_j}\, dP_j\right) + \frac{\partial \mathcal{H}}{\partial Z}\, dZ + \mathcal{H}_t\, dt\right\} \\
&\qquad - \frac{\partial \mathcal{H}}{\partial Z} \left\{\sum_{j=1}^{n} P_j\, dX_j\right\} + \frac{\partial \mathcal{H}}{\partial Z}\, dZ + \mathcal{H}_t\, dt \\
&= -d\mathcal{H} - \frac{\partial \mathcal{H}}{\partial Z} \left\{\sum_{j=1}^{n} P_j\, dX_j - dZ - \mathcal{H}\, dt\right\} - \frac{\partial \mathcal{H}}{\partial Z} \mathcal{H}\, dt + \mathcal{H}_t\, dt \\
&= -d\mathcal{H} - \frac{\partial \mathcal{H}}{\partial Z} \Omega - \frac{\partial \mathcal{H}}{\partial Z} \mathcal{H}\, dt + \mathcal{H}_t\, dt.
\end{aligned}$$

Next we calculate using (2.14)

$$\begin{aligned}\frac{d\mathcal{H}}{dt} &= \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial X_j}\frac{dX_j}{dt} + \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial P_j}\frac{dP_j}{dt} + \frac{\partial \mathcal{H}}{\partial Z}\frac{dZ}{dt} + \mathcal{H}_t \\ &= \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial X_j}\frac{\partial \mathcal{H}}{\partial P_j} - \sum_{j=1}^{n} \frac{\partial \mathcal{H}}{\partial P_j}\left(\frac{\partial \mathcal{H}}{\partial X_j} + P_j\frac{\partial \mathcal{H}}{\partial Z}\right) \\ &\quad + \frac{\partial \mathcal{H}}{\partial Z}\left(\sum_{j=1}^{n} P_j \frac{\partial \mathcal{H}}{\partial P_j} - \mathcal{H}\right) + \mathcal{H}_t \\ &= -\mathcal{H}\frac{\partial \mathcal{H}}{\partial Z} + \mathcal{H}_t.\end{aligned}$$

Thus,

$$d\mathcal{H} = -\mathcal{H}\frac{\partial \mathcal{H}}{\partial Z}\,dt + \mathcal{H}_t\,dt$$

and so from the previous calculation

$$\dot{\Omega} = -\frac{\partial \mathcal{H}}{\partial Z}\Omega.$$

We integrate to obtain

$$\Omega = \rho\omega \tag{2.15}$$

where

$$\rho = \exp\left(-\int_0^t \frac{\partial \mathcal{H}}{\partial Z}\,dt\right) \tag{2.16}$$

which proves the assertion. □

We combine this result with Theorem 2 of section 1 to obtain the following theorem.

Theorem 2.3. *A family of element transformations*

$$X = X(x, z, p, t), \qquad Z = Z(x, z, p, t), \qquad P = P(x, z, p, t)$$

is

i) *a family of contact transformations containing the identity for* $t = 0$ *in* $\mathbb{R}_{n+1}$ *if and only if it represents a solution of the canonical differential equations (2.14) and satisfies (2.2) at* $t = 0$,

ii) *a group of contact transformations if and only if the characteristic function,* $\mathcal{H}$, *of the canonical system does not depend explicitly on the parameter* t, *i.e.* $\mathcal{H}$ *is autonomous.*

We shall close this section with a few remarks on the characteristic function $\mathcal{H} = \mathcal{H}(X, Z, P, t)$. From the fourth equation in (2.12), we have

$$\rho \frac{\partial \mathcal{H}}{\partial t} = \rho \dot{\mathcal{H}} - \dot{\rho} \mathcal{H}.$$

Divide by ρ^2 to find

$$\frac{1}{\rho} \frac{\partial \mathcal{H}}{\partial t} = \frac{\rho \dot{\mathcal{H}} - \dot{\rho} \mathcal{H}}{\rho^2} = \frac{d}{dt} \left(\frac{\mathcal{H}}{\rho} \right).$$

Integrate with respect to t to find

$$\frac{\mathcal{H}}{\rho} - \frac{\mathcal{H}^0}{\rho^0} = \int_0^t \frac{1}{\rho} \mathcal{H}_t \, dt, \tag{2.17}$$

where the superscript indicates that the arguments of $\mathcal{H}$ and ρ are to be taken at $t = 0$:

$$\rho^0 = \rho(x, z, p, 0) = 1, \qquad \mathcal{H}^0 = \mathcal{H}(x, z, p, 0).$$

The fact that $\rho^0 = 1$ is a consequence of (2.16).

We consider two special cases.

Case 1. $\partial \mathcal{H} / \partial t = 0$ so that $\mathcal{H}$ does not depend explicitly on t.

Then the family, $\{S_t\}$, represents a one parameter group of contact transformations. The relation (v2.17) implies that

$$\mathcal{H}(X, Z, P) = \mathcal{H}^0(x, z, p) \, \rho(x, z, p). \tag{2.18}$$

(2.18) has a geometric interpretation. Let us think of the parameter, t, as the time and the curve along which

$$(X, Z, P) = S_t(x, z, p)$$

moves in $\mathbb{R}_{2n+1}$ as its orbit under the group of contact transformations. Along this orbit, the function $\mathcal{H}(X, Z, P)$, up to the factor $\mathcal{H}^0$, coincides with $\rho(X, Z, P)$.

If in particular $\mathcal{H}^0 = 0$ at a point (x, z, p), then $\mathcal{H}(X, Z, P) = 0$ along the whole orbit through it. The strip condition is along the orbit. If we think of (X, Z, P) as an element in $\mathbb{R}_{n+1}$, then we refer to the orbit as an orbital strip of the group of contact transformations in $\mathbb{R}_{n+1}$. For points on the orbital strip, the second equation in (2.14) simplifies to

$$\frac{dZ}{dt} = \sum_{j=1}^{n} P_j \frac{\partial \mathcal{H}}{\partial P_j}.$$

Case 2 $\partial\mathcal{H}/\partial Z = 0$ so that $\mathcal{H}$ does not depend explicitly on Z and by (2.16) $\rho = \rho(X, Z, P, t) \equiv 1$.

The canonical equations (2.14) reduce to

$$\frac{dX_j}{dt} = \frac{\partial\mathcal{H}}{\partial P_j}, \qquad \frac{dP_j}{dt} = -\frac{\partial\mathcal{H}}{\partial X_j} \tag{2.19}$$

together with the additional equation

$$\frac{dZ}{dt} = \sum_{j=1}^{n} P_j \frac{\partial\mathcal{H}}{\partial P_j} - \mathcal{H} \tag{2.20}$$

for the construction of Z.

The transformations determined by (2.19) are the special, or xp–transformations which commute with translations along the z–axis. The equation (2.8) in this case reads

$$\sum_{j=1}^{n} (P_j\, dX_j - p_j\, dx_j) = d(Z - z) + \mathcal{H}\, dt.$$

If in addition, $\partial\mathcal{H}/\partial t = 0$, then $\mathcal{H} = \mathcal{H}^0$. The family determined by solutions to (2.19) is a group of contact transformations which on the orbit passing through (x, z, p) satisfies $\mathcal{H}(X, Z, P) = \mathcal{H}^0(x, z, p)$.

§ 4.3 Transformations of Canonical Differential Equations

Since a canonical system is always uniquely determined by the characteristic function, $\mathcal{H}$, we shall begin this discussion with the canonical differential equations written as

$$\left\{\begin{aligned} \dot{x}_j &= \frac{\partial\mathcal{H}}{\partial p_j}, \qquad j = 1, \ldots, n \\ \dot{z} &= \sum_{j=1}^{n} p_j \frac{\partial\mathcal{H}}{\partial p_j} - \mathcal{H} \\ \dot{p}_j &= -\frac{\partial\mathcal{H}}{\partial x_j} - p_j \frac{\partial\mathcal{H}}{\partial z}, \qquad j = 1, \ldots, n \end{aligned}\right. \tag{3.1}$$

and denote the initial values by

$$x_j(0) = x_j^0, \quad z(0) = z^0, \quad p_j = p_j^0, \qquad j = 1, \ldots, n \tag{3.2}$$

We wish to investigate how these equations transform under a one parameter family of contact transformations. For this purpose, let

$$\left.\begin{aligned} X_j &= X_j(x,z,p,t), \\ Z &= Z(x,z,p,t), \\ P_j &= P_j(x,z,p,t), \end{aligned}\right\} \quad j = 1,\dots,n \tag{3.3}$$

or more succinctly, let

$$(X,Z,P) = T_t(x,z,p) \tag{3.4}$$

denote such a transformation and let $K = K(X,Z,P,t)$ and $\sigma = \sigma(X,Z,P,t)$ be the functions such that

$$\sum_{j=1}^{n} P_j\, dX_j - dZ - K\, dt = \sigma\left(\sum_{j=1}^{n} p_j\, dx_j - dz\right) \tag{3.5}$$

We have seen that the solution (x,z,p,t) to (3.1), (3.2) depends upon (x^0,z^0,p^0,t) and can be written as

$$(x,z,p) = S_t(x^0,z^0,p^0). \tag{3.6}$$

Now carry out the substitution indicated by (3.3) or, equivalently, by (3.4). The initial values transform to functions of

$$\begin{aligned} (X^0,Z^0,P^0) &= T_0(x^0,z^0,p^0) \\ &= \Big(X(x^0,z^0,p^0,0), Z(x^0,z^0,p^0,0), P(x^0,z^0,p^0,0)\Big) \end{aligned}$$

and the solutions to (3.1), (3.2) transform to functions of (X^0,Z^0,P^0,t) according to

$$(X,Z,P) = T_t S_t T_0^{-1}(X^0,Z^0,P^0). \tag{3.7}$$

Let

$$S_t^* = T_t S_t T_0^{-1}. \tag{3.8}$$

$\{S_t^*\}$ is a one parameter family of contact transformations, so there exists a canonical system for it which is determined by a characteristic function

$$\mathcal{H}^* = \mathcal{H}^*(X,Z,P,t).$$

We must determine $\mathcal{H}^*$ in terms of known quantities.

Since T_0 is a contact transformation, we find from (3.5)

$$\sum_{j=1}^{n} P_j^0 \, dX_j^0 - dZ^0 = \sigma^0 \left(\sum_{j=1}^{n} p_j^0 \, dx_j^0 - dz^0 \right) \tag{3.9}$$

where $\sigma^0 = \sigma(X^0, Z^0, P^0, 0)$. Further,

$$\sum_{j=1}^{n} p_j \, dx_j - dz = \rho \left(\sum_{j=1}^{n} p_j^0 \, dx_j^0 - dz^0 \right) + \mathcal{H} \, dt. \tag{3.10}$$

Now, using (3.5), (3.10) and (3.9) we find

$$\begin{aligned}
\sum_{j=1}^{n} P_j \, dX_j - dZ &= \sigma \left(\sum_{j=1}^{n} p_j \, dx_j - dz \right) + K \, dt \\
&= \sigma \left\{ \rho \left(\sum_{j=1}^{n} p_j^0 \, dx_j^0 - dz^0 \right) + \mathcal{H} \, dt \right\} + K \, dt \\
&= \sigma \left\{ \frac{\rho}{\sigma^0} \left(\sum_{j=1}^{n} P_j^0 \, dX_j^0 - dZ^0 \right) + \mathcal{H} \, dt \right\} + K \, dt \\
&= \frac{\sigma\rho}{\sigma^0} \left(\sum_{j=1}^{n} P_j^0 \, dX_j^0 - dZ^0 \right) + \Big(\sigma\mathcal{H} + K \Big) dt.
\end{aligned}$$

The coefficient of dt represents the desired characteristic function

$$\mathcal{H}^* = \mathcal{H}^*(X, Z, P, t) = \Big(\sigma\mathcal{H} + K \Big). \tag{3.11}$$

Observe that σ and K are already evaluated at (X, Z, P, t). The function $\mathcal{H}$, initially evaluated at (x, z, p, t), must simply be rewritten in terms of the new variables $(X, Z, P) = T_t^{-1}(x, z, p)$.

Having determined the characteristic function, $\mathcal{H}^*$, we can rewrite the system (3.1) immediately in terms of the new variables.

Example 3.1. Suppose $\mathcal{H} = (x^2 + p^2)/2$. The canonical equations are

$$\dot{x} = p, \qquad \dot{p} = -x.$$

Let us make the substitution

$$X = x^2/2, \qquad P = p/x.$$

Then

$$P\,dX = p\,dx$$

so that $\sigma = 1$ and $K = 0$. Consequently, $\mathcal{H}^*$ is obtained by evaluating $\mathcal{H}$ at $x = \sqrt{2X}$, $p = \sqrt{2X}\,P$ which yields

$$\mathcal{H}^*(X,P) = X + XP^2$$

and the canonical equations are

$$\dot{X} = 2XP \qquad \dot{P} = -(1+P^2).$$

This transformation essentially uncouples the original system. This system integrates directly to

$$X = \beta\cos^2(t-\alpha), \qquad P = -\tan(t-\alpha)$$

where α, β are arbitrary constants. x, and p are obtained by

$$\begin{aligned} x &= \sqrt{2X} = \sqrt{2\beta}\,\cos(t-\alpha) = \left(\sqrt{2\beta}\,\cos\alpha\right)\cos t + \left(\sqrt{2\beta}\,\sin\alpha\right)\sin t \\ p &= \sqrt{2X}P = -\sqrt{2\beta}\,\sin(t-\alpha) = -\left(\sqrt{2\beta}\,\cos\alpha\right)\sin t + \left(\sqrt{2\beta}\,\sin\alpha\right)\cos t \end{aligned}$$

or taking $A = \sqrt{2\beta}\,\cos\alpha$, $B = \sqrt{2\beta}\,\sin\alpha$

$$\begin{aligned} x &= A\cos t + B\sin t \\ p &= -A\sin t + B\cos t \end{aligned}$$

which is the standard solution.

We close this section with a final remark. Suppose $\mathcal{H}$ is independent of z so that the canonical equations are

$$\dot{x}_j = \frac{\partial \mathcal{H}}{\partial p_j}, \qquad \dot{p}_j = -\frac{\partial \mathcal{H}}{\partial x_j}. \tag{3.12}$$

Now make the substitution

$$\begin{aligned} &X = X(x,p), \qquad P = P(x,p) \\ &\text{with} \quad P\cdot dX = p\cdot dx. \end{aligned} \tag{3.13}$$

This is a special contact transformation which is independent of the parameter t. Then $\sigma = 1$, $K = 0$ and $\mathcal{H}^*$ is obtained by evaluating $\mathcal{H}$ at $x = x(X,P)$, $p = p(X,P)$ and the canonical equations in the (X,P) variables are

$$\dot{X}_j = \frac{\partial \mathcal{H}^*}{\partial P_j}, \qquad \dot{P}_j = -\frac{\partial \mathcal{H}^*}{\partial X_j}. \tag{3.14}$$

Since (3.12) transforms in (3.14) with $\mathcal{H}^*$ arising from $\mathcal{H}$ by means of (3.13), the special contact transformation is also called a canonical transformation.

§ 4.4 The Theorems of Liouville and Poisson

In this section, we prove some results of Liouville and Poisson. They are special cases of a more general approach due to Lie, but are interesting and important in their own right.

We shall first suppose that $\mathcal{H}$ is independent of z so that the canonical system is

$$(4.1)\qquad \begin{cases} \dot{x} = \dfrac{\partial \mathcal{H}}{\partial p_j}, & j = 1, \dots, n \\ \dot{p} = -\dfrac{\partial \mathcal{H}}{\partial x_j} & j = 1, \dots, n \end{cases}$$

and the resulting contact transformations are the special or xp-transformations. Let us suppose further that $X_i(x, p, t)$ is a (first) integral for (4.1), that is along a solution,

$$(4.2)\qquad X_i(x, p, t) = c_i$$

where the c_i are constants. From the theory of ordinary differential equations, we might expect that we would need $2n$ independent first integrals, X_i, P_i, in order to construct the solution to (4.1). However, we have seen in III.3.2 that, given the X_i components of a special contact transformation, we can construct the P_i components by means of a quadrature. The assertion of the first theorem is, therefore, plausible.

Theorem 4.1. *(Liouville). If $X_1(x, p, t), \dots, X_n(x, p, t)$ are n independent first integrals for (4.1) which are pairwise in involution:*

$$(4.3)\qquad [\![X_i, X_j]\!]_{xp} = \sum_{k=1}^{n} \left[\frac{\partial X_i}{\partial x_k} \frac{\partial X_j}{\partial p_k} - \frac{\partial X_i}{\partial p_k} \frac{\partial X_j}{\partial x_k} \right] = 0, \qquad i, j = 1, \dots, n$$

then the general solution of the canonical system can be constructed by means of a quadrature.

Proof. We begin by thinking of the functions $P_1, (x, p, t), \dots, P_n(x, p, t)$ as being constructed from the $X_1, (x, p, t), \dots, X_n(x, p, t)$ so that (X, P) is a special contact transformation. Introduce (X, P) as new variables and let $\mathcal{H}^*$ denote the characteristic function obtained from $\mathcal{H}$ in the XP-variables. Along a solution, the X_i are constant so that the characteristic equations are

$$\begin{cases} \dot{X}_j = \dfrac{\partial \mathcal{H}^*}{\partial P_j} = 0, & j = 1, \dots, n \\ \dot{P}_j = -\dfrac{\partial \mathcal{H}^*}{\partial X_j} & j = 1, \dots, n. \end{cases}$$

These equations imply that $\mathcal{H}^*$ is independent of P_j so $\mathcal{H}^* = \mathcal{H}^*(X,t)$ and that $X_i = c_i$ so that with $c = (c_1, \ldots, c_n)$, $\mathcal{H}^* = \mathcal{H}^*(c,t)$. Thus,

$$\begin{aligned} X_j &= c_j, & j &= 1, \ldots, n \\ P_j &= -\int \frac{\partial \mathcal{H}^*(c,t)}{\partial c_j}\, dt, & j &= 1, \ldots, n \end{aligned}$$

and the Z component is

$$Z = -\int \mathcal{H}^*(c,t)\, dt.$$

This completes the proof of Liouvilles theorem. □

The importance of Liouvilles theorem lies in the fact that only n first integrals are needed to construct the complete solution to the canonical system (4.1).

Let us now suppose that $F(x,p,t)$ is an integral of (4.1). Then along a solution it is a constant, so

$$\begin{aligned} 0 = \frac{d}{dt}F(x,p,t) &= F_t + \sum_{j=1}^{n}\left[\frac{\partial F}{\partial x_j}\dot{x}_j + \frac{\partial F}{\partial p_j}\dot{p}_j\right] \\ &= F_t + \sum_{j=1}^{n}\left[\frac{\partial F}{\partial x_j}\frac{\partial \mathcal{H}}{\partial p_j} - \frac{\partial F}{\partial p_j}\frac{\partial \mathcal{H}}{\partial x_j}\right] \end{aligned}$$

from which it follows that

$$F_t + [\![F, H]\!]^{\cdot}_{xp} = 0. \tag{4.4}$$

Conversely, the equation (4.4) can be regarded as a first order, partial differential equation for F. The characteristic equations for F are

$$\frac{dt}{d\tau} = 1, \qquad \frac{dx_j}{d\tau} = \frac{\partial \mathcal{H}}{\partial p_j}, \qquad \frac{dp_j}{d\tau} = -\frac{\partial \mathcal{H}}{\partial x_j}, \qquad j = 1, \ldots, n \tag{4.5}$$

The first equation has $t = \tau + \tau_0$ for its solution and if we take $\tau_0 = 0$, $t = \tau$ then (4.5) is the canonical system (4.1). We summarize this discussion as

Theorem 4.2. *$F(x,p,t)$ =constant is a first integral for (4.1) if and only if it satisfies (4.4).*

Now suppose that both F and G are integrals for (4.1). Then along a solution,

$$F(x,p,t) = \alpha, \qquad G(x,p,t) = \beta,$$

where α and β are constants, both F and G satisfy Theorem 1.4 of §3.1, and (4.4). The Jacobi identity yields

$$\begin{aligned} 0 &= [\![F,[\![G,H]\!]]\!] + [\![G,[\![H,F]\!]]\!] + [\![H,[\![F,G]\!]]\!] \\ &= -[\![F,G_t]\!] + [\![G,F_t]\!] - [\![[\![F,G]\!],H]\!] \\ &= -\left\{\frac{\partial}{\partial t}[\![f,g]\!] + [\![[\![F,G]\!],H]\!]\right\}. \end{aligned}$$

This argument proves

Theorem 4.3. *(Poisson) If $F(x,p,t)$ and $G(x,p,t)$ are two integrals for (4.1), then so is $[\![F,G]\!]$.*

Poisson's theorem tells us that if we know two integrals for (4.1), a third can be found by differentiation processes. However, the integral so obtained may or may not be independent of the first two integrals.

These theorems can be generalized to the more general canonical system

$$\text{(4.6)} \qquad \begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j} & j = 1,\ldots,n \\ \dot{z} = \displaystyle\sum_{k=1}^{n} p_k \frac{\partial \mathcal{H}}{\partial p_k} - \mathcal{H} \\ \dot{p}_j = -\dfrac{\partial \mathcal{H}}{\partial x_j} - p_j \dfrac{\partial \mathcal{H}}{\partial z}, & j = 1,\ldots,n \end{cases}$$

where $\mathcal{H} = \mathcal{H}(x,z,p,t)$.

Again, an integral is a function $F = F(x,z,p,t)$ which is constant along a solution. However, the Poisson brackets must now be replaced by the Mayer brackets.

Theorem 4.4. *Suppose $X_1,\ldots,X_n,Z$ are $n+1$ independent first integrals of (4.6) which satisfy*

$$\begin{aligned} \{X_i,X_j\}_{xzp} &= 0, \qquad i,j = 1,\ldots,n \\ \{X_i,Z\}_{xzp} &= 0, \qquad i = 1,\ldots,n. \end{aligned}$$

Then the general solution to (4.6) can be constructed by means of a quadrature.

PROOF. Construct the functions $P_1, \dots, P_n$ so that (X, Z, P) is a contact transformation. Along a solution, $X_i = c_i$, $Z = \gamma$ where c_i, and γ are constants so that $\dot{X}_i = \dot{Z} = 0$ as before. These relations and (4.6) imply that the function $\mathcal{H}^*$, expressed in terms of (X, Z, P) is independent of P, so that $\mathcal{H}^* = \mathcal{H}^*(c, \gamma, t)$. The third set of equations in (4.6) is in the new variables

$$\dot{P}_j = -\frac{\partial \mathcal{H}^*}{\partial c_j} - P_j \frac{\partial \mathcal{H}^*}{\partial \gamma}$$

which is immediately solvable. The complete solution is given by

$$\begin{aligned} X_i(x, z, p, t) &= c_i, \qquad i = 1, \dots, n \\ Z(x, z, p, t) &= \gamma, \\ P_i(x, z, p, t) &= \frac{-\left\{\int \exp\left(\int \frac{\partial \mathcal{H}^*(c,\gamma,t)}{\partial \gamma}\, dt\right) \frac{\partial \mathcal{H}^*}{\partial c_j}(c, \gamma, t)\, dt\right\}}{\exp\left(\int \frac{\partial \mathcal{H}^*}{\partial \gamma}\, dt\right)} \qquad i = 1, \dots, n \end{aligned}$$

which was the assertion of the theorem. □

REMARK: If n first integrals, $X_1, \dots, X_n$ are known which are pairwise in involution, the equations uncouple. Introduce new coordinates (X, Z, P) as before. $\mathcal{H}$ goes into $\mathcal{H}^*(c, Z, t)$. A nonlinear equation for Z must be solved. Once that is done, we can proceed to construct the functions P_i as in the proof of the theorem.

Again, we observe that if $F(x, z, p, t)$ is a first integral for (4.6), then

$$\begin{aligned} 0 &= F_t + \sum_j^n \left(\frac{\partial F}{\partial x_j}\right) \dot{x}_j + \frac{\partial F}{\partial z} \dot{z} + \sum_j^n \left(\frac{\partial F}{\partial p_j}\right) \dot{p}_j \\ &= F_t + \sum_j^n F_{x_j} \mathcal{H}_{p_j} + F_z \left(\sum_j^n p_j \mathcal{H}_{p_j} - \mathcal{H}\right) + \sum_j^n F_{p_j} \left(-\mathcal{H}_{x_j} - p_j \mathcal{H}_z\right) \\ &= F_t + \sum_j^n \left(F_{x_j} + p_j F_z\right) \mathcal{H}_{p_j} - \sum_j^n \left(\mathcal{H}_{x_j} + p_j \mathcal{H}_z\right) F_{p_j} - F_z \mathcal{H} \end{aligned}$$

we, therefore, obtain the equality

$$F_t + \{F, \mathcal{H}\} - F_z \mathcal{H} = 0. \tag{4.7}$$

This equation can also be regarded as a first order partial differential equation for F having (4.6) as its system of characteristic equations. The analogue of Theorem 4.2 is

Theorem 4.5. *$F(x, z, p, t)$ =const. is a first integral for (4.6) if and only if it satisfies (4.7).*

Finally, we derive an analogue for Theorem 4.3.

Let $F(x, z, p, t) = \alpha$, $G(x, z, p, t) = \beta$, α, and β constants, be two first integrals for (4.6). The Jacobi identity, (see Theorem III.3.1) is

$$\begin{aligned}&\{F, \{G, \mathcal{H}\}\} + \{G, \{\mathcal{H}, F\}\} \\ &\quad + \{\mathcal{H}, \{F, G\}\} + F_z\{G, \mathcal{H}\} + G_z\{\mathcal{H}, F\} + \mathcal{H}_z\{F, G\} = 0.\end{aligned}$$

Now replace the expressions for $\{F, \mathcal{H}\}$ and $\{G, \mathcal{H}\}$ using (4.7), expand the resulting expression using Theorem III.3.1, and rearrange the result to obtain the identity

$$-\frac{\partial}{\partial t}\{F, G\} - \{\{F, G\}, \mathcal{H}\} + \mathcal{H}\frac{\partial}{\partial z}\{F, G\} + \frac{\partial \mathcal{H}}{\partial z}\{F, G\} = 0.$$

We can rewrite this identity as

$$\frac{d}{dt}(\rho\{F, G\}) = 0$$

where $\rho = \exp\left(-\int_0^t \frac{\partial \mathcal{H}}{\partial z}\, d\tau\right)$ and conclude that along a solution, $\rho\{F, G\}$ is a constant, that is,

Theorem 4.6. *(Jacobi-Poisson). If $F(x, z, p, t) = \alpha$, $G(x, z, p, t) = \beta$ are first integrals for (4.6), (α and β constants), then $\rho\{F, G\}$ is also a first integral for (4.6).*

§ 4.5 The Theorems of Liouville and Poincaré

The theorems in this section deal with properties of the solutions to the canonical system

$$(5.1) \qquad \begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j} & j = 1, \ldots, n, \\[2ex] \dot{p}_j = -\dfrac{\partial \mathcal{H}}{\partial x_j} & j = 1, \ldots, n, \end{cases}$$

where $\mathcal{H} = \mathcal{H}(x, p)$. The initial conditions for (5.1) are

$$(5.2) \qquad \begin{cases} x_j(0) = x_j^0, & j = 1, \ldots, n, \\ p_j(0) = p_j^0, & j = 1, \ldots, n. \end{cases}$$

We have proven that for each t, the family of transformations

$$(x, p) = S_t(x^0, p^0) \tag{5.3}$$

is a contact transformation and the Jacobian determinant

$$\frac{\partial(x,p)}{\partial(x^0,p^0)} \neq 0.$$

However, we can in the present circumstances be more precise. For this computation, it is simplest to introduce the notation

$$\begin{aligned} \xi &= (\xi_1, \dots, \xi_{2n}) = (x_1, \dots, x_n, p_1, \dots, p_n) \\ \xi^0 &= (\xi_1^0, \dots, \xi_{2n}^0) = (x_1^0, \dots, x_n^0, p_1^0, \dots, p_n^0). \end{aligned}$$

Denote the solution, (5.3), by

$$\phi = \phi(\xi^0, t), \tag{5.4}$$

and the inverse by

$$\Phi = \Phi(\xi, t). \tag{5.5}$$

Let

$$J(\xi^0, t) = \frac{\partial(\phi(\xi^0, t))}{\partial(\xi^0)} = \det\left(\frac{\partial \phi_i(\xi^0, t)}{\partial \xi_j^0}\right) \tag{5.6}$$

and the inverse by

$$J^{-1}(\xi, t) = \frac{\partial(\Phi(\xi, t))}{\partial(\xi)} = \det\left(\frac{\partial \Phi_i(\xi, t)}{\partial \xi_j}\right) \tag{5.7}$$

and recall that if $\mathfrak{X}$ and $\mathfrak{Y}$ are two matrices,

$$\det(\mathfrak{X}\mathfrak{Y}) = \det(\mathfrak{X}) \det(\mathfrak{Y}).$$

Now use the formula for the derivative of a determinant to conclude

$$J_t J^{-1} = \sum_{i,j=1}^{2n} \frac{\partial \dot{\phi}_i}{\partial \xi_j^0} \frac{\partial \Phi_i}{\partial \xi_j} \tag{5.8}$$

where, as usual, the subscript t refers to the partial derivative with respect to t and the dot refers to the time derivative. $\phi_i = \phi_i(\xi^0, t)$ and so

$$\frac{d}{dt}\phi_i(\xi^0, t) = \dot{\phi}_i(\xi^0, t) = \dot{\phi}_i\big(\Phi(\xi, t), t\big).$$

Whence

$$\frac{\partial \dot{\phi}_i}{\partial \xi_j} = \sum_{k=1}^{2n} \frac{\partial \dot{\phi}_i}{\partial \xi_k^0} \frac{\partial \Phi_k}{\partial \xi_j}$$

and so

$$J_t J^{-1} = \sum_{i=1}^{2n} \frac{\partial \dot{\phi}_i}{\partial \xi_j}.$$

Finally we make use of (5.1) and the definition of ξ to conclude that

$$\begin{aligned}\sum_{i=1}^{2n} \frac{\partial \dot{\phi}_i}{\partial \xi_i} &= \sum_{i}^{n} \frac{\partial \dot{\phi}_i}{\partial \xi_i} + \sum_{i=n+1}^{2n} \frac{\partial \dot{\phi}_i}{\partial \xi_i} \\ &= \sum_{i}^{n} \frac{\partial^2 \mathcal{H}(x,p)}{\partial x_i \partial p_i} - \sum_{i}^{n} \frac{\partial^2 \mathcal{H}(x,p)}{\partial p_i \partial x_i} = 0\end{aligned},$$

that is $J_t J^{-1} = 0$ or $J_t = 0$. This implies that J is independent of t, that is

$$J(\xi^0, t) = J(\xi^0, 0) = 1,$$

since $\phi(\xi^0, 0) = \xi^0$ is the identity transformation.

Theorem 5.1.

$$J(\xi^0, t) = 1.$$

Next let us envision the initial conditions for (5.1) as being taken from some bounded domain, V_0, in the $\xi^0 = (x^0, p^0)$ space. The domain V_0, will be transformed to a domain

$$V_t = \phi(V_0, t) = S_t(V_0).$$

Theorem 5.1 implies that the measure or volume, $|V_t|$, remains constant.

Theorem 5.2. *(Liouville)*

$$|V_t| = |V_0|.$$

Proof.

$$|V_t| = \int_{V_t} dx\, dp = \int_{V_0} J\, dx^0\, dp^0 = \int_{V_0} dx^0\, dp^0 = |V_0|.$$

□

REMARK: If we envision $\{S_t\}$ as generating a flow, then Theorem 5.2 states that the flow is volume preserving.

Theorem 5.2 has another consequence which is due to Liouville. To formulate it, we need to introduce the concept of a density function, $\delta = \delta(\xi, t) = \delta(x, p, t)$ which describes the number of particles in a bounded domain, V_t. Liouvilles theorem states that this number is a constant, and that $\delta(x, p, t)$ is a constant of motion. More precisely,

Corollary 5.3. *(Liouville)*

$$\frac{d}{dt}\int_{V_t} \delta(x,p,t)\,dx\,dp = 0,$$

and

$$\delta_t + [\![\delta, \mathcal{H}]\!] = 0.$$

PROOF. Let $\delta_0 = \delta(x^0, p^0, 0)$.

first we find that

$$\int_{V_t} \delta(x,p,t)\,dx\,dp = \int_{V_0} \delta(x^0,p^0,t)\,J\,dx^0\,dp^0 = \int_{V_0} \delta_0\,dx^0\,dp^0 \tag{5.9}$$

which is a constant, and the first part of Liouville's theorem follows. Next, divide (5.9) by $|V_0|$ and use Theorem 5.2 to conclude that

$$\delta(x,p,t) = \delta_0,$$

and δ_0 is independent of t. We calculate

$$0 = \frac{d}{dt}\delta(x,p,t) = \delta_t + \sum_j^n \left(\frac{\partial\delta}{\partial x_j}\dot{x}_j + \frac{\partial\delta}{\partial p_j}\dot{p}_j\right)$$

$$= \delta_t + \sum_j^n \left(\frac{\partial\delta}{\partial x_j}\frac{\partial\mathcal{H}}{\partial p_j} - \frac{\partial\delta}{\partial p_j}\frac{\partial\mathcal{H}}{\partial x_j}\right) = \delta_t + [\![\delta, \mathcal{H}]\!]$$

which completes the proof. □

Liouvilles theorem can be combined with the following theorem to obtain some surprising information about systems governed by (5.1).

Theorem 5.4. *(Poincaré Recurrence Theorem) Suppose S is a continuous, one-to-one, volume preserving mapping of a bounded set $\mathfrak{M}$ in a Euclidean space onto itself, i.e. $S\mathfrak{M} = \mathfrak{M}$. Suppose ξ^0 is an arbitrary point in $\mathfrak{M}$ and U is any neighborhood of ξ^0. Then there is a point $\xi \in U$ and an integer $k \geq 1$ such that $S^k\xi \in U$.*

PROOF. Let U be the neighborhood of ξ^0, and consider the sets U, SU, S^2U, ... all of which lie in $\mathfrak{M}$. By hypothesis, they all have the same nonempty volume. Thus, nonzero intersections must now occur. For if they did not $\mathfrak{M}$ would be unbounded, which contradicts our assumption. Consequently, there are integers $j \geq 0$, $l \geq 0$, such that

$$S^{j+l}U \bigcap S^j U \neq \emptyset,$$

and whence

$$S^l U \bigcap U \neq \emptyset.$$

Suppose

$$\eta \in S^l U \bigcap U.$$

Then there is a $\xi \in U$ such that

$$\eta = S^l \xi \in S^l U \bigcap U \subset U.$$

The theorem follows with $k = l$.

Example 5.1. Let $\mathfrak{M}$ be a unit circle in two dimensional space, and let S_α denote a rotation in the counterclockwise direction through an angle α. There are two cases to consider.

i) Suppose $\alpha = 2\pi r$, where $r = m/n$ is a rational number. Then S_α^n is the identity and Poincaré's theorem is obvious.
ii) Suppose $\alpha = 2\pi r$, where r is irrational. Then starting with any point θ on the unit circle, a certain power $S^n\theta$ will be arbitrarily close to any given point, that is the set $\{S^n\theta\}$ is dense on the circle.

Observe that if $x = x(t)$, $p = p(t)$ is a solution to (5.1) then

(5.10)
$$\frac{d}{dt}\mathcal{H}(x,p) = [\![\mathcal{H},\mathcal{H}]\!] = 0 \qquad \text{or} \qquad \mathcal{H}(x,p) = \mathcal{H}(x^0,p^0) \qquad \text{a constant.}$$

We shall find in our applications to mechanical systems that $\mathcal{H}(x,p)$ represents the energy of the system and (5.10) is simply the statement that energy is conserved. The function $\mathcal{H}$ in that context is called the **Hamiltonian**. Theorem 5.2 asserts that the solution operator for (5.1) is volume preserving. As the set $\mathfrak{M}$, we can take $\{(x,p)\colon |\mathcal{H}(x,p)| \leq E < \infty\}$, which is a bounded set in xp-space. Theorem 5.4 contains, therefore, the following theorem as a special case.

Theorem 5.5. *(Poincaré). A mechanical system governed by (5.1) enclosed in a bounded set in xp-space and having finite energy, will return to an arbitrarily small neighborhood of almost any given initial state, in a finite period of time.*

§ 4.6 The Euler-Lagrange Equations and a Generalized Variational Calculus

In this section, we determine the relationship between the canonical systems and the classical Euler-Lagrange equations of mechanics and the calculus of variations. We proceed in a general way and the classical results will be special cases of the theory developed here.

Let us denote by

$$\mathcal{L} \equiv \mathcal{L}(x, \dot{x}, z, t) \equiv \mathcal{L}(x_1, \dots, x_n, \dot{x}_1, \dots, \dot{x}_n, z, t) \tag{6.1}$$

the Lagrange function, or Lagrangian, of the variables $(x, \dot{x}, z, t)$. The dot denotes, as usual, the time derivative of $x = x(t)$. While $z = z(t)$ is a scalar valued function of t. In classical mechanics, the Lagrangian is a function of $(x, \dot{x})$ and perhaps t, and is the difference between the kinetic and potential energies of the system. The variable z is to be determined as the solution to the differential equation

$$\dot{z} = \mathcal{L}(x, \dot{x}, z, t). \tag{6.2}$$

Observe that (6.2) represents a family of differential equations, since for each $x(t)$ a different differential equation arises, that is, given $x(t)$, $z(t)$ is determined by (6.2) so that $z(t)$ depends on $x(t)$. A fact which we make explicit by writing

$$z = z(x; t) = z(x, \dot{x}, t). \tag{6.3}$$

The problem (6.2) is a kind of control problem. The differential equation for z describes a process which depends on $(x, \dot{x})$ and which in turn can be chosen, that is they give us the opportunity to control or guide the process and are therefore referred to as controls. In the classical theory of the calculus of variations, $\mathcal{L}$ is independent of z and can be integrated immediately from 0 to T say, to obtain

$$z(x, \dot{x}, T) - z(x^0, \dot{x}^0, 0) = \int_0^T \mathcal{L}\big(x(t), \dot{x}(t), t\big)\, dt. \tag{6.4}$$

One then seeks stationary values for (6.4). We shall proceed similarly in our case. However, we need to work directly with (6.2).

To determine stationary values, we introduce the idea of the "rate of change" of $z(x; t)$ in the "direction" of a function, $\xi = (\xi_1, \dots, \xi_n)$. Let $x(t)$ denote a fixed curve, ϵ a real parameter, and ξ an "arbitrary" curve, also called a variation. Then $x + \epsilon\xi$ is a curve which for small ϵ is close to x. The rate of change of z in the direction of ξ, also called the first variation of z, is defined to be

$$\frac{d}{d\epsilon} z(x + \epsilon\xi, \dot{x} + \epsilon\dot{\xi}, t)|_{\epsilon=0} \equiv \zeta(t) \tag{6.5}$$

Using the definition (6.5), we find from (6.2) that ζ satisfies the differential equation

$$\dot{\zeta} = \sum_j^n \left(\frac{\partial \mathcal{L}}{\partial x_j}\xi_j + \frac{\partial \mathcal{L}}{\partial \dot{x}_j}\dot{\xi}_j \right) + \frac{\partial \mathcal{L}}{\partial z}\zeta$$

which is a linear differential equation for ζ.

The solution is

$$\begin{aligned} &\exp\left(-\int_0^t \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\zeta - \zeta^0 \\ &= \int_0^t \exp\left(-\int_0^\tau \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\sum_j^n \left[\frac{\partial \mathcal{L}}{\partial x_j}\xi_j + \frac{\partial \mathcal{L}}{\partial \dot{x}_j}\dot{\xi}_j\right] d\tau \end{aligned} \tag{6.6}$$

where ζ^0 means that ζ is to be evaluated at $t = 0$. To simplify the notation, let

$$\mathcal{L}_j = \frac{\partial \mathcal{L}}{\partial x_j}, \qquad p_j = \frac{\partial \mathcal{L}}{\partial \dot{x}_j}$$

and rewrite (6.6) using integration by parts as

$$\begin{aligned} &\exp\left(-\int_0^t \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\zeta - \zeta^0 \\ &= \exp\left(-\int_0^\tau \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\sum_j^n p_j\xi_j\Bigg|_0^t \\ &+ \int_0^t \exp\left(-\int_0^\tau \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\sum_j^n \left[\mathcal{L}_j + \mathcal{L}_z p_j - \dot{p}_j\right]\xi_j\,d\tau \end{aligned} \tag{6.7}$$

Now let us suppose that the values of x are prescribed at an initial state, $t = 0$, and at some final state at $t = T$. Then $z(x;t)$ can only be defined for those x which assume those prescribed values so that the variations ξ, must satisfy

$$\xi(0) = \xi(T) = 0 \tag{6.8}$$

Evaluate (6.7) at $t = T$ and use (6.8) to obtain

$$\exp\left(-\int_0^T \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\zeta(T) - \zeta^0 = \int_0^T \exp\left(-\int_0^\tau \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\sum_j^n [\mathcal{L}_j + \mathcal{L}_z p_j - \dot{p}_j]\,\xi_j\,d\tau \tag{6.9}$$

for the first variation, ζ, of z evaluated at $t = T$.

We seek functions, $x(t)$, such that $z(x;t)$ is stationary, that is, such that

$$\exp\left(-\int_0^t \frac{\partial \mathcal{L}}{\partial z}\,d\theta\right)\zeta = \zeta^0 \qquad \text{for} \quad 0 \le t \le T,$$

for all variations, ξ_j, so that the left hand side of (6.9) vanishes. By the fundamental lemma of the calculus of variations, this means that the integrand on the right hand side of (6.9) must vanish. Since the exponential never vanishes, we conclude that for each j

$$\mathcal{L}_j + \mathcal{L}_z p_j - \dot{p}_j = 0.$$

Thus the (x, z) must satisfy the following system of ordinary differential equations

$$\begin{cases} \dot{p}_j = \mathcal{L}_j + \mathcal{L}_z p_j, \qquad j = 1, \ldots, n \\ \\ \dot{z} = \mathcal{L} \qquad \text{with} \quad \mathcal{L}_j = \dfrac{\partial \mathcal{L}}{\partial x_j}, \quad p_j = \dfrac{\partial \mathcal{L}}{\partial \dot{x}_j}. \end{cases} \tag{6.10}$$

We prove the following

Theorem 6.1. *Let $\mathcal{L} = \mathcal{L}(x, \dot{x}, z, t)$ and suppose* $\det(\partial^2 \mathcal{L}/\partial \dot{x}_i \partial \dot{x}_j) \neq 0$. *Then the solutions to (6.10) determine a family of contact transformations. If $\mathcal{L}$ is independent of t, the family is a one parameter group.*

Proof. Denote the initial values for (6.10) by $(x^0, \dot{x}^0, z^0)$. Then the solution to (6.10) subject to these initial conditions will be given by

$$\begin{cases} x = x(x^0, \dot{x}^0, z^0, t) \\ \dot{x} = \dot{x}(x^0, \dot{x}^0, z^0, t) \\ z = z(x^0, \dot{x}^0, z^0, t) \end{cases} \tag{6.11}$$

where the dependence on the initial conditions is explicitly displayed. Now differentiate the equation $\dot{z} = \mathcal{L}$ with respect to t to get

$$\ddot{z} = \sum_j^n \mathcal{L}_j \dot{x}_j + \mathcal{L}_z \dot{z} + \sum_j^n p_j \ddot{x}_j + \mathcal{L}_t.$$

Set

$$\lambda = \exp\left(-\int_0^t \mathcal{L}_z \, d\tau\right) \tag{6.12}$$

and write this equation as

$$\begin{aligned} \frac{d}{dt}\left(\lambda \dot{z}\right) &= \lambda \sum_j^n [\mathcal{L}_j \dot{x}_j + p_j \ddot{x}_j] + \lambda \mathcal{L}_t \\ &= \lambda \sum_j^n \mathcal{L}_j \dot{x}_j - \sum_j^n \dot{x}_j \frac{d}{dt}(\lambda p_j) + \frac{d}{dt} \sum_j^n \lambda p_j \dot{x}_j + \lambda \mathcal{L}_t \\ &= \frac{d}{dt} \sum_j^n \lambda p_j \dot{x}_j + \lambda \mathcal{L}_t \end{aligned}$$

by (6.10). Hence

$$\frac{d}{dt}\left[\lambda \left(\dot{z} - \sum_j^n p_j \dot{x}_j\right)\right] = \lambda \mathcal{L}_t$$

and upon integration and division by λ

$$\sum_j^n p_j \dot{x}_j - \dot{z} = \frac{1}{\lambda}\left(\sum_j^n p_j^0 \dot{x}_j^0 - \dot{z}^0\right) - \frac{1}{\lambda}\int_0^t \lambda \mathcal{L}_\tau \, d\tau.$$

Let

$$\mathcal{H} = -\frac{1}{\lambda}\int_0^t \lambda \mathcal{L}_\tau \, d\tau \tag{6.13}$$

and rewrite this relationship in terms of differentials to find

$$\sum_j^n p_j dx_j - dz = \frac{1}{\lambda}\left(\sum_j^n p_j^0 dx_j^0 - dz^0\right) + \mathcal{H}\, dt. \tag{6.14}$$

Equation (6.14) characterizes the contact transformations which depend on a parameter, t. We must rewrite the transformations in terms of p instead of $\dot{x}$. For that it suffices to observe that the system

$$p_j = \frac{\partial \mathcal{L}}{\partial \dot{x}_j}(x, \dot{x}, z, t) \tag{6.15}$$

can be solved for $\dot{x}$ in terms of p because the Jacobi determinant is nonzero. This completes the proof of the theorem. □

The expression $\mathcal{H}$ is, at the moment, in an inconvenient form. We can put it in a more familiar form as follows. Let λ continue to be given by (6.12) and use (6.10) to find after some rearranging

$$\frac{d}{dt}(\lambda \mathcal{L}) = \frac{d}{dt}\sum_j^n \lambda \mathcal{L}_{\dot{x}_j}\dot{x}_j + \lambda \mathcal{L}_t$$

so that

$$\lambda \mathcal{L}_t = \frac{d}{dt}\left\{\lambda \mathcal{L} - \sum_j^n \lambda \mathcal{L}_{\dot{x}_j}\dot{x}_j\right\}.$$

By (6.13), we find

$$\mathcal{H} = \frac{1}{\lambda}\left\{\lambda \sum_j^n \mathcal{L}_{\dot{x}_j}\dot{x}_j - \lambda \mathcal{L} - \sum_j^n \mathcal{L}^0_{\dot{x}_j}\dot{x}^0_j + \mathcal{L}^0\right\}$$

where the superscripts indicate that the expressions are to be evaluated at the initial condition. Let us think of the $\mathcal{L}_{\dot{x}_j}$ as being replaced by the p_j so that

$$\mathcal{H} = \mathcal{H}(x, p, z, t) = \left(\sum_j^n p_j \dot{x}_j - \mathcal{L}\right) - \lambda^{-1}\left(\sum_j^n p^0_j \dot{x}^0_j - \mathcal{L}^0\right).$$

The last term is a function of t. We will be differentiating $\mathcal{H}$ with respect to the (x, p, z) variables to set up the canonical equations. This term will therefore contribute a zero. Thus, we define the function

$$\mathcal{H}(x, p, z, t) = \sum_j^n p_j \dot{x}_j - \mathcal{L}(x, \dot{x}, z, t), \qquad p_j = \frac{\partial \mathcal{L}}{\partial \dot{x}_j}. \tag{6.16}$$

This means that the previous expression for $\mathcal{H}$ must be replaced by $\mathcal{H} - \lambda^{-1}\mathcal{H}^0$.

The equations (6.10) can be written in terms of $\mathcal{H}$.

$$(6.17)\qquad \begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j} \\ \dot{z} = \displaystyle\sum_k^n p_k \frac{\partial \mathcal{H}}{\partial p_k} - \mathcal{H}, \\ \dot{p}_j = -\left(\dfrac{\partial \mathcal{H}}{\partial x_j} + p_j \dfrac{\partial \mathcal{H}}{\partial z}\right) \end{cases}$$

On the other hand, if one has a family of contact transformations, corresponding to it is a characteristic function, $\mathcal{H}$, and an associated canonical system.

We summarize these considerations in the following general statement.

The following four kinds of problems are equivalent:

i) Variational Problems
ii) Lagrangian Differential Equations
iii) Canonical Differential Equations
iv) One parameter Families of Contact Transformations

Example 6.1. Let $n = 1$ and define a Lagrangian function by

$$\mathcal{L} = \frac{m\dot{x}^2}{2} - \frac{lx^2}{2} - \alpha z, \qquad \text{where} \quad m, l, \alpha \quad \text{are all positive constants.}$$

Then

$$\begin{aligned} \mathcal{L}_{\dot{x}} &= m\dot{x} = p \\ \mathcal{L}_x &= -lx \\ \mathcal{L}_z &= -\alpha \end{aligned}$$

The Hamiltonian or characteristic function $\mathcal{H}$ is

$$\mathcal{H} = \mathcal{H}(x, p, z) = \frac{p^2}{2m} + \frac{lx^2}{2} + \alpha z.$$

The canonical system (6.17) is

$$\begin{aligned} \dot{x} &= \frac{p}{m} \\ \dot{z} &= \left(\frac{p^2}{2m} - \frac{lx^2}{2} - \alpha z\right) \\ \dot{p} &= -(lx + \alpha p) \end{aligned}$$

and the Lagrange equation, (6.10), is

$$m\ddot{x} = -lx - \alpha m\dot{x}$$

or

$$\ddot{x} + \alpha\dot{x} + \omega^2 x = 0, \qquad \text{where} \quad \omega^2 = \frac{l}{m}$$

which is the equation for a damped, harmonic oscillator.

Example 6.2. More generally, consider the Lagrangian in one spatial dimension

$$\mathcal{L}(x,\dot{x},z) = \frac{m\dot{x}^2}{2} - \frac{lx^2}{2} - f(z).$$

the system (6.10) is given by:

$$\mathcal{L}_x = -lx, \qquad \mathcal{L}_{\dot{x}} = m\dot{x} = p \qquad \mathcal{L}_z = -f'(z)$$

so

$$\begin{aligned} m\ddot{x} &= -lx - f'(z)m\dot{x} \\ \dot{z} &= \frac{m\dot{x}^2}{2} - \frac{lx^2}{2} - f(z). \end{aligned}$$

The Hamiltonian formulation is also easily obtained. Here

$$\mathcal{H}(x,\dot{x},z) = \frac{p^2}{2m} + \frac{lx^2}{2} + f(z)$$

and so

$$\begin{aligned} \dot{x} &= \frac{p}{m} \\ \dot{p} &= -lx - f'(z)p \\ \dot{z} &= \frac{p^2}{2m} - \frac{lx^2}{2} - f(z). \end{aligned}$$

§ 4.7 Partial Differential Equations and Canonical Systems

In this section we develop the one-to-one correspondence between partial differential equations of the first order and canonical systems based on the concept of a one-parameter group of contact transformations.

Let $\mathcal{H}$ be a function of $(2n+1)$ variables and consider the partial differential equation

$$\mathcal{H}(x,z,z_x) = 0 \tag{7.1}$$

where $x = x(x_1,\ldots,x_n)$ and $z_x = \nabla z = (\partial z/\partial x_1,\ldots,\partial z/\partial x_n)$ is the gradient of z. By a solution to (7.1) we mean a function

$$z = f(x) \tag{7.2}$$

which when inserted into $\mathcal{H}$ satisfies (7.1). In general, a function (7.2) which satisfies (7.1) is an n-dimensional, integral surface, E_n in $\mathbb{R}_{n+1}$. By adjoining direction quantities, we shall be able to think of E_n as an n-parameter union of elements. As in Chapter I, certain degenerate cases may arise. In $\mathbb{R}_3$ special cases of an integral surface were curves and points. In the case of $\mathbb{R}_{n+1}$, there are lower dimensional degeneracies, E_{n-1}, E_{n-2},

..., which may arise. Here we shall only treat the case when the integral surface is an n-dimensional object in $\mathbb{R}_{n+1}$.

We now take up the problem of finding a nondegenerate integral surface E_n, for (7.1). We shall think of E_n as given in terms of the u parameters $u = (u_1, \ldots, u_n)$. Thus, we seek functions

$$x = x(u), \qquad z = z(u), \qquad p = p(u), \tag{7.3}$$

such that

$$\mathcal{H}\big(x(u), z(u), p(u)\big) = 0 \tag{7.4}$$

where x, z, p satisfy

$$\sum_i^n p_i\, dx_i - dz = 0$$

or more explicitly

$$\sum_i^n p_i \frac{\partial x_i}{\partial u_k} - \frac{\partial z}{\partial u_k} = 0, \qquad k = 1, \ldots, n. \tag{7.5}$$

In order that functions (7.3) be independent and depend on the parameters $(u_1, \ldots, u_n)$ we must require that the rank of the $n \times (2n+1)$ matrix

$$\mathfrak{A} = \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_1} & \frac{\partial z}{\partial u_1} & \frac{\partial p_1}{\partial u_1} & \cdots & \frac{\partial p_n}{\partial u_1} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \frac{\partial x_1}{\partial u_n} & \cdots & \frac{\partial x_n}{\partial u_n} & \frac{\partial z}{\partial u_n} & \frac{\partial p_1}{\partial u_n} & \cdots & \frac{\partial p_n}{\partial u_n} \end{bmatrix}$$

be n. In view of (7.5), it is in fact only necessary to demand this from the $n \times 2n$ matrix obtained by deleting the column containing the $\partial z/\partial u_i$.

We now take the function $\mathcal{H}$ and set up the canonical system

$$\begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j}, \qquad j = 1, \ldots, n \\ \dot{z} = \displaystyle\sum_k^n p_k \frac{\partial \mathcal{H}}{\partial p_k} - \mathcal{H}, \\ \dot{p}_j = -\left(\dfrac{\partial \mathcal{H}}{\partial x_j} + p_j \dfrac{\partial \mathcal{H}}{\partial z}\right). \qquad j = 1, \ldots, n \end{cases} \tag{7.6}$$

Since $\mathcal{H}$ is independent of t, the solutions to (7.6) form a one parameter group of contact transformations, assuming of course that the initial conditions are chosen so that the identity transformation belongs to the family $\{S_t\}$. On a solution curve to (7.6),

$$\frac{d}{dt}\mathcal{H} = \sum_j^n \frac{\partial \mathcal{H}}{\partial x_j}\dot{x}_j + \frac{\partial \mathcal{H}}{\partial z}\dot{z} + \sum_j^n \frac{\partial \mathcal{H}}{\partial p_j}\dot{p}_j = -\mathcal{H}_z\mathcal{H}$$

so that

$$\mathcal{H} = \mathcal{H}^0 \exp\left(-\int_0^t \mathcal{H}_z\, dt\right) \tag{7.7}$$

where

$$\mathcal{H}^0 = \mathcal{H}(x^0, z^0, p^0). \tag{7.8}$$

If the initial element, (x^0, z^0, p^0) is chosen so that $\mathcal{H}^0 = 0$, then by (7.7) the elements

$$\big(x(t), z(t), p(t)\big) = S_t(x^0, z^0, p^0)$$

all satisfy $\mathcal{H} = 0$, that is, the elements (x, z, p) all lie on an orbital strip of the group which is characterized by $\mathcal{H} = 0$. All the elements satisfy, therefore, the partial differential equation. We come now to the main theorem of the theory.

Theorem 7.1. *Every integral surface E_n, of the partial differential equation (7.1) is transformed by the application of the group of contact transformations arising from (7.6) back onto itself.*

REMARK: In other words, if one has any element of the integral surface E_n, i.e. a point together with a tangent plane for which $\mathcal{H} = 0$, then the group determined by (7.6) generates a characteristic strip containing this element, and this strip lies in E_n.

PROOF. We must show that we can find functions $u = u(t)$ such that:

i) The element (x^0, z^0, p^0) corresponding to the initial values

$$u(0) = u^0$$

lies on E_n.

ii) The family of elements (x, z, p) determined by

$$u = u(t)$$

making up an orbital strip belongs to E_n. In other words, through every element of E_n, one can pass an orbital strip which has the same elements as E_n.

Let us formulate the conditions which $u(t)$ must satisfy more precisely. Differentiate (7.3) with respect to t to find for $j = 1, \ldots, n$

$$\dot{x}_j = \sum_k^n \frac{\partial x_j}{\partial u_k}\dot{u}_k, \qquad \dot{z} = \sum_k^n \frac{\partial z}{\partial u_k}\dot{u}_k, \qquad \dot{p}_j = \sum_k^n \frac{\partial p_j}{\partial u_k}\dot{u}_k. \tag{7.9}$$

Now make use of (7.6) and observe that the summand, $\mathcal{H}$, in the expression for $\dot{z}$ can be dropped since it is zero for the elements with which we are dealing. We find

$$\begin{cases} \mathfrak{X}_j \equiv \displaystyle\sum_k^n \frac{\partial x_j}{\partial u_k}\dot{u}_k - \frac{\partial \mathcal{H}}{\partial p_j} = 0 \qquad j = 1, \ldots, n \\ \mathfrak{Z} \equiv \displaystyle\sum_k^n \frac{\partial z}{\partial u_k}\dot{u}_k - \sum_k^n p_k \frac{\partial \mathcal{H}}{\partial p_k} = 0 \\ \mathfrak{P}_j \equiv \displaystyle\sum_k^n \frac{\partial p_j}{\partial u_k}\dot{u}_k + \frac{\partial \mathcal{H}}{\partial x_j} + p_j \frac{\partial \mathcal{H}}{\partial z} = 0. \qquad j = 1, \ldots, n \end{cases} \tag{7.10}$$

We are regarding the integral surface E_n, as known, so that only known functions of u occur in (7.10). Thus, (7.10) represents a system of $(2n+1)$ equations in n unknowns. They are not all independent of each other. First, if all the $\mathfrak{X}_j$ vanish, then by (7.5) and the definition of $\mathfrak{Z}$

$$0 = \sum_j^n p_j \mathfrak{X}_j = \sum_{j,k}^n p_j \frac{\partial x_j}{\partial u_k}\dot{u}_k - \sum_j^n p_j \frac{\partial \mathcal{H}}{\partial p_j} = \sum_k^n \frac{\partial z}{\partial u_k}\dot{u}_k - \sum_j^n p_j \frac{\partial \mathcal{H}}{\partial p_j} = \mathfrak{Z}.$$

Next the equations

$$\mathfrak{X}_j = 0, \qquad \mathfrak{P}_j = 0, \qquad j = 1, \ldots, n$$

represent a system of $2n$ equations in the n unknowns $\dot{u}_j$. The rank of the coefficient matrix is n by assumption, so we may select n equations which, possibly by renumbering, are given by

$$\begin{cases} \mathfrak{X}_j = 0, & j = 1, \ldots, r, \qquad 0 \le r \le n \\ \mathfrak{P}_j = 0, & j = 1, \ldots, s, \qquad 0 \le s \le n \\ & r + s = n \end{cases} \tag{7.11}$$

and are such that the coefficient matrix of $\dot{u}_1, \ldots, \dot{u}_n$ is nonsingular. Solve this system for $\dot{u}_1, \ldots, \dot{u}_n$. We obtain in this way n linearly independent differential equations for $u_1, \ldots, u_n$, which we solve subject to the initial conditions. To complete the proof of the theorem, we must show that the remaining equations

$$\begin{aligned} \mathfrak{X}_j &= 0, \qquad j = r+1, \ldots, n \\ \mathfrak{P}_j &= 0, \qquad j = s+1, \ldots, n. \end{aligned}$$

are satisfied.

There are three separate cases to consider: $r = n, s = 0$; $r = 0, s = n$; $1 \le r < n, 1 \le s < n$. We only consider the case $r = n, s = 0$. In this case the matrix $(\partial x_i/\partial u_j)$ is invertible. Its entries can be obtained by solving $x = x(u)$ for u in terms of x to obtain $u = u(x)$ and then forming $(\partial u_i/\partial x_j)$. The equation $\mathfrak{X}_j = 0$ then implies that

$$\dot{u}_j = \sum_k^n \frac{\partial u_j}{\partial x_k}\frac{\partial \mathfrak{X}}{\partial p_k}.$$

Consequently,

$$\begin{aligned} \mathfrak{P}_j &= \sum_k^n \frac{\partial p_j}{\partial u_k}\dot{u}_k + \frac{\partial \mathcal{H}}{\partial x_j} + p_j\frac{\partial \mathcal{H}}{\partial z} \\ &= \sum_{l,k}^n \frac{\partial p_j}{\partial u_k}\frac{\partial u_k}{\partial x_l}\frac{\partial \mathcal{H}}{\partial p_l} + \frac{\partial \mathcal{H}}{\partial x_j} + p_j\frac{\partial \mathcal{H}}{\partial z} \\ &= \sum_l^n \frac{\partial p_j}{\partial x_l}\frac{\partial \mathcal{H}}{\partial p_l} + \frac{\partial \mathcal{H}}{\partial x_j} + p_j\frac{\partial \mathcal{H}}{\partial z} \\ &= \frac{\partial}{\partial x_j}\mathcal{H}(x, z, p) = 0 \end{aligned}$$

since $\mathcal{H} = 0$ and $p_j = \partial z/\partial x_j$.

The other cases are treated similarly. This completes the proof of the theorem. □

These considerations yield the following corollaries.

Corollary 7.2. *If E_n is an integral surface for both partial differential equations*

$$\mathcal{H}(x, z, z_x) = 0 \qquad \textit{and} \qquad \mathcal{K}(x, z, z_x) = 0,$$

then it is an integral surface for the equation

$$L(x, z, z_x) = \{\mathcal{H}, \mathcal{K}\}_{xzp} = 0, \qquad p = z_x,$$

where $\{\mathcal{H}, \mathcal{K}\}_{xzp}$ is the Mayer bracket for $\mathcal{H}$ and $\mathcal{K}$.

PROOF. Simply form $\{\mathcal{H}, \mathcal{K}\}_{xzp}$ and use (7.6). □

Corollary 7.3. *Suppose E_n and E'_n are two distinct n dimensional integral surfaces of the differential equation*

$$\mathcal{H}(x, z, z_x) = 0.$$

Then the intersection in the set theoretic sense,

$$E_m = E_n \bigcap E'_n, \qquad m \leq n$$

is an integral surface for the differential equation.

We close this section with a discussion of a concrete integration problem. Suppose we have an $n-1$ dimensional integral surface, E_{n-1}, for the partial differential equation

$$\mathcal{H}(x, z, p) = 0, \qquad p = z_x, \tag{7.12}$$

that is we are given functions of $s = (s_1, \ldots, s_{n-1})$

$$\begin{cases} x^0 = x^0(s) \\ z^0 = z^0(s) \\ p^0 = p^0(s) \end{cases} \tag{7.13}$$

such that

$$\mathcal{H}(x^0, z^0, p^0) = 0 \tag{7.14}$$

and

$$\sum_j^n p_j^0(s) \frac{\partial x_j^0}{\partial s_k} - \frac{\partial z^0}{\partial s_k} = 0, \qquad k = 1, \ldots, n-1. \tag{7.15}$$

In a concrete case, the functions (x^0, z^0) are usually given and the functions, p^0 are chosen by the n conditions, (7.14) and (7.15). In order to guarantee the solvability of (7.14) and (7.15), we require that

$$\begin{vmatrix} \frac{\partial x_1^0}{\partial s_1} & \ldots & \frac{\partial x_1^0}{\partial s_{n-1}} & \frac{\partial \mathcal{H}(x^0,z^0,p^0)}{\partial p_1} \\ \vdots & \ldots & \vdots & \vdots \\ \frac{\partial x_n^0}{\partial s_1} & \ldots & \frac{\partial x_n^0}{\partial s_{n-1}} & \frac{\partial \mathcal{H}(x^0,z^0,p^0)}{\partial p_n} \end{vmatrix} \neq 0. \tag{7.16}$$

When $n = 2$, that is when the dimension of the space is 3, these conditions amount to giving data on a union of elements, that is a curve with direction coefficients. This problem was treated in detail in our geometric theory of partial differential equations. Here we treat the general case.

Set up the canonical equations

$$\begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j} \\ \dot{z} = \displaystyle\sum_k^n p_k \frac{\partial \mathcal{H}}{\partial p_k} - \mathcal{H} \\ \dot{p}_j = -\dfrac{\partial \mathcal{H}}{\partial x_j} - p_j \dfrac{\partial \mathcal{H}}{\partial z} \end{cases} \tag{7.17}$$

and solve them subject to the initial conditions

$$x(s,0) = x^0(s), \qquad z(s,0) = z^0(s), \qquad p(s,0) = p^0(s), \tag{7.18}$$

to obtain the solution

$$\big(x(s,t), z(s,t), p(s,t)\big) = S_t\big(x^0(s), z^0(s), p^0(s)\big). \tag{7.19}$$

This represents an n parameter family with $u = (s,t)$. We show that it is an integral manifold. This is most easily seen by noting that by (2.18)

$$\mathcal{H} = \rho \mathcal{H}(x^0, z^0, p^0) = 0$$

and

$$\sum_j^n p_j \, dx_j - dz = \rho \left(\sum_j^n p_j^0 \, dx_j^0 - dz^0 \right)$$

so that the solutions form a union of elements.

It is also possible to construct the solution, (7.2), to (7.1). For that we must require that the determinant

$$\left(\frac{\partial(x_1, \ldots, x_n)}{\partial(s_1, \ldots, s_n)} \right) \neq 0$$

at least near the initial manifold, that is

$$\left.\frac{\partial(x_1, \ldots, x_n)}{\partial(s_1, \ldots, s_n)}\right|_{t=0} = \left| \begin{matrix} \frac{\partial x_1}{\partial s_1} & \cdots & \frac{\partial x_1}{\partial s_{n-1}} & \frac{\partial x_1}{\partial t} \\ \vdots & \cdots & \vdots & \vdots \\ \frac{\partial x_n}{\partial s_1} & \cdots & \frac{\partial x_n}{\partial s_{n-1}} & \frac{\partial x_n}{\partial t} \end{matrix} \right|_{t=0}$$

$$= \left| \begin{matrix} \frac{\partial x_1^0}{\partial s_1} & \cdots & \frac{\partial x_1^0}{\partial s_{n-1}} & \frac{\partial \mathcal{H}(x^0, z^0, p^0)}{\partial p_1} \\ \vdots & \cdots & \vdots & \vdots \\ \frac{\partial x_n^0}{\partial s_1} & \cdots & \frac{\partial x_n^0}{\partial s_{n-1}} & \frac{\partial \mathcal{H}(x^0, z^0, p^0)}{\partial p_n} \end{matrix} \right|$$

is nonzero, which is precisely the condition (7.16). If that is the case, we can solve for (s,t) in terms of x and insert the result into the expression for z to obtain

$$z = z\big(s(x), t(x)\big) \equiv f(x). \tag{7.19}$$

Example 7.1. Solve

$$z_y(1+z_x) - z = x + y$$
$$z(x,0) = 1.$$

First, let

$$\mathcal{H}(x,y,z,p,q) = q(1+p) - z - x - y.$$

And let

$$x^0(s) = s, \qquad y^0(s) = 0, \qquad z^0(s) = 1$$

and determine p^0 and q^0 by

$$q^0(1+p^0) - 1 - s = 0$$
$$p^0\frac{dx^0}{ds} + q^0\frac{dy^0}{ds} - \frac{dz^0}{ds} = p^0 = 0$$

so

$$q^0 = 1 + s$$

The Jacobi determinant is

$$\begin{vmatrix} 1 & \mathcal{H}_p(x^0,y^0,z^0,p^0,q^0) \\ 0 & \mathcal{H}_q(x^0,y^0,z^0,p^0,q^0) \end{vmatrix} = 1 \neq 0.$$

The characteristic equations are

$$\begin{aligned}
&\dot{x} = q, && x(s,0) = s\\
&\dot{y} = 1 + p, && y(s,0) = 0\\
&\dot{z} = pq + q(1+p) - (q(1+p) - z - x - y), && z(s,0) = 1\\
&\dot{p} = 1 + p && p(s,0) = 0\\
&\dot{q} = 1 + q, && q(s,0) = 1 + s
\end{aligned}$$

The solution to the system is

$$\begin{aligned}
x(s,t) &= e^t(s+2) - t - 2,\\
y(s,t) &= e^t - 1,\\
z(s,t) &= e^{2t}(s+2) - e^t(s+4) + t + 3,\\
p(s,t) &= e^t - 1,\\
q(s,t) &= e^t(s+2) - 1.
\end{aligned}$$

The solution for z in terms of x and y is

$$z = -\ln(y+1) + 1 + (x + \ln(y+1) + 1 - y)y.$$

The fact that (7.19) satisfies (7.12) can be checked directly at this point as well. We summarize this discussion in

Theorem 7.4. *There exists a unique integral manifold satisfying (7.12), (7.13).*

If (7.16) holds, only x^0 and z^0 must be specified and p^0 is determined by (7.14) and (7.15). In that case the solution surface (7.2) can be constructed by solving (7.17) and (7.18), expressing (s,t) in terms of x and inserting that result into the expression for $z(s,t)$.

§ 4.8 Integration of the Canonical System given a Complete Integral of the Partial Differential Equation

In this section we take up the converse problem to that treated in the previous section. We wish to integrate the canonical system

$$\begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j}, & j = 1, \ldots, n \\ \\ \dot{p}_j = -\dfrac{\partial \mathcal{H}}{\partial x_j}, & j = 1, \ldots, n \end{cases} \tag{8.1}$$

where the Hamiltonian

$$\mathcal{H} = \mathcal{H}(x, p, t) \tag{8.2}$$

is given. We suppose that we can find a complete integral for the associated partial differential equation

$$z_t + \mathcal{H}(x, z_x, t) = 0. \tag{8.3}$$

Before proceeding with the construction, we must make precise what we mean by a complete integral for the partial differential equation (8.3). First, note that if z is a solution to (8.3), then $z + \text{const.}$ is also a solution.

We shall say that

$$z = f(x, t, X) + c, \qquad X = X(X_1, \ldots, X_n), \tag{8.4}$$

is a **complete integral** for (8.3), for all choices of the parameters (X, c) taken from a region in $\mathbb{R}_{n+1}$, if z satisfies (8.3) in the variables (x, t), and if

$$\det \left(\frac{\partial^2 f(x, t, X)}{\partial x_i \, \partial X_j} \right) \neq 0 \tag{8.5}$$

To obtain the general solution for (8.1), fix the parameters X. For definiteness, set

$$X_i = a_i, \qquad i = 1, \ldots, n$$

and now solve the system

$$\frac{\partial f(x,t,a)}{\partial X_j} = b_j, \qquad j = 1,\dots,n \tag{8.6}$$

for x in terms of (a,b,t), where the b_j are arbitrary constants. This is possible because of (8.5). We obtain in this way the functions

$$x_j = x_j(t) = \phi_j(a,b,t), \qquad j = 1,\dots,n \tag{8.7}$$

Next, define $p_j(t)$ by

$$p_j = p_j(t) = \psi_j(a,b,t) = \frac{\partial f\big(\phi(a,b,t),t,a\big)}{\partial x_j}, \qquad j = 1,\dots,n \tag{8.8}$$

We show that the x_j, p_j defined in this way satisfies (8.1). To see this, observe first that f satisfies by definition

$$f_t(x,t,X) + \mathcal{H}\big(x, f_x(x,t,X), t\big) = 0 \qquad \text{with} \qquad X = a. \tag{8.9}$$

Differentiate with respect to X_j and set $X = a$ to find

$$\frac{\partial^2 f}{\partial t \partial X_j} + \sum_k^n \frac{\partial \mathcal{H}}{\partial p_k} \frac{\partial^2 f}{\partial x_k \partial X_j} = 0. \tag{8.10}$$

Next, differentiate the identity

$$\frac{\partial}{\partial X_j} f\big(\phi(a,b,t),t,a\big) = b_j$$

with respect to t to obtain (since b_j is a constant),

$$\frac{\partial^2 f}{\partial X_j \partial t} + \sum_k^n \frac{\partial^2 f}{\partial X_j \partial x_k} \dot{\phi}_k = 0. \tag{8.11}$$

Subtract (8.10) from (8.11) to conclude

$$\sum_k^n \frac{\partial^2 f}{\partial x_k \partial X_j} \left\{ \dot{\phi}_k - \frac{\partial \mathcal{H}}{\partial p_k} \right\} = 0$$

and by (8.5), ϕ_k satisfies the first equation in (8.1).

Next, differentiate (8.9) with respect to x_j to find

$$\frac{\partial^2 f}{\partial x_j \partial t} + \frac{\partial \mathcal{H}}{\partial x_j} + \sum_k^n \frac{\partial \mathcal{H}}{\partial p_k} \frac{\partial^2 f}{\partial x_k \partial x_j} = 0, \qquad j = 1,\dots,n \tag{8.12}$$

By the definition (8.8),

$$\dot{p}_j = \frac{d}{dt}\frac{\partial f}{\partial x_j} = \frac{\partial^2 f}{\partial x_j \partial t} + \sum_k^n \frac{\partial^2 f}{\partial x_j \partial x_k}\dot{\phi}_k \tag{8.13}$$

Subtract (8.12) from (8.13) to get

$$\dot{p}_j = -\frac{\partial \mathcal{H}}{\partial x_j} + \sum_k^n \frac{\partial^2 f}{\partial x_j \partial x_k}\left[\dot{\phi}_k - \frac{\partial \mathcal{H}}{\partial p_k}\right] = -\frac{\partial \mathcal{H}}{\partial x_j}$$

since $\dot{\phi}_k = \partial\mathcal{H}/\partial p_k$ by what we have just proven. This is the second equation in (8.1).

We summarize this discussion in the following theorem.

Theorem 8.1. *(Jacobi) Suppose $z = f(x,t,a)+c$ is a complete integral for the partial differential equation*

$$z_t + \mathcal{H}(x,p,t) = 0, \qquad p = z_x.$$

Then the general solution for the canonical system

$$\dot{x}_j = \frac{\partial \mathcal{H}}{\partial p_j}$$
$$\dot{p}_j = -\frac{\partial \mathcal{H}}{\partial x_j}$$

is given by $x = \phi(a,b,t)$, $p = \psi(a,b,t)$ where ϕ is obtained as the solution to the system

$$\frac{\partial f(x,t,a)}{\partial X_j} = b_j, \qquad j = 1,\dots,n$$

for x in terms of (a,b,t) and ψ is defined by

$$\psi_j(a,b,t) = \frac{\partial f}{\partial x_j}(\phi(a,b,t),t,a).$$

The main weakness of the theory in this section is the determination of a complete integral. Nevertheless, if a complete integral can be found easily, the method can be very convenient.

Example 8.1. The system

$$\dot{x}_1 = p_1, \qquad \dot{x}_2 = p_2$$
$$\dot{p}_1 = 0 \qquad \dot{p}_2 = 0$$

arises from the Hamiltonian

$$\mathcal{H} = \frac{1}{2}(p_1^2 + p_2^2).$$

The associated partial differential equation is

$$z_t + \frac{1}{2}\left(\frac{\partial z}{\partial x_1}\right)^2 + \frac{1}{2}\left(\frac{\partial z}{\partial x_2}\right)^2 = 0.$$

The complete integral can be obtained by separation of variables. Assume a solution in the form

$$z = T(t) + S_1(x_1) + S_2(x_2)$$

so that

$$T'(t) + \frac{1}{2}\left([S_1'(x_1)]^2 + [S_2'(x_2)]^2\right) = 0.$$

Solve for $[S_1'(x_1)]^2$ to obtain

$$T'(t) + \frac{1}{2}[S_2'(x_2)]^2 = -\frac{1}{2}[S_1'(x_1)]^2.$$

The right and left hand sides vary independently of each other and so must be constant. With malice of forethought, let the constant be $-(1/2)a_1^2$ so that

$$S_1'(x_1) = a_1 \qquad \text{or} \qquad S_1(x_1) = a_1 x_1 + \alpha_1.$$

Next

$$T'(t) = -\frac{1}{2}[S_2'(x_2)]^2 - \frac{1}{2}a_1^2.$$

Again the right and left sides vary independently of each other and so must be constant. Denote the constant by $-(1/2)a_2^2$. Then

$$S_2'(x_2) = \sqrt{a_2^2 - a_1^2}$$

or

$$S_2(x_2) = \left[\sqrt{a_2^2 - a_1^2}\right] x_2 + \alpha_2$$

and

$$T(t) = -\frac{1}{2}a_2^2\, t + \alpha_3.$$

The complete integral is

$$z = a_1 x_1 + \left[\sqrt{a_2^2 - a_1^2}\right] x_2 - \frac{1}{2}a_2^2\, t + c,$$

where $c = \alpha_1 + \alpha_2 + \alpha_3$ is a constant.

Next set up the system

$$x_1 - \frac{a_1}{\sqrt{a_2^2 - a_1^2}}\, x_2 = b_1$$

$$\frac{a_2}{\sqrt{a_2^2 - a_1^2}}\, x_2 - a_2\, t = b_2$$

from which we obtain

$$\begin{aligned} x_1 &= \frac{a_1}{a_2}(a_2 t + b_2) + b_1 \\ x_2 &= \frac{\sqrt{a_2^2 - a_1^2}}{a_2}(a_2 t + b_2) \\ p_1 &= a_1 \\ p_2 &= \sqrt{a_2^2 - a_1^2}. \end{aligned}$$

If we set

$$\begin{aligned} \alpha_1 &= a_1, \\ \alpha_2 &= \sqrt{a_2^2 - a_1^2}, \\ \beta_1 &= \frac{a_1 b_2}{a_2} + b_1, \\ \beta_2 &= \frac{b_2\sqrt{a_2^2 - a_1^2}}{a_2}, \end{aligned}$$

we obtain the result

$$\begin{aligned} x_1 &= \alpha_1 t + \beta_1, \\ x_2 &= \alpha_2 t + \beta_2, \\ p_1 &= \alpha_1, \\ p_2 &= \alpha_2. \end{aligned}$$

Example 8.2. The Hamiltonian for the undamped Harmonic oscillator is

$$\mathcal{H}(x,p) = \frac{1}{2}(x^2 + p^2)$$

and the canonical equations are

$$\begin{aligned} \dot{x} &= p \\ \dot{p} &= -x. \end{aligned}$$

The associated partial differential equation is

$$z_t + \frac{1}{2}(z_x^2 + x^2) = 0.$$

The complete integral is again constructed by separation of variables Set

$$z(x,t) = T(t) + S(x).$$

Then

$$T'(t) = -\frac{1}{2}\left(\left[S'(x)\right]^2 + x^2\right) = -\frac{a^2}{2}$$

where $-a^2/2$ is the separation constant. The solutions for T and S are

$$\begin{aligned} T(t) &= -\frac{a^2}{2}t + c_1 \\ S(x) &= \frac{a^2}{2}\left(\arcsin\frac{x}{a} + \frac{x(a^2 - x^2)^{1/2}}{a^2}\right) + c_2 \end{aligned}$$

so that a complete integral is

$$z(x,t) = -\frac{a^2}{2}t + \frac{a^2}{2}\left(\arcsin\frac{x}{a} + \frac{x(a^2-x^2)^{1/2}}{a^2}\right) + c. \tag{8.14}$$

Differentiate with respect to a and set the result equal to b to obtain, after cancellation

$$-at + a\arcsin\frac{x}{a} = b,$$

whence

$$x = a\sin\left(t + \frac{b}{a}\right)$$

Now differentiate (8.14) with respect to x and insert the expression just obtained to find

$$p = \frac{a^2}{2\sqrt{a^2-x^2}} + \frac{1}{2}\sqrt{a^2-x^2} - \frac{x^2}{2\sqrt{a^2-x^2}} = \sqrt{a^2-x^2} = a\cos\left(t+\frac{b}{a}\right).$$

If we set $b/a = -\phi$, we obtain the solution for the harmonic oscillator in the standard form

$$x = a\sin(t-\phi), \qquad p = a\cos(t-\phi).$$

§ 4.9 Multiparameter Families of Contact Transformations

Let $x = (x_1, \ldots, x_n)$, $p = (p_1, \ldots, p_n)$ denote as usual points in $\mathbb{R}_n$ so that (x, z, p) is a point in a $(2n+1)$ dimensional space. $t = (t_1, \ldots, t_r)$ will denote a system of r parameters and $f = f(f_1, \ldots, f_n)$, g, and $h = (h_1, \ldots, h_n)$ are functions of (x^0, z^0, p^0, t). We call

$$\begin{cases} x = f(x^0, z^0, p^0, t) \\ z = g(x^0, z^0, p^0, t) \\ p = h(x^0, z^0, p^0, t) \end{cases} \tag{9.1}$$

an r parameter family of contact transformations if, for each fixed t, the functions f, g, and h satisfy the strip condition. It is often convenient to write the transformation (9.1) in the form

$$(x, z, p) = S_t(x^0, z^0, p^0) \tag{9.2}$$

to bring out the fact that the point (x^0, z^0, p^0) is carried into the point (x, z, p). We do not at this point demand that the family of transformations $\{S_t\}$ contains the identity, nor that (x^0, z^0, p^0) represent initial values. Rather, (x^0, z^0, p^0) is a generic point in the $(2n+1)$ dimensional space where the transformations are defined.

By repeating the arguments given in §2 for each parameter t_j, while holding the others fixed, we arrive at the following theorem.

Theorem 9.1. *If $\{S_t\}$ is an r parameter family of contact transformations, then there exists functions,*

$$\mathcal{H}_j = \mathcal{H}_j(x, z, p, t), \qquad j = 1, \ldots, r \tag{9.3}$$

such that the (x, z, p) of (9.1) satisfy the total canonical system

$$\begin{cases} dx_j = \displaystyle\sum_{k=1}^{r} \frac{\partial \mathcal{H}_k}{\partial p_j}\, dt_k, & j = 1, \ldots, n \\ dz = \displaystyle\sum_{k=1}^{r} \left\{ \sum_{j}^{n} p_j \frac{\partial \mathcal{H}_k}{\partial p_j} - \mathcal{H}_k \right\} dt_k & \\ dp_j = -\displaystyle\sum_{k=1}^{r} \left(\frac{\partial \mathcal{H}_k}{\partial x_j} + p_j \frac{\partial \mathcal{H}_k}{\partial z} \right) dt_k, & j = 1, \ldots, n. \end{cases} \tag{9.4}$$

The functions, $\mathcal{H}_j(x, z, p, t)$ of (9.3), characterize the particular family of contact transformations and are again called characteristic or Hamiltonian functions. Although they may be derived from (9.1) as indicated, in practical problems one is usually faced with the converse problem of constructing the family (9.1) or (9.2) from (9.4) given the (9.3). In order to carry out the integrations, the $\mathcal{H}_j$ must satisfy certain integrability conditions. To obtain them, it is convenient to rewrite the system (9.4) as

$$\begin{cases} \dfrac{\partial x_j}{\partial t_k} = \dfrac{\partial \mathcal{H}_k}{\partial p_j}, & j = 1, \ldots, n \quad k = 1, \ldots, r \\ \dfrac{\partial z}{\partial t_k} = \displaystyle\sum_{j}^{n} p_j \frac{\partial \mathcal{H}_k}{\partial p_j} - \mathcal{H}_k, & k = 1, \ldots, r \\ \dfrac{\partial p_j}{\partial t_k} = -\dfrac{\partial \mathcal{H}_k}{\partial x_j} - p_j \dfrac{\partial \mathcal{H}_k}{\partial z}, & j = 1, \ldots, n \quad k = 1, \ldots, r \end{cases} \tag{9.5}$$

To formulate the integrability conditions, it is advantageous to introduce one more bracket symbol. Let us continue to denote the Mayer bracket of two functions, F and G, by $\{F, G\}_{xzp}$. Define

$$[F, G]_{xzp} = \{F, G\}_{xzp} + F \frac{\partial G}{\partial z} - G \frac{\partial F}{\partial z}, \tag{9.6}$$

which when written out becomes

$$[F, G]_{xzp} = \sum_i^n \left(\frac{\partial F}{\partial x_i} \frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial x_i} \right) + \frac{\partial F}{\partial z} \left(\sum_i^n p_i \frac{\partial G}{\partial p_i} - G \right) - \frac{\partial G}{\partial z} \left(\sum_i^n p_i \frac{\partial F}{\partial p_i} - F \right).$$

This bracket symbol also satisfies the Jacobi identity

(9.7) $$[F, [G, H]] + [G, [H, F]] + [H, [F, G]] = 0$$

Now let us define the symbols

(9.8) $$\mathcal{H}_{kl} = [\mathcal{H}_k, \mathcal{H}_l]_{xzp} + \frac{\partial \mathcal{H}_k}{\partial t_l} - \frac{\partial \mathcal{H}_l}{\partial t_k}.$$

The integrability conditions require that the second mixed partials of the functions x_j, z, p_j with respect to the t variables are equal. An unpleasant calculation making use of (9.5) and the definition (9.8) yields the relations

(9.9) $$\begin{cases} \dfrac{\partial^2}{\partial t_l \partial t_k} x_j - \dfrac{\partial^2}{\partial t_k \partial t_l} x_j = \dfrac{\partial}{\partial p_j} \mathcal{H}_{kl} \\ \dfrac{\partial^2}{\partial t_l \partial t_k} p_j - \dfrac{\partial^2}{\partial t_k \partial t_l} p_j = -\dfrac{\partial}{\partial x_j} \mathcal{H}_{kl} + p_j \dfrac{\partial}{\partial z} \mathcal{H}_{kl} \\ \dfrac{\partial^2}{\partial t_l \partial t_k} z - \dfrac{\partial^2}{\partial t_k \partial t_l} z = \sum_j^n p_j \dfrac{\partial}{\partial p_j} \mathcal{H}_{kl} - \mathcal{H}_{kl}. \end{cases}$$

In order to force the right hand sides to be zero in these expressions, we see that the $\mathcal{H}_{kl}$ must vanish, which in view of (9.8) says

(9.10) $$[\mathcal{H}_k, \mathcal{H}_l]_{xzp} = \frac{\partial \mathcal{H}_l}{\partial t_k} - \frac{\partial \mathcal{H}_k}{\partial t_l},$$

which are the integrability conditions.

To formulate the next result, we need the following remark. Suppose

$$F = F(x, z, p, t)$$

where (x, z, p) satisfy (9.4) or equivalently, (9.5). Calculate the differential using (9.4) to obtain

$$
\begin{aligned}
dF &= \sum_j^n \frac{\partial F}{\partial x_j}\, dx_j + \frac{\partial F}{\partial z}\, dz + \sum_j^n \frac{\partial F}{\partial p_j}\, dp_j + \sum_{i=1}^r \frac{\partial F}{\partial t_i}\, dt_i \\
&= \sum_j^n \sum_{i=1}^r \frac{\partial F}{\partial x_j}\frac{\partial \mathcal{H}_i}{\partial p_j}\, dt_i + \frac{\partial F}{\partial z}\sum_{i=1}^r \left\{ \sum_j^n p_j \frac{\partial \mathcal{H}_i}{\partial p_j} - \mathcal{H}_i \right\} dt_i \\
&\quad - \sum_j^n \sum_{i=1}^r \frac{\partial F}{\partial p_j}\left(\frac{\partial \mathcal{H}_i}{\partial x_j} + p_j \frac{\partial \mathcal{H}_i}{\partial z} \right) dt_i + \sum_{i=1}^r \frac{\partial F}{\partial t_i}\, dt_i
\end{aligned}
$$

We conclude after some rearranging

$$
dF = \sum_{i=1}^r \left\{ [F, \mathcal{H}_i]_{xzp} - F\frac{\partial \mathcal{H}_i}{\partial z} + \frac{\partial F}{\partial t_i} \right\} dt_i \tag{9.11}
$$

or in terms of components

$$
\frac{\partial F}{\partial t_i} = [F, \mathcal{H}_i]_{xzp} - F\frac{\partial \mathcal{H}_i}{\partial z} + \frac{\partial F}{\partial t_i}. \tag{9.12}
$$

If we now calculate the second derivative and form the difference, we find

$$
\frac{\partial^2 F}{\partial t_l \partial t_k} - \frac{\partial^2 F}{\partial t_k \partial t_l} = [F, \mathcal{H}_{kl}]_{xzp} - F\frac{\partial \mathcal{H}_{kl}}{\partial z} \tag{9.13}
$$

We now state and prove the converse of Theorem 9.1.

Theorem 9.2. *Suppose the total canonical system (9.4) is given where the characteristic functions satisfy the integrability conditions, (9.10). Then the family of transformations $\{S_t\}$ obtained by solving (9.4) subject to the initial conditions*

$$
(x, z, p)\Big|_{t=0} = (x^0, z^0, p^0)
$$

is an r-parameter family of contact transformations.

The proof is similar to the one parameter case, see §IV.2.

We again define the linear differential form

$$
\omega = \sum_j^n p_j\, dx_j - dz - \sum_{i=1}^r \mathcal{H}_i\, dt_i \tag{9.14}
$$

when $t = 0$, i.e. $t = (t_1, \ldots, t_n) = (0, \ldots, 0)$, ω goes over into

$$\omega^0 = \sum_j^n p_j^0 \, dx_j^0 - dz^0 \tag{9.15}$$

We apply the arguments leading to equation (IV.2.15) for each t_i to obtain

$$\frac{\partial \omega}{\partial t_i} = -\frac{\partial \mathcal{H}_i}{\partial z}\omega, \qquad i = 1, \ldots, r \tag{9.16}$$

and consequently the total differential equation

$$d\omega = -\omega \sum_{i=1}^r \frac{\partial \mathcal{H}_i}{\partial z} \, dt_i \tag{9.17}$$

This equation is integrable because it satisfies (9.13) by hypothesis, i.e.

$$\frac{\partial^2 \omega}{\partial t_k \partial t_i} - \frac{\partial^2 \omega}{\partial t_i \partial t_k} = [\omega, \mathcal{H}_{ik}]_{xzp} - \omega \frac{\partial}{\partial z} \mathcal{H}_{ik} = 0$$

Now let t be a permissible value for the functions in question. We determine the function

$$\rho = \rho(x^0, z^0, p^0, t)$$

from the equation

$$\ln \rho = -\int_{\Gamma[0,t]} \sum_{k=1}^r \frac{\partial \mathcal{H}_k}{\partial z} \, dt_k$$

where the integral is taken over a path, Γ, connecting 0 and t. Because of the integrability conditions, the integral is independent of the path. Exponentiate to find for ρ the expression

$$\rho = \exp\left[-\int_{\Gamma[0,t]} \sum_{k=1}^r \frac{\partial \mathcal{H}_k}{\partial z} \, dt_k\right] \tag{9.18}$$

and set

$$\omega = \rho \omega^0. \tag{9.19}$$

By carrying out the differentiations, it is easy to verify that ω defined by (9.19) satisfies the total differential equation (9.4). But (9.19) is simply

$$\sum_j^n p_j \, dx_j - dz = \rho \left(\sum_j^n p_j^0 \, dx_j^0 - dz^0 \right) + \sum_{i=1}^r \mathcal{H}_i \, dt_i$$

that is, the strip condition holds, which completes the proof of the assertion. □

We close this section with a derivation of several formulas which will be of use to us in the sequel, and a discussion of two special cases similar to the ones of §IV.2.

To begin with, observe that from (9.18)

$$d\rho = -\rho \sum_{k=1}^{r} \frac{\partial \mathcal{H}_k}{\partial z}\, dt_k \tag{9.20}$$

so that, combining (9.20) with (9.12) where F is replaced by ρ, we obtain

$$\sum_{j=1}^{r} \left([\rho, \mathcal{H}_j]_{xzp} + \frac{\partial \rho}{\partial t_j} \right) dt_j = 0 \tag{9.21}$$

Next, calculate using (9.12) and (9.20) again to get

$$\begin{aligned} F\, d\ln(\frac{F}{\rho}) &= F\, d(\ln F - \ln \rho) = dF - \frac{F}{\rho}\, d\rho \\ &= \sum_{i=1}^{r} \left([F, \mathcal{H}_i]_{xzp} - F\frac{\partial \mathcal{H}_i}{\partial z} + \frac{\partial F}{\partial t_i} \right) dt_i - \frac{F}{\rho} \left(-\rho \sum_{i=1}^{r} \frac{\partial \mathcal{H}_i}{\partial z}\, dt_i \right) \\ &= \sum_{i=1}^{r} \left([F, \mathcal{H}_i]_{xzp} + \frac{\partial F}{\partial t_i} \right) dt_i \end{aligned}$$

and so we find the general formula

$$F\, d\ln(\frac{F}{\rho}) = \sum_{i=1}^{r} \left([F, \mathcal{H}_i]_{xzp} + \frac{\partial F}{\partial t_i} \right) dt_i \tag{9.22}$$

In particular if F is replaced by $\mathcal{H}_k$, we get

$$\mathcal{H}_k\, d\ln(\frac{\mathcal{H}_k}{\rho}) = \sum_{i=1}^{r} \left([\mathcal{H}_k, \mathcal{H}_i]_{xzp} + \frac{\partial \mathcal{H}_k}{\partial t_i} \right) dt_i \tag{9.23}$$

Now by (9.8) and the fact that $\mathcal{H}_{ki} = 0$ by the integrability conditions,

$$[\mathcal{H}_k, \mathcal{H}_i]_{xzp} + \frac{\partial \mathcal{H}_k}{\partial t_i} = \mathcal{H}_{ki} + \frac{\partial \mathcal{H}_i}{\partial t_k} = \frac{\partial \mathcal{H}_i}{\partial t_k},$$

we see that (9.20) becomes

$$\mathcal{H}_k \, d\ln\left(\frac{\mathcal{H}_k}{\rho}\right) = \sum_{i=1}^{r} \frac{\partial \mathcal{H}_i}{\partial t_k}\, dt_i \tag{9.24}$$

For ease in reference, we gather the more important of these formulas together in

Theorem 9.3. *Let ρ, $\mathcal{H}_k$, F be as above. Then*

$$F\, d(F/\rho) = \sum_{i=1}^{r} \left([F, \mathcal{H}_i]_{xzp} + \frac{\partial F}{\partial t_i} \right) dt_i$$

$$\sum_{i}^{n} \left([\rho, \mathcal{H}_i]_{xzp} + \frac{\partial \rho}{\partial t_i} \right) dt_i = 0$$

$$\mathcal{H}_k \, d\ln\left(\frac{\mathcal{H}_k}{\rho}\right) = \sum_{i=1}^{r} \left(\frac{\partial \mathcal{H}_i}{\partial t_k}\right) dt_i.$$

We now turn to the special cases.

Case 1.

$$\frac{\partial \mathcal{H}_i}{\partial z} = 0, \qquad i = 1, \ldots, r.$$

The functions, $\mathcal{H}_i$, are all independent of z and the total canonical system (9.4) reduces to the $2n$ equations

$$\begin{cases} dx_j = \displaystyle\sum_{k=1}^{r} \frac{\partial \mathcal{H}_k}{\partial p_j}\, dt_k, & j = 1, \ldots, n \\ dp_j = -\displaystyle\sum_{k=1}^{r} \frac{\partial \mathcal{H}_k}{\partial x_j}\, dt_k, & j = 1, \ldots, n \end{cases}$$

and once the x_j and p_j are calculated the z is obtained by quadrature.

$$z - z^0 = \sum_{k=1}^{r} \int_{\Gamma[0,t]} \left(\sum_{j}^{n} p_j \frac{\partial \mathcal{H}_k}{\partial p_j} - \mathcal{H}_k \right) dt_k.$$

In addition the function

$$\rho \equiv 1$$

and the strip condition reduces to

$$\sum_j^n p_j\,dx_j - \sum_j^n p_j^0\,dx_j^0 - \sum_{k=1}^r \mathcal{H}_k\,dt_k = d(z - z^0).$$

Case 2.

$$\frac{\partial \mathcal{H}_i}{\partial t_k} = 0, \qquad i, k = 1, \ldots, r.$$

The functions $\mathcal{H}_i$ depend only on (x, z, p) and are independent of the t variables. Theorem 9.3 (or equation 9.24) implies

$$\mathcal{H}_k\,d\ln(\mathcal{H}_k/\rho) = 0.$$

Expand and rewrite this equality to obtain

$$d\mathcal{H}_k = \mathcal{H}_k \frac{d\rho}{\rho}$$

or

$$d\ln \mathcal{H}_k = d\ln \rho. \tag{9.25}$$

Set $\mathcal{H}_k^0 = \mathcal{H}_k(x^0, z^0, p^0)$ and recall that at $t = 0$, $\rho = 1$. Integrate (9.25) to find

$$\ln \mathcal{H}_k - \ln \mathcal{H}_k^0 = \ln \rho,$$

that is

$$\mathcal{H}_k = \rho \mathcal{H}_k^0, \qquad k = 1, \ldots, r. \tag{9.26}$$

Case 3.

$$\frac{\partial \mathcal{H}_k}{\partial z} = 0 \qquad \text{and} \qquad \frac{\partial \mathcal{H}_k}{\partial t_l} = 0.$$

In this case,

$$\mathcal{H}_k = \mathcal{H}_k(x, p).$$

Again $\rho \equiv 1$ and by (9.26) we obtain immediately r integrals of the form

$$\mathcal{H}_k(x, p) = \mathcal{H}_k^0, \qquad k = 1, \ldots, r.$$

§ 4.10 The Infinitesimal Contact Transformations

Before introducing the concept of an infinitesimal contact transformation, we wish to list the properties of the bracket symbol

$$
\begin{aligned}
[F,G]_{xzp} = \sum_j^n \left(\frac{\partial F}{\partial x_j}\frac{\partial G}{\partial p_j} - \frac{\partial F}{\partial p_j}\frac{\partial G}{\partial x_j} \right) + \frac{\partial F}{\partial z}\left(\sum_j^n p_j \frac{\partial G}{\partial p_j} - G \right) \\
- \frac{\partial G}{\partial z}\left(\sum_j^n p_j \frac{\partial F}{\partial p_j} - F \right)
\end{aligned} \tag{10.1}
$$

introduced in the previous section. We group these properties together as

Theorem 10.1.

i) $[F+G,H] = [F,H]+[G,H]$ *and* $[\alpha F,G] = \alpha[F,G]$, $[F,\alpha G] = \alpha[F,G]$ *where* α *is a constant. In other words,* $[\cdot,\cdot]$ *is linear in each argument.*

ii) $[F,F] = 0$ *and* $[F,G] = -[G,F]$.

iii) $[Fg,H] = F[G,H] + G[F,H] - FG\frac{\partial H}{\partial z}$.

iv) $\frac{\partial}{\partial z}[F,G] = [\frac{\partial F}{\partial z},G] + [F,\frac{\partial G}{\partial z}] = [\frac{\partial F}{\partial z},G] - [\frac{\partial g}{\partial z},F]$.

v) $[F,[G,H]] + [G,[H,F]] + [H,[F,G]] = 0$ *(Jacobi identity).*

vi) *If* $(x,z,p) \longrightarrow (X,Z,P)$ *is a contact transformation with*

$$
\sum_j^n P_j\,dX_j - dZ = \rho\left(\sum_j^n p_j\,dx_j - dz \right),
$$

then

$$
\frac{1}{\rho}[\rho F,\rho G]_{XZP} = [F,G]_{xzp}.
$$

Remark: In the statement of this theorem, the subscripts, xzp, have been dropped except in item (vi). We will consistently drop the xzp subscripts in this section, and the bracket symbol will refer to the bracket defined by (10.1).

The proofs of these assertions are straightforward computations. The proof of (vi) can be simplified somewhat by using Theorem III.3.2 for the Mayer brackets.

After these preliminary remarks, let us consider again a one parameter family of contact transformations, which are characterized by means of the Hamiltonian, $\mathcal{H} = \mathcal{H}(x, z, p, t)$, as the solution to the canonical system

$$\tag{10.2} \begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j}, & j = 1, \ldots, n \\ \dot{z} = \displaystyle\sum_j^n p_j \frac{\partial \mathcal{H}}{\partial p_j} - \mathcal{H} \\ \dot{p}_j = -\dfrac{\partial \mathcal{H}}{\partial x_j} - p_j \dfrac{\partial \mathcal{H}}{\partial z}, & j = 1, \ldots, n \end{cases}$$

We call the quantities

$$\tag{10.3} \xi_j = \frac{\partial \mathcal{H}}{\partial p_j}, \qquad \zeta = \sum_j^n p_j \frac{\partial \mathcal{H}}{\partial p_j} - \mathcal{H}, \qquad \pi_j = -\left(\frac{\partial \mathcal{H}}{\partial x_j} + p_j \frac{\partial \mathcal{H}}{\partial z},\right)$$

$j = 1, \ldots, n$, the components of the transformation of the contact transformation generated by $\mathcal{H}$. It is customary in the theory of continuous groups to define the symbol, $\mathfrak{H}$, as a differential operator by setting

$$\tag{10.4} \mathfrak{H}(F) = \sum_j^n \xi_j \frac{\partial F}{\partial x_j} + \zeta \frac{\partial F}{\partial z} + \sum_j^n \pi_j \frac{\partial F}{\partial p_j},$$

where F is a differentiable function of x, z, p. In terms of the bracket symbols, $\mathfrak{H}$, can also be written

$$\tag{10.5} \mathfrak{H}(F) = [F, \mathcal{H}]_{xzp} - F \frac{\partial \mathcal{H}}{\partial z} = \{F, \mathcal{H}\}_{xzp} - \mathcal{H} \frac{\partial F}{\partial z}.$$

Let $\mathfrak{K}$ be the symbol for the Hamiltonian $\mathcal{K}$. The product of the differential operators is defined by

$$\tag{10.6} \mathfrak{K}\mathfrak{H}(F) = \mathfrak{K}\big(\mathfrak{H}(F)\big)$$

Making use of Theorem 10.1 we find

$$\begin{aligned} \mathfrak{K}\big(\mathfrak{H}(F)\big) &= \mathfrak{K}\Big([F, \mathcal{H}] - F\mathcal{H}_z\Big) \\ &= \Big[[F, \mathcal{H}] - F\mathcal{H}_z, \mathcal{K}\Big] - \Big([F, \mathcal{H}] - F\mathcal{H}_z\Big)\mathcal{K}_z \\ &= -\Big[\mathcal{K}, [F, \mathcal{H}]\Big] - [F\mathcal{H}_z, \mathcal{K}] - \mathcal{K}_z[F, \mathcal{H}] + F\mathcal{H}_z\mathcal{K}_z \end{aligned}$$

and similarly

$$\mathfrak{H}(\mathfrak{R}(F)) = \mathfrak{H}\Big([F,\mathcal{K}] - F\mathcal{K}_z\Big)$$
$$= -\Big[\mathcal{H},[F,\mathcal{K}]\Big] - [F\mathcal{K}_z,\mathcal{H}] - \mathcal{H}_z[F,\mathcal{K}] + F\mathcal{K}_z\mathcal{H}_z.$$

The commutator is the difference

$$\mathfrak{C}(F) = \mathfrak{R}\mathfrak{H}(F) - \mathfrak{H}\mathfrak{R}(F). \tag{10.7}$$

Making use of Theorem 10.1 and (iv), we obtain the explicit expression

$$\mathfrak{C}(F) = \Big[F,[\mathcal{H},\mathcal{K}]\Big] - F\frac{\partial}{\partial z}[\mathcal{H},\mathcal{K}] \tag{10.8}$$

From (10.8) it is immediately clear that

$$\mathfrak{C}(F) = 0 \tag{10.9}$$

is the condition that the infinitesimal transformations, $\mathfrak{H}$ and $\mathfrak{R}$, commute. We shall see below that (10.9) has a deeper significance for the relationship between the solutions to the canonical equations (10.1) generated by $\mathcal{H}$ and those generated by $\mathcal{K}$, i.e. those where in (10.1) the $\mathcal{H}$ is replaced by $\mathcal{K}$.

We would also like to make some remarks concerning the designation "infinitesimal contact transformation". This arises from the fact that the element (x,z,p) is transformed during a time span, ϵ, into the neighboring element $(x+\epsilon\xi, z+\epsilon\zeta, p+\epsilon\pi)$, where ϵ is thought of as being so small that terms multiplied by higher powers of ϵ may be neglected.

We now take up the multiparameter case and for this purpose consider the system of total differentiable equations

$$dy_j = \sum_{k=1}^{r} \mathcal{A}_{jk}\, dt_k, \qquad j = 1,\dots,m \tag{10.10}$$

where for $j = 1,\dots,m$, $k = 1,\dots,r$

$$\mathcal{A}_{jk} = \mathcal{A}_{jk}(y,t), \qquad y = (y_1,\dots,y_m), \qquad t = (t_1,\dots,t_r). \tag{10.11}$$

We shall later identify y with (x,z,p) and m will be $2n+1$. The $\mathcal{A}_{jk}$ are, as usual, assumed to be smooth functions on their domains of definition and to satisfy the integrability conditions

$$\frac{\partial^2 y_j}{\partial t_k \partial t_l} = \frac{\partial^2 y_j}{\partial t_l \partial t_k},$$

which when written out in terms of the $\mathcal{A}'s$, is by (10.10)

$$\frac{\partial \mathcal{A}_{jl}}{\partial t_k} - \frac{\partial \mathcal{A}_{jk}}{\partial t_l} + \sum_{i=1}^{m} \left(\mathcal{A}_{ik} \frac{\partial \mathcal{A}_{jl}}{\partial y_i} - \mathcal{A}_{il} \frac{\partial \mathcal{A}_{jk}}{\partial y_i} \right) = 0 \tag{10.12}$$

We can rewrite (10.12) more compactly as follows. Let

$$F = F(y; t)$$

be a differentiable function and set

$$\mathfrak{H}_k(F) = \frac{\partial F}{\partial t_k} + \sum_{j=1}^{m} \mathcal{A}_{jk} \frac{\partial F}{\partial y_j}. \tag{10.13}$$

The commutator is

$$\mathfrak{H}_k(\mathfrak{H}_l(F)) - \mathfrak{H}_l(\mathfrak{H}_k(F)) =$$
$$\sum_{j=1}^{m} \left\{ \frac{\partial \mathcal{A}_{jl}}{\partial t_k} - \frac{\partial \mathcal{A}_{jk}}{\partial t_l} + \sum_{i=1}^{m} \left(\mathcal{A}_{ik} \frac{\partial \mathcal{A}_{jl}}{\partial y_i} - \mathcal{A}_{il} \frac{\partial \mathcal{A}_{jk}}{\partial y_i} \right) \right\} \frac{\partial F}{\partial y_i}$$

and it vanishes for all F if and only if the integrability condition (10.12) holds; we conclude that (10.12) is equivalent to requiring

$$\mathfrak{H}_k(\mathfrak{H}_l(F)) - \mathfrak{H}_l(\mathfrak{H}_k(F)) = 0 \quad \text{for all} \quad F. \tag{10.14}$$

We specialize to the case $r = 1$. Consider the system of ordinary differential equations

$$\dot{y}_j = A_j(y, t), \qquad j = 1, \dots, m \tag{10.15}$$

We define the symbol, $\mathfrak{A}$, of an infinitesimal transformation by

$$\mathfrak{A}(F) = \sum_{i=1}^{m} A_i \frac{\partial F}{\partial y_i} + \frac{\partial F}{\partial t}, \tag{10.16}$$

where A_i is the right hand side of (10.15). Note that the symbol of an infinitesimal transformation is obtained by computing the total derivative of $F(y, t)$ with respect to t and making use of (10.15).

Now suppose the $A_i = A_i(y)$, that is the A_i are independent of t in (10.15) and let $\mathfrak{B}$ denote the symbol for the infinitesimal transformation generated by $(B_1(y), \dots, B_m(y))$, where again the $B_i = B_i(y)$. We say that $\mathfrak{B}$ is **compatible** with $\mathfrak{A}$ if, whenever y is a solution to (10.15), $y + \epsilon B(y)$ is as well, where ϵ is a parameter which ranges over some interval $\{|\epsilon| < \bar{\epsilon}\}$.

We may express this definition analytically by observing first that

$$\frac{d}{dt}\big(y_j + \epsilon B_j(y)\big) = A_j(y + \epsilon B)$$

so that

$$\dot{y}_j + \epsilon \sum_{k=1}^{m} \frac{\partial B_j}{\partial y_k} \dot{y}_k = A_j(y) + \epsilon \sum_{k=1}^{m} \frac{\partial B_j}{\partial y_k} A_k(y) = A_j(y + \epsilon B).$$

Now differentiate with respect to ϵ and set $\epsilon = 0$ to find

$$\sum_{k=1}^{m} \frac{\partial B_j}{\partial y_k} A_k = \sum_{k=1}^{m} \frac{\partial A_j}{\partial y_k} B_k, \tag{10.17}$$

or in terms of the symbols

$$\mathfrak{A}\big(\mathfrak{B}(F)\big) - \mathfrak{B}\big(\mathfrak{A}(F)\big) = \sum_{j=1}^{m} \left\{ \sum_{k=1}^{m} \left(A_k \frac{\partial B_j}{\partial y_k} - B_k \frac{\partial A_j}{\partial y_k} \right) \right\} \frac{\partial F}{\partial y_j} = 0$$

for all F, i.e.

$$\mathfrak{A}\mathfrak{B} - \mathfrak{B}\mathfrak{A} = 0.$$

(10.17) is the integrability condition for the system

$$dy_j = A_j\, dt_1 + B_j\, dt_2.$$

§ 4.11 Multiparameter Groups of Contact Transformations

The definition of a multiparameter group is modeled on that of a one parameter group. In this section, the letters t, s, etc., will denote the set of parameters $(t_1, \ldots, t_r)$, $(s_1, \ldots, s_r)$, etc. In the development which follows, the value $t = 0$ will correspond to the identity transformation. Strictly speaking these considerations take place in some neighborhood of $t = 0$, but that fact will be suppressed in the sequel.

Definition 11.1

An r-parameter group of contact transformations is a family, $\{S_t\}$, of contact transformations which satisfy the following conditions:

i) The family includes an element, S_0, called the identity.

ii) There is an operation called multiplication such that if S_t and S_s are elements of the family, there exists an element, S_σ, of the family such that

$$S_\sigma = S_t S_s.$$

This multiplication is determined by a smooth function

$$\phi = (\phi_1, \ldots, \phi_r)$$

of the variables (t, s).

iii) $S_t S_0 = S_0 S_t = S_t$, that is,

$$\phi(t, 0) = \phi(0, t) = t,$$

and the Jacobi determinant

$$\frac{\partial\big(\phi_1(t,s), \ldots, \phi_r(t,s)\big)}{\partial(t_1, \ldots, t_r)} \neq 0$$

for t, s near 0. In particular, $\phi(0,0) \neq 0$.

iv) The associative law holds, that is

$$S_t(S_s S_\sigma) = (S_t S_s) S_\sigma,$$

in other words, ϕ satisfies

$$\phi\big(t, \phi(s,\sigma)\big) = \phi\big(\phi(t,s), \sigma\big).$$

The condition (iii) implies the existence of an inverse, because the equation

$$S_\sigma S_t = S_0,$$

or more precisely

$$\phi(\sigma, t) = 0$$

is solvable for σ in terms of t. In operator notation, let

$$S_\sigma = S_t^{-1}$$

denote that solution. We must show that also

$$S_t S_\sigma = S_0.$$

For this calculation, let S_σ^* be such that $S_\sigma^* S_\sigma = S_0$. Then

$$S_t S_\sigma = S_0(S_t S_\sigma) = (S_\sigma^* S_\sigma)(S_t S_\sigma) = S_\sigma^*(S_\sigma S_t) S_\sigma = S_\sigma^* S_\sigma = S_0$$

so that $S_\sigma = S_t^{-1}$ is both a right and a left inverse, and the standard group axioms hold. S_t^{-1} is easily seen to be unique and moreover we find that

$$\frac{\partial\big(\phi(t,s)\big)}{\partial(s)} \neq 0$$

for t, s near 0.

After these preliminaries, we state the main theorem of this section.

Theorem 11.1. *In order that an r-parameter family, $\{S_t\}$ of contact transformations be a group, it is necessary and sufficient that the characteristic functions, $\mathcal{H}_k$, have the form*

$$\mathcal{H}_k = \mathcal{H}_k(x,z,p,t) = \sum_{i=1}^{r} \mathcal{K}_i(x,z,p)\omega_{ik}(t), \qquad k = 1,\ldots,r. \tag{11.1}$$

Here the $\mathcal{K}_i$ are independent of t and the ω_{ik} depend only on t. Moreover, the functions, $\mathcal{K}_1, \ldots \mathcal{K}_r$ are linearly independent, and the determinant of the $r \times r$ matrix (ω_{ik}) is nonzero.

Before giving a proof of this theorem, we first make a few observations and introduce some notation.

Let us set

$$d\omega_i = \sum_k^r \omega_{ik}\, dt_k, \qquad i = 1,\ldots,r \tag{11.2}$$

so that (11.1) takes the form

$$\sum_k^r \mathcal{H}_k\, dt_k = \sum_i^r \mathcal{K}_i\, d\omega_i. \tag{11.3}$$

The differential form

$$\sum_k^r \mathcal{H}_k\, dt_k$$

is integrable and by (9.10) the integrability conditions are

$$[\mathcal{H}_k, \mathcal{H}_l]_{xzp} = \frac{\partial \mathcal{H}_l}{\partial t_k} - \frac{\partial \mathcal{H}_k}{\partial t_l}. \tag{11.4}$$

If the $\mathcal{H}_k$ are given by (11.1), then (11.4) has the form

$$\sum_{k,l}^r [\mathcal{K}_k, \mathcal{K}_l]_{xzp}\, \omega_{k\alpha}\, \omega_{l\beta} = \sum_k^r \mathcal{K}_k \left[\frac{\partial \omega_{k\beta}}{\partial t_\alpha} - \frac{\partial \omega_{k\alpha}}{\partial t_\beta}\right]. \tag{11.5}$$

Since $\det(\omega_{ij}) \neq 0$, the matrix (ω_{ij}) has an inverse which we denote by (η_{ij}). Consequently,

$$(11.6) \qquad \sum_k^r \omega_{ik}\eta_{kj} = \delta_{ij}, \qquad \sum_k^r \eta_{ik}\omega_{kj} = \delta_{ij} \qquad i,j = 1,\ldots,r,$$

where δ_{ij} is the Kronecker delta. Multiply (11.5) by $\eta_{\alpha\rho}$ and $\eta_{\beta\sigma}$, sum over α and β, and use (11.6) to get

$$(11.7) \qquad [\mathcal{K}_\rho, \mathcal{K}_\sigma] = \sum_j^r c_{\rho\sigma j}\mathcal{K}_j, \qquad \rho,\sigma = 1,\ldots,r,$$

where

$$(11.8) \qquad c_{\rho\sigma j} = \sum_{\alpha,\beta}^r \left(\frac{\partial\omega_{j\beta}}{\partial t_\alpha} - \frac{\partial\omega_{j\alpha}}{\partial t_\beta}\right)\eta_{\alpha\rho}\eta_{\beta\sigma}$$

Now multiply (11.8) by $\omega_{\rho k}\omega_{\sigma l}$, sum over ρ and σ and use (11.6) to find

$$(11.9) \qquad \frac{\partial\omega_{jl}}{\partial t_k} - \frac{\partial\omega_{jk}}{\partial t_k} = \sum_{\rho,\sigma}^r c_{\rho\sigma j}\omega_{\rho k}\omega_{\sigma l}.$$

The formulas (11.7) and (11.9) are called the Maurer relations.

The $c_{\rho\sigma j}$ are independent of (x, z, p) by their definition, but apparently may depend upon t. We assert that the $c_{\rho\sigma j}$ are all constant. To see that, we note that the left hand side of (11.7) is independent of t. Differentiate with respect to t_i to get

$$\sum_j^r \left(\frac{\partial}{\partial t_i} c_{\rho\sigma j}\right)\mathcal{K}_j = 0.$$

The $\mathcal{K}_i$ are linearly independent so that

$$\frac{\partial}{\partial t_i} c_{\rho\sigma j} = 0,$$

that is the $c_{\rho\sigma j}$ are independent of t and hence are constant.

The $c_{\rho\sigma j}$ are called the **structure constants** of the group. From the definition

$$\begin{aligned} c_{\sigma\rho j} &= \sum_{\alpha,\beta}^r \left(\frac{\partial\omega_{j\beta}}{\partial t_\alpha} - \frac{\partial\omega_{j\alpha}}{\partial t_\beta}\right)\eta_{\alpha\sigma}\eta_{\beta\rho} \\ &= -\sum_{\alpha,\beta}^r \left(\frac{\partial\omega_{j\alpha}}{\partial t_\beta} - \frac{\partial\omega_{j\beta}}{\partial t_\alpha}\right)\eta_{\beta\rho}\eta_{\alpha\sigma} = -c_{\rho\sigma j} \end{aligned}$$

so the $c_{\rho\sigma j}$ are antisymmetric in the first two indices:

$$c_{\rho\sigma j} + c_{\sigma\rho j} = 0. \tag{11.10}$$

Moreover, the $c_{\rho\sigma j}$ satisfy a Jacobi type identity:

$$\sum_{\alpha}^{r} \big(c_{ik\alpha} c_{j\alpha m} + c_{kj\alpha} c_{i\alpha m} + c_{ji\alpha} c_{k\alpha m} \big) = 0 \tag{11.11}$$

The next theorem characterizes the function, $\phi = \phi(t, s)$, which describes the multiplication rule for the multiparameter group of transformations.

Theorem 11.2. *The functions describing the group operation*

$$t' = \phi(t, s)$$

is determined by the Maurer-Cartan system of total differential equations

$$\sum_{j}^{r} \omega_{ij}(t')\, dt'_j = \sum_{j}^{r} \omega_{ij}(t)\, dt_j, \quad \textit{briefly}, \quad d\omega'_i = d\omega_i \tag{11.12}$$

which satisfy the initial conditions

$$t' = s \qquad \textit{when} \quad t = 0.$$

PROOF. Let

$$P(t, s) = \left(\frac{\partial \phi_i(t, s)}{\partial t_j} \right) \qquad \text{and} \qquad Q(t, s) = \left(\frac{\partial \phi_i(t, s)}{\partial s_j} \right)$$

denote $r \times r$ matrices and consider the relations

$$\phi_i\big(\sigma, \phi(t, s)\big) = \phi_i\big(\phi(\sigma, t), s\big), \qquad i = 1, \dots, r. \tag{11.13}$$

Differentiate (11.13) successively with respect to $t_1, \dots, t_r$ to obtain the relationship

$$Q\big(\sigma, \phi(t, s)\big) P(t, s) = P\big(\phi(\sigma, t), s\big) Q(\sigma, t) \tag{11.14}$$

and then with respect to $\sigma_1, \dots, \sigma_r$ to find

$$P\big(\sigma, \phi(t, s)\big) = P\big(\phi(\sigma, t), s\big) P(\sigma, t). \tag{11.15}$$

The matrices, P and Q, are invertible. Set

$$\Omega(t,s) = P^{-1}(\sigma,t)Q(\sigma,t) \tag{11.16}$$

and in the computation below, let

$$t' = \phi(t,s) \qquad \text{and} \qquad t'' = \phi(\sigma,t).$$

Then by (11.14) and (11.15) and the definition (11.16),

$$\begin{aligned}
\Omega(t',\phi)P(t,s) &= P^{-1}(\sigma,t')\,Q(\sigma,t')\,P(t,s)\\
&= P^{-1}(\sigma,t')\,P(t'',s)\,Q(\sigma,t)\\
&= P^{-1}(\sigma,t')\,P(\sigma,t')\,P^{-1}(\sigma,t)\,Q(\sigma,t)\\
&= P^{-1}(\sigma,t)\,Q(\sigma,t)\\
&= \Omega(t,\sigma)
\end{aligned}$$

so that

$$\Omega(t',\sigma)P(t,s) = \Omega(t,\sigma). \tag{11.17}$$

Set $\sigma = 0$ and let

$$\Omega(t,0) = \big(\omega_{ij}(t)\big).$$

then (11.17) becomes

$$\sum_{j}^{r} \omega_{ij}(t')\frac{\partial\phi_i}{\partial t_k} = \omega_{ik}(t). \tag{11.18}$$

For s fixed,

$$dt'_j = \sum_{k}^{r} \frac{\partial\phi_j}{\partial t_k}\,dt_k$$

so that if we multiply (11.18) by dt_k and sum over k, we get

$$\sum_{j}^{r} \omega_j(t')\,dt'_j = \sum_{j}^{r} \omega_j(t)\,dt_j$$

which was to be proven. □

On the other hand, we can derive the associativity of the solution system, $\phi(t,s)$, from these differential equations. To see that, suppose

$$d\omega'_j = d\omega_j$$

and let

$$t'' = \phi(t', \sigma), \qquad t' = \phi(t, s).$$

Then from what we have just proven

$$d\omega_j'' = d\omega_j' = d\omega_j.$$

In particular, when $t = 0$, $t'' = \phi(s, \sigma)$. By the uniqueness of the solutions

$$t'' = \phi\bigl(t, \phi(s, \sigma)\bigr)$$

and by the definition of t''

$$t'' = \phi\bigl(\phi(t, s), \sigma\bigr)$$

which proves the associativity of the system of functions $\phi(t, s)$, which appear as solutions to the Maurer-Cartan equations.

Finally, we show that the integrability conditions for the Maurer-Cartan equations are precisely the equations (11.9).

First, rewrite the conditions $d\omega_i' = d\omega_i$ where $t' = \phi(t, s)$ as

$$\sum_j^r \omega_{ij}(t')\, dt_j' = \sum_{j,k}^r \omega_{ij}(t') \frac{\partial \phi_i}{\partial t_k}\, dt_k = \sum_k^r \omega_{ij}(t)\, dt_k$$

so that the integrability conditions is

$$\frac{\partial}{\partial t_l} \left[\sum_j^r \omega_{ij}(t') \frac{\partial \phi_j}{\partial t_k} - \omega_{ik}(t) \right] = \frac{\partial}{\partial t_k} \left[\sum_j^r \omega_{ij}(t') \frac{\partial \phi_j}{\partial t_l} - \omega_{il}(t) \right],$$

that is

$$\sum_{m,j}^r \left[\frac{\partial \omega_{ij}(t')}{\partial t_m'} \frac{\partial \phi_m}{\partial t_l} \frac{\partial \phi_j}{\partial t_k} - \frac{\partial \omega_{ij}(t')}{\partial t_m'} \frac{\partial \phi_m}{\partial t_k} \frac{\partial \phi_j}{\partial t_l} \right] = \frac{\partial \omega_{ik}}{\partial t_l} - \frac{\partial \omega_{il}}{\partial t_k}$$

by the chain rule. In the first part of the summation, sum first with respect to j and then with respect to m, and in the second, sum first with respect to m and then with respect to j. Rewriting as a single sum now yields

$$\frac{\partial \omega_{ik}}{\partial t_l} - \frac{\partial \omega_{il}}{\partial t_k} = \sum_m^r \sum_j^r \left[\frac{\partial \omega_{ij}(t')}{\partial t_m'} - \frac{\partial \omega_{im}(t')}{\partial t_j'} \right] \frac{\partial \phi_j}{\partial t_k} \frac{\partial \phi_m}{\partial t_l}. \tag{11.17}$$

Now by (11.17), the $\partial \phi_j / \partial t_k$ are the components of a matrix given by

$$P(t, s) = \left(\frac{\partial \phi_j(t, s)}{\partial t_k} \right) = \Omega^{-1}(t', 0)\omega(t, 0).$$

The matrix $\Omega^{-1}(t',0)$ is given by

$$\Omega^{-1}(t',0) = \big(\eta_{ij}(t')\big)$$

Moreover, $\Omega(t,0) = \big(\omega_{ij}(t)\big)$ so that after inserting these expressions into (11.17) and using the definition (11.8) of the structure constants, we see that (11.17) is precisely the condition (11.9).

REMARK: If the function defining the group operation satisfies

$$\phi(t,s) = \phi(s,t),$$

then the group is abelian and we can show that

$$d\omega_i(t) = \sum_j^r \omega_{ij}(t)\, dt_j$$

is a total differential. The solution to the Maurer-Cartan equations is obtained by a quadrature and one gets

$$\omega_i(t') = \omega_i(t) + \omega_i(s).$$

If we introduce the parameter

$$\tau_i = \omega_i(t),$$

then

$$\tau_i' = \tau_i + \sigma_i$$

where

$$\tau_i' = \omega_i(t'), \qquad \sigma_i = \omega_i(s),$$

which are

$$S_\tau S_\sigma = S_{\tau+\sigma}$$

In the case $r = 1$, the possibility of introducing an additive parameter follows from the associative law, but if $r \geq 2$, the commutativity condition on the group multiplication must be required in addition to associativity.

We now take up the proof of Theorem 11.1. We begin by proving that the condition

$$\mathcal{H}_k(x,z,p,t) = \sum_j^r \mathcal{K}_j(x,z,p)\omega_{jk}(t) \tag{11.19}$$

is necessary in order that the $\mathcal{H}_j$ generate a group of contact transformations.

We assume, therefore, that the family of contact transformations generated by the $\mathcal{H}_j$ forms a group and denote the function describing the group operation by ϕ so that

$$S_t S_s = S_{t'} \qquad \text{where} \quad t' = \phi(t, s). \tag{11.20}$$

Let (x^0, z^0, p^0) and s be fixed but arbitrary, and set

$$(x, z, p) = S_{t'}(x^0, z^0, p^0) = S_t S_s(x^0, z^0, p^0).$$

Then

$$\sum_{\nu}^{n} p_\nu \, dx_\nu - dz = \sum_{j}^{r} \mathcal{H}_j(x, z, p, t) \, dt'$$

and also

$$\sum_{\nu}^{n} p_\nu dx_\nu - dz = \sum_{j}^{r} \mathcal{H}_j(x, z, p, t) \, dt$$

whence, together with (11.20)

$$\sum_{l,j}^{r} \mathcal{H}_j(x, z, p, t') \frac{\partial \phi_j(t, s)}{\partial t_l} \, dt_l = \sum_{l}^{r} \mathcal{H}_l \, dt_l$$

and consequently,

$$\sum_{j}^{r} \mathcal{H}_j\big(x, z, p, \phi(t, s)\big) \frac{\partial \phi(t, s)}{\partial t_l} = \mathcal{H}_l(x, z, p, t).$$

Set $t = 0$ to find

$$\sum_{j}^{r} \mathcal{H}_j(x, z, p, s) \frac{\partial \phi(0, s)}{\partial t_l} = \mathcal{H}_l(x, z, p, 0). \tag{11.21}$$

Now let

$$\mathcal{K}_l(x, z, p) = \mathcal{H}_l(x, z, p, 0)$$

and $\big(\omega_{jk}(s)\big)$ denote the components of the matrix inverse of $\big(\partial \phi(0, s)/\partial t_l\big)$. Then (11.21) becomes with s now replaced by t

$$\mathcal{H}_k(x, z, p, t) = \sum_{j}^{r} \mathcal{K}_j(x, z, p) \omega_{jk}(t). \tag{11.22}$$

The $(\omega_{jk}(t))$ obviously has a nonzero determinant. The linear independence of the $\mathcal{K}_j$ follows immediately from that of the $\mathcal{H}_j$.

Next, let us show that the condition (11.1) is sufficient. In that case we are assuming that the canonical system generated by the $\mathcal{H}_j$ is integrable. We have seen that this implies the validity of the Maurer relations (11.9), that is the system

$$d\omega'_j = d\omega_j \tag{11.23}$$

is integrable. Let

$$t' = \phi(t, s)$$

be a solution to (11.23) satisfying

$$\phi(0, s) = s.$$

We must prove that

$$(x, z, p) = S_{t'}(x^0, z^0, p^0) \tag{11.24}$$

and

$$(x^*, z^*, p^*) = S_t S_s(x^0, z^0, p^0) \tag{11.25}$$

are equal when $t' = \phi(t, s)$. Let s be fixed and arbitrary. We consider $S_{\phi(t,s)}$ and $S_t S_s$ as functions of t. For $t = 0$,

$$(x, z, p) = (x^*, z^*, p^*) = S_s(x^0, z^0, p^0).$$

Both the (x, z, p) and (x^*, z^*, p^*) satisfy the same canonical equations, e.g. x satisfies

$$dx_\nu = \sum_j^r \frac{\partial \mathcal{H}'_j}{\partial p_\nu} dt'_j = \sum_j^r \frac{\partial \mathcal{K}_j}{\partial p_\nu}, d\omega'_j$$

$$dx^*_\nu = \sum_j^r \frac{\partial \mathcal{H}^*_j}{\partial p_\nu} dt_j = \sum_j^r \frac{\partial \mathcal{K}_j}{\partial p_\nu} d\omega_j$$

and by (11.23) these systems are the same, whence by the uniqueness, x and x^* are equal. The other cases are similar, which proves the theorem.

Theorem 11.1 is Lie's first fundamental theorem which we have proven in the setting of contact transformations. The formula (11.7) with the structure constants defined by (11.8) is called Lie's second fundamental theorem and the formulas (11.9) and (11.10) make up Lie's third fundamental theorem.

V

Selected Applications

§ 5.0 Introduction

In this chapter we take up some specific applications which illustrate the use of the theory developed in the previous chapters. In the sequel,

$$\mathrm{L} = \mathrm{L}(x, \dot{x}, t) \tag{0.1}$$

will denote the Lagrangian for a mechanical system. The coordinates describing the system are

$$x = (x_1, \dots, x_n)$$

and when $n = 1$, the subscripts will be dropped. The passage to the canonical form is effected by solving the system

$$p_j = \frac{\partial \mathrm{L}}{\partial \dot{x}_j}(x, \dot{x}, t), \qquad j = 1, \dots, n \tag{0.2}$$

for $\dot{x}$ in terms of p and defining the Hamiltonian function by

$$\mathcal{H}(x, p, t) = p \cdot \dot{x} - \mathrm{L}(x, \dot{x}, t) \tag{0.3}$$

where the $\dot{x}$ are all replaced by their expression involving (x, p, t). The resulting canonical system which describes the physical situation is

$$\begin{cases} \dot{x}_j = \dfrac{\partial \mathcal{H}}{\partial p_j}, & j = 1, \dots, n \\ \dot{p}_j = -\dfrac{\partial \mathcal{H}}{\partial x_j}, & j = 1, \dots, n. \end{cases} \tag{0.4}$$

As we noted earlier, the basic idea is to transform (0.4) into a simpler system using a contact transformation. The choice of a proper contact transformation is not always apparent. Even the concept of simplification is a subjective one. Moreover, it can occur that the work of actually carrying through the construction of a desired contact transformation may be prohibitive. Nevertheless, these examples may provide a guide in some cases to allow one to arrive at a desired simplification of (0.4).

§ 5.1 Judicious Guessing and the Undamped Harmonic Oscillator

It is often possible to arrive at a desirable contact transformation by assuming a general form for the transformation, and then choosing free parameters appropriately. We illustrate the procedure by considering the undamped harmonic oscillator. See Figure 5.1.

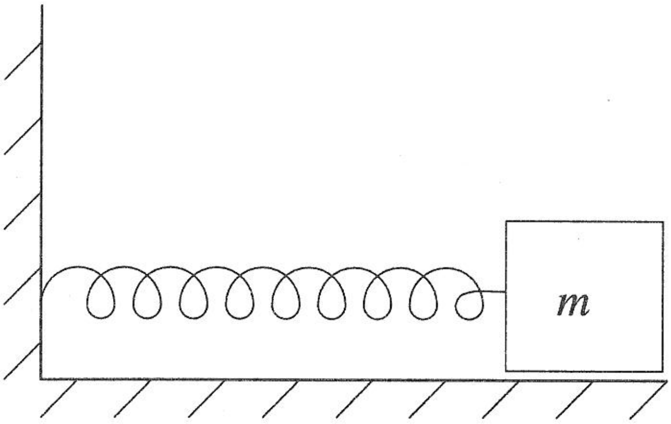

Figure 5.1

Assume that the spring is a linear, Hookian spring with spring constant k, and thst it is attached to a frictionless mass m. We also assume that the motion takes place in one dimension. Let $x = x(t)$ denote the displacement of the mass from its equilibrium position. The Lagrangian is given by

$$Ł = \frac{m\dot{x}^2}{2} - \frac{kx^2}{2},$$

and the Hamiltonian is

$$\mathcal{H}(x,p) = \frac{p^2}{2m} + \frac{kx^2}{2}, \tag{1.1}$$

where $p = m\dot{x}$.

$\mathcal{H}(x,p)$ denotes the total energy of the spring-mass system and

$$\mathcal{H}(x,p) = \text{ constant}$$

represents an ellipse in the xp-space. Let us take one of the variables to be the energy in the new coordinate system, say

$$P = \frac{p^2}{2m} + \frac{kx^2}{2}, \tag{1.2}$$

The X variable has to be brought in so that the transformation is a contact transformation, i.e. canonical. Since P represents the total energy, with a little malice of forethought, we try setting

$$\begin{cases} \dfrac{p^2}{2m} = P\cos^2(\omega X), \\ \dfrac{kx^2}{2} = P\sin^2(\omega X), \end{cases}$$

or after taking square roots

$$\begin{cases} p = (2mP)^{1/2}\cos(\omega X), \\ x = \left(\dfrac{2}{k}P\right)^{1/2}\sin(\omega X). \end{cases} \tag{1.3}$$

Obviously, (1.3) satisfies (1.2). ω is a free parameter which we will choose so that (1.3) is canonical. This means that

$$p\,dx - P\,dX = \left[\sqrt{\frac{m}{k}}\cos(\omega X)\sin(\omega X)\right]dP + \left[2\omega\sqrt{\frac{m}{k}}P\cos^2(\omega X) - P\right]dX$$

must be a perfect differential. If we equate the derivative with respect to X of the coefficient of dX, we find that $\omega = \sqrt{k/m}$. In that case the transformation (1.3) is canonical. The Hamiltonian (1.1) transforms to the simple Hamiltonian

$$\mathcal{H}(X,P) = P \tag{1.4}$$

and the corresponding canonical equations to

$$\dot{X} = 1,$$
$$\dot{P} = 0,$$

which have the simple solutions

$$X = t + \beta,$$
$$P = \alpha,$$

where α, β are constants. The solution to the original problem is

$$(1.5)\qquad \begin{cases} x = \left(\dfrac{2}{k}\alpha\right)^{1/2} \sin\left(\omega(t+\beta)\right) \\ p = (2m\alpha)^{1/2} \cos\left(\omega(t+\beta)\right) \\ \omega = \sqrt{\dfrac{k}{m}}. \end{cases}$$

From a geometrical standpoint, the solution curves (x, p, θ) with $\theta = t$, represent curves winding up an elliptical cylinder. Those curves have been transformed to straight lines embedded in a plane perpendicular to the XP-plane. See Figure 5.2.

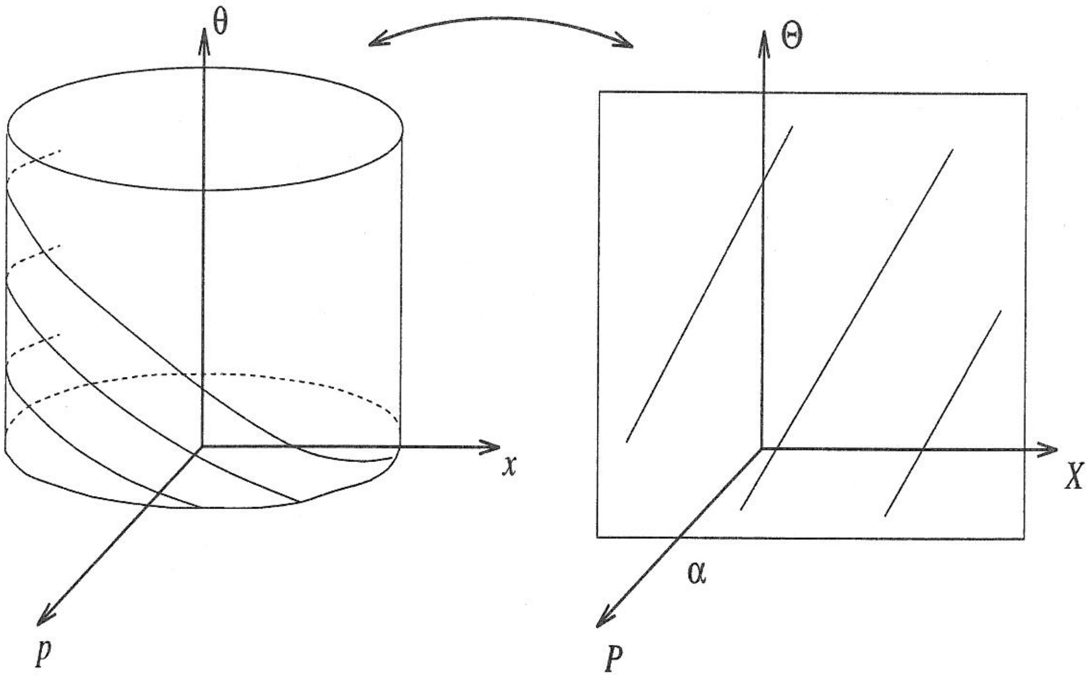

Figure 5.2

In the next section, we shall develop this transformation in a more systematic way. Rather than using the X, P variables, we shall take $P = \omega I$

and $X = (1/\omega)w$. The canonical transformation in that case is given by

$$\begin{cases} p = (2m\omega I)^{1/2}\cos w, \\ x = \left(\frac{2}{k}\omega I\right)\sin w, \end{cases} \tag{1.6}$$

with $\omega^2 = k/m$. In this case the variable I has the units of action, that is (mass)(length)^2/time and w is an angle. The variables (w, I) are therefore referred to as action-angle variables.

We could, of course, have taken

$$\mathcal{H}(x,p) = \frac{p^2}{2m} + \frac{kx^2}{2} = X$$

instead of P and would have arrived at the contact transformation

$$\begin{cases} p = (2mX)^{1/2}\cos(\omega P) \\ x = -\left(\frac{2X}{k}\right)^{1/2}\sin(\omega P) \end{cases}$$

with $\omega^2 = k/m$.

Another possibility is to manipulate the generating function of the form

$$\Omega(X,x) = \alpha x^2\cot(\beta X),$$

where α and β are parameters to be chosen. From Chapter III, section 2, we know that P and p must be chosen so that

$$P = \frac{\partial\Omega}{\partial X} = -\alpha\beta x^2\csc^2(\beta X),$$
$$p = -\frac{\partial\Omega}{\partial x} = -2\alpha x\cot(\beta X).$$

Now solve for x and p in terms of X and P to obtain

$$x = \sqrt{-\frac{P}{\alpha\beta}}\sin(\beta X),$$
$$p = -2\alpha\sqrt{-\frac{P}{\alpha\beta}}\cos(\beta X)$$

Next we choose α, β to achieve a simplification of the Hamiltonian. This is

$$\mathcal{H}(x,p) = \frac{p^2}{2m} + \frac{kx^2}{2} = \frac{4\alpha^2}{2m}\left(\frac{-P}{\alpha\beta}\right)\cos^2(\beta X) + \frac{k}{2}\left(\frac{-P}{\alpha\beta}\right)\sin^2(\beta X)$$
$$\equiv \mathcal{H}(X,P).$$

We can eliminate the X variable if we take

$$\frac{4\alpha^2}{2m(-\alpha\beta)} = 1 \qquad \text{and} \qquad \frac{k}{2(-\alpha\beta)} = 1.$$

α and β must have opposite signs and $\alpha\beta = -k/2$. Thus,

$$\alpha^2 = \frac{mk}{4}.$$

If we take $\alpha = -\sqrt{mk}/2$, then

$$\beta = -\frac{k}{2}\frac{-2}{\sqrt{mk}} = \sqrt{\frac{k}{m}} \equiv \omega.$$

The generating function is, therefore,

$$\Omega(X, x) = -\frac{\sqrt{mk}}{2}x^2 \cot(\omega X), \qquad \omega = \sqrt{\frac{k}{m}},$$

and the resulting contact transformation is (1.3).

As a final remark, we reiterate that although this approach of judicious guessing does have a certain ad hoc character, it often leads quickly to the desirable contact transformation.

§ 5.2 Action-Angle Variables

We begin this section by treating one dimensional Hamiltonians of the form

$$\mathcal{H}(x,p) = \frac{p^2}{2m} + V(x), \qquad p = m\dot{x}. \tag{2.1}$$

The potential energy function, $V(x)$, is an even function of x. $\mathcal{H}(x,p)$ represents the total energy of the system and for a given motion, it will be equal to some value E. For each permissible value of E, we suppose that

$$C_E \colon \mathcal{H}(x,p) = E \tag{2.2}$$

represents a simple closed curve in the xp phase plane so that the motion is periodic, with the period in general depending upon E. See Figure 5.3.

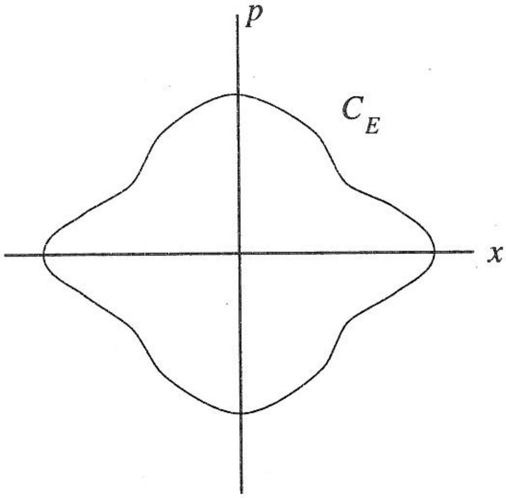

Figure 5.3

The time it will take for a particle governed by (2.2) to traverse C_E is given by

(2.3)
$$T(E) = \int_{C_E} \frac{m\,dx}{p} = \int_{C_E} \frac{m\,dx}{\pm\sqrt{2m(E-V)}} = \frac{d}{dE}\int_{C_E} \pm\sqrt{2m(E-V)}\,dx,$$

where the plus sign is to be taken when $p > 0$ and the minus sign when $p < 0$. Set

(2.4) $$I(E) = \int_{C_E} p\,dx,$$

so that (2.3) becomes

(2.5) $$T(E) = \frac{d}{dE} I(E).$$

The integral, $I(E)$, of (2.4) has the units of action. By (2.5), $I'(E) \neq 0$ so that I has an inverse

(2.6) $$I = I(E), \qquad E = E(I).$$

We propose to introduce it as a new canonical variable. It will play the role of P in the discussion below. The new canonical variables will be denoted by (w, I) which is the traditional notation. Therefore, if we can introduce new canonical variables, (w, I), by

$$(x, p) \longleftrightarrow (w, I),$$

the canonical system will transform to

$$\begin{cases} \dot{w} = E'(I), \\ \dot{I} = 0, \end{cases} \tag{2.7}$$

since $E(I) = \mathcal{H}(I)$. The units of $E'(I)$ are one divided by the time, i.e. $\dot{w}$ has units of frequency so that w is an angular measure. The variables I and w are therefore referred to as action-angle variables.

More generally, suppose $\mathcal{H} = \mathcal{H}(x,p)$ is the Hamiltonian for a mechanical system, and suppose that for energies, E, belonging to an interval, the canonical system has periodic solutions.

Let

$$C_E\colon \mathcal{H}(x,p) = E$$

be a simple closed orbit in the xp phase space and set

$$I = I(E) = \int_{C_E} p\,dx.$$

The canonical transformation

$$(x,p) \longleftrightarrow (w,I)$$

is characterized by a generating function $S = S(x,I)$. (See III.2).

Then

$$p = \frac{\partial S}{\partial x}, \qquad w = \frac{\partial S}{\partial I},$$

and the generating function, $S(s,I)$, is obtained by solving the Hamilton-Jacobi differential equation

$$\mathcal{H}\left(x, \frac{\partial S}{\partial x}\right) = E$$

for S.

Example 2.1. The Linear Harmonic Oscillator

The Hamiltonian is

$$\mathcal{H}(x,p) = \frac{p^2}{2m} + \frac{kx^2}{2}$$

and the curves $C_E\colon \mathcal{H}(x,p) = E$ are ellipses. See Figure 5.4

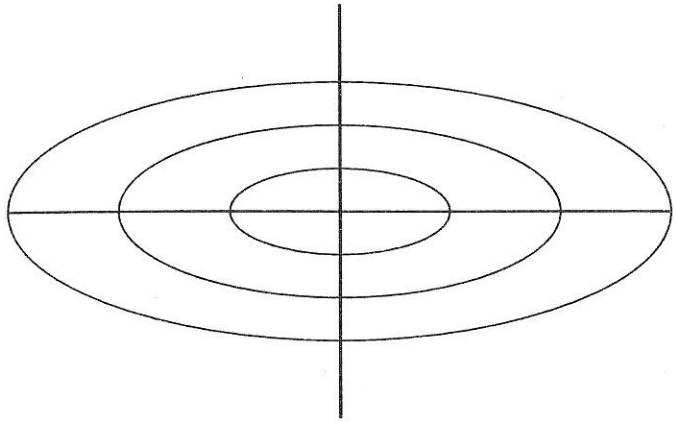

Figure 5.4

A straight forward calculation yields

$$T(E) = \frac{2\pi}{\omega},$$
$$I(E) = \frac{2\pi}{\omega}E, \qquad \text{so} \qquad E = \frac{\omega I}{2\pi},$$

where $\omega = \sqrt{k/m}$. The Hamilton-Jacobi differential equation for S_k is

$$\frac{S_k^2}{2m} + \frac{kx^2}{2} = E = \frac{\omega I}{2\pi}.$$

Thus,

$$p = S_k = \left[2m\left(\frac{\omega I}{2\pi} - \frac{kx^2}{2}\right)\right]^{1/2} = \left[\frac{m\omega I}{\pi}\left(1 - \frac{k\pi}{\omega I}x^2\right)\right]^{1/2}$$

Let $x\sqrt{k\pi/\omega I} = \sin\theta$ and integrate to obtain

$$S(x, I) = \frac{I}{2\pi}\left[\theta + 2\sin\theta\cos\theta\right],$$

$$w = \frac{\partial S}{\partial I} = \frac{1}{2\pi}\arcsin\left(xk\pi/(\omega I)\right)$$

so that

$$x = \frac{\omega I}{k\pi}\sin(2\pi w),$$
$$p = \frac{m\omega I}{\pi}\cos(2\pi w).$$

The substitution $w^* = 2\pi w$, $I^* = I/(2\pi)$ puts this into the form of (1.6) of the first section.

An easy generalization of these ideas to a system with n degrees of freedom occurs with the Hamiltonian $\mathcal{H}(x,p) = \mathcal{H}(x_1,\ldots,x_n,p_1,\ldots,p_n)$ can be written as the sum

$$\mathcal{H}(x,p) = \sum_j^n \mathcal{H}_j(x_j,p_j)$$

If $\mathcal{H}(x,p) = E$, then one can write

$$\mathcal{H}_1(x_1,p_1) = E - \sum_{j=2}^n \mathcal{H}_j(x_j,p_j)$$

and the motion of (x_1,p_1) is independent of the motions of $(x_2,p_2),\ldots,$ (x_n,p_n). Thus,

$$\mathcal{H}_1(x_1,p_1) = \alpha_1$$

a constant. The variables in this case separate and one precedes now, if possible, to introduce the action-angle coordinates

$$I_j = \int_{C_{\alpha j}} p_j\,dx_j$$

for each coordinate. Examples of this procedure are given in the next section.

§ 5.3 Separation of Variables

We shall illustrate this procedure by considering two special but typical problems where the computations can be done explicitly.

Let us treat first the case where we model a thrown ball, neglecting air resistance. See Figure 5.5.

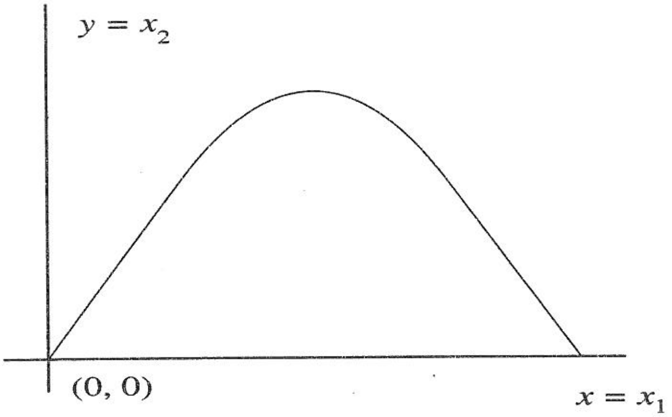

Figure 5.5

The kinetic energy is

$$T = \frac{m}{2}\left(\dot{x}^2 + \dot{y}^2\right) = \frac{m}{2}\left(\dot{x}_1^2 + \dot{x}_2^2\right),$$

and the potential energy is

$$V = mgy = mgx_2$$

The Hamiltonian is, therefore,

$$\mathcal{H}(x_1, x_2, p_1, p_2) = \frac{1}{2m}\left[\left(\frac{\partial S}{\partial x_1} + \frac{\partial S}{\partial x_2}\right)\right] + mgx_2 = E. \tag{3.1}$$

We assume that

$$S = S_1(x_1) + S_2(x_2)$$

so that (3.1) becomes

$$\frac{1}{2m}\left[\left(S_1'\right)^2(x_1) + \left(S_2'\right)^2(x_2)\right] + mgx_2 = E. \tag{3.2}$$

We put the terms depending only upon x_1 on one side of the equation and those depending upon x_2 on the other. We obtain

$$S_1'^2(x_1) = 2mE - 2m^2 gx_2 - S_2'^2(x_2). \tag{3.3}$$

The left hand side depends only upon x_1, and the right hand side on x_2. Since they vary independently of each other, they must be constant. Denote the constant by α^2. Then

$$S_1' = \alpha \qquad \text{and} \qquad 2mE - 2m^2 gx_2 - S_2'^2(x_2) = \alpha^2.$$

The function $S_1'(x_1)$ is now found to be

$$S_1(x_1) = \alpha x_1 + \beta_1.$$

Next consider

$$S_2'^2(x_2) = 2mE - \alpha^2 - 2m^2 g x_2.$$

We integrate this differential equation to find

$$S_2(x_2) = -\frac{1}{3m^2 g}\left(2mE - \alpha^2 - 2m^2 g x_2\right)^{3/2} + \beta_2.$$

The complete solution to the Hamilton Jacobi equation is given by the sum of S_1 and S_2, thus

$$S = \alpha x_1 - \frac{1}{3m^2 g}\left(2mE - \alpha^2 - 2m^2 g x_2\right)^{3/2} + \beta.$$

The generating function is therefore

$$\Omega = S - Et$$

The solution to the problem

$$\frac{\partial \Omega}{\partial x_1} = p_1, \qquad \frac{\partial \Omega}{\partial x_2} = p_2,$$

$$\frac{\partial \Omega}{\partial E} = -t - \frac{1}{mg}\left[\left(2mE - \alpha^2\right) - 2m^2 g x_2\right]^{1/2} \equiv \gamma_1$$

$$\frac{\partial \Omega}{\partial \alpha} = x_1 + \frac{\alpha}{m^2 g}\left[\left(2m^2 E - \alpha^2\right) - 2m^2 g x_2\right]^{1/2} \equiv \gamma_2.$$

Square the equation for $\partial\Omega/\partial E$ and rearrange the terms to find

$$x_2 = -\frac{gt^2}{2} + c_1 t + c_2, \qquad c_1, c_2 \quad \text{constant}$$

and by eliminating the radical from the $\partial\Omega/\partial E$ and $\partial\Omega/\partial\alpha$ terms, we find for x_1

$$x_1 = \frac{\alpha}{m} t + k, \qquad k \quad \text{a constant.}$$

Consider next the motion of a planet m which moves about a fixed sum of mass, M. The Hamiltionian in spherical coordinates (r, θ, ϕ)

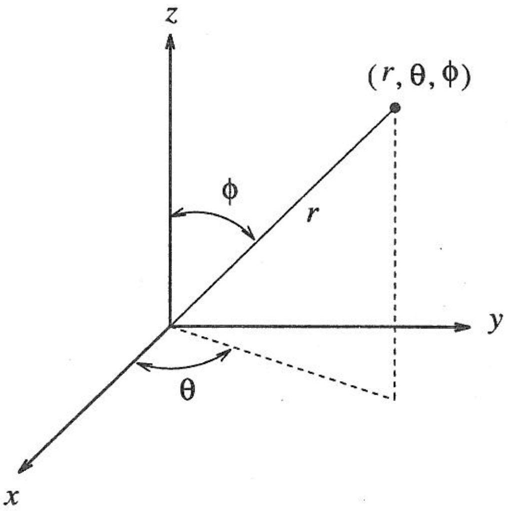

Figure 5.6

is given by
(3.4)

$$\mathcal{H} = \mathcal{H}(p_r, p_\theta, p_\phi, r, \theta) = \frac{1}{2m}\left(p_r^2 + \frac{1}{r^2}p_\theta^2 + \frac{p_\phi^2}{r^2\sin^2\theta}\right) - \frac{GmM}{r} = E.$$

According to Jacobi's method, we seek a general solution to (3.4). This is done by assuming a solution in the form

$$S = R(r) + \Theta(\theta) + \Phi(\phi), \tag{3.5}$$

and determining the unknown functions R, Θ, Φ. We find

$$\frac{1}{2m}\left[R'^2(r) + \frac{1}{r^2}\Theta'^2(\theta) + \frac{1}{r^2\sin^2\theta}\Phi'^2(\phi)\right] - \frac{GmM}{r} = E.$$

We first separate out the $\Phi(\phi)$ variable

$$\frac{1}{2m}\left[R'^2(r) + \frac{1}{r^2}\Theta'^2(\theta)\right] - \frac{GmM}{r} - E = -\frac{1}{2mr^2\sin^2\theta}\Phi'^2(\phi),$$

whence

$$\left\{\frac{1}{2m}\left[R'^2(r) + \frac{1}{r^2}\Theta'^2(\theta)\right] - \frac{GmM}{r} - E\right\}2mr^2\sin^2\theta = -\Phi'^2(\phi). \tag{3.6}$$

The left hand side depends only on (r, θ), while the right hand side only on ϕ. Thus, they are constant. Set

$$\Phi'(\phi) = a_3$$

so that

$$\Phi(\phi) = a_3\phi + b_3,$$

where a_3, b_3 are constants. Next, return to (3.6) and separate the terms depending on r from those depending on θ. We find first

$$\frac{1}{2m}R'^2(r) - \frac{GmM}{r} - E = -\frac{1}{2mr^2}\Theta'^2(\theta) - \frac{a_3^2}{2mr^2\sin^2\theta}$$

so that

$$2mr^2\left[\frac{1}{2m}R'^2(r) - \frac{GmM}{r} - E\right] = -\left[\Theta'^2 + \frac{a_3^2}{\sin^2\theta}\right].$$

Once again the variables have been successfully separated, each side therefore is constant, denote it by $-a_2$. The equation for $\Theta(\theta)$ is

$$\Theta'^2 + \frac{a_3^2}{\sin^2\theta} = a_2.$$

Integrate this equation to find

$$\Theta(\theta) = \int\left[a_2 - \frac{a_3^2}{\sin^2\theta}\right]^{1/2} d\theta + b_2.$$

Finally we obtain for $R(r)$,

$$R(r) = \int\left[2m\left(E + \frac{GmM}{r}\right) - \frac{a_2}{r^2}\right]^{1/2} dr + b_1.$$

The sum of $R(r)$, $\Theta(\theta)$, $\Phi(\phi)$ gives the complete integral of the Hamilton Jacobi differential equation. The (E, a_2, a_3) correspond to the (P_1, P_2, P_3). We find for S the expression

$$S = -Et + a_3\phi + \int\left[a_2 - \frac{a_3^2}{\sin^2\theta}\right]^{1/2} d\theta$$
$$+ \int\left[2m\left(E + \frac{GmM}{r}\right) - \frac{a_2}{r^2}\right]^{1/2} dr + b,$$

where $b = b_1 + b_2 + b_3$. Differentiation with respect to (E, a_2, a_3) yields the integrated form of the equations of motion

$$Q_1 = -\frac{\partial S}{\partial E} = t - \int \left[2m\left(E + \frac{GmM}{r}\right) - \frac{a_2}{r^2}\right]^{-1/2} m\,dr,$$

$$\begin{aligned} Q_2 &= \frac{\partial S}{\partial a_2} \\ &= \frac{1}{2}\int \left[a_2 - \frac{a_3^2}{\sin^2\theta}\right]^{-1/2} d\theta - \int \left[2m\left(E + \frac{GmM}{r}\right) - \frac{a_2}{r^2}\right]^{-1/2} \frac{dr}{r^2}, \end{aligned}$$

$$Q_3 = \frac{\partial S}{\partial a_3} = \phi + \int \left[a_2 - \frac{a_3^2}{\sin^2\theta}\right]^{-1/2} \frac{a_3^2}{\sin^2\theta}\,d\theta.$$

§ 5.4 Perturbation Theory

In perturbation theory, the Hamiltonian function, $\mathcal{H}$, is thought of as being written as the sum of two terms, a principal part which makes the major contribution to the solution plus a small remainder term, the perturbation:

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2, \tag{4.1}$$

where $\mathcal{H}_1$ is the principal term and is itself the characteristic function for a canonical system. The canonical equations arising from $\mathcal{H}_1$ are supposed to be simple in the sense that they can be solved. The idea is to use the solutions arising from $\mathcal{H}_1$ to get some information about the behavior of the solutions to the system generated by $\mathcal{H}$.

We illustrate the procedure for a spring-mass system where the restoring force of the spring is given by

$$f(x) = -(kx + lx^3) \tag{4.2}$$

This nonlinear spring is referred to as a stiff spring. The parameters k and l are assumed to be constant. $x = x(t)$ is the position of the mass, m, measured from the equilibrium position. We assume that there are no external forces acting on m, and that friction is absent as well. See Figure 5.7

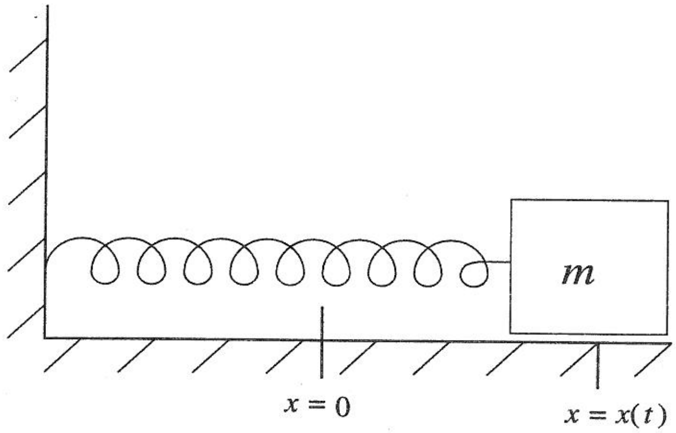

Figure 5.7

The kinetic energy for the system is

$$T = \frac{m\dot{x}^2}{2},$$

and the potential energy is

$$V = \frac{kx^2}{2} + \frac{lx^4}{4}$$

so that the Hamiltonian function is

$$\mathcal{H}(x,p) = \frac{1}{2m}p^2 + \frac{k}{2}x^2 + \frac{l}{4}x^4. \tag{4.3}$$

We can write (4.3) in the form of (4.1) by setting

$$\mathcal{H}_1(x,p) = \frac{1}{2m}p^2 + \frac{k}{2}x^2, \qquad \text{and} \qquad \mathcal{H}_2(x,p) = \frac{l}{4}x^4. \tag{4.4}$$

$\mathcal{H}_1$ is the characteristic function for the linear harmonic operator.

We have solved the canonical system

$$\begin{cases} \dot{x} = \dfrac{p}{m} \\ \dot{p} = -kx \end{cases} \tag{4.5}$$

by reducing it to a simpler system using the contact transformation

$$\begin{cases} p = (2mP)^{1/2}\cos(\omega X) \\ x = \left(\dfrac{2}{k}P\right)\sin(\omega X) \end{cases} \tag{4.6}$$

with $\omega^2 = k/m$. Under (4.6) the function $\mathcal{H}_1(x,p)$ transforms to P. The total Hamiltonian, (4.3) transforms to

$$H(X,P) = P + \frac{\epsilon}{2}P^2 \sin^2(\omega X) \tag{4.7}$$

where we have taken

$$\epsilon = \frac{2l}{k^2}, \tag{4.8}$$

and we are thinking of ϵ as a small parameter. The resulting canonical system generated by (4.7) is

$$\begin{aligned} \dot{X} &= 1 + \epsilon P \sin^2(\omega X) \\ \dot{P} &= \omega\epsilon P^2 \sin(\omega X)\cos(\omega X) \end{aligned} \tag{4.9}$$

The initial values

$$x(0) = x^0, \qquad p(0) = p^0, \tag{4.10}$$

transform to

$$X(0) = X^0, \qquad P(0) = P^0, \tag{4.11}$$

and the system (4.9) should be solved subject to (4.11). Observe that when $\epsilon = 0$, (4.9) reduces to the problem for the harmonic oscillator. There are several ways we could procede to get some information about solutions to (4.9), (4.11) and how they compare, at least for small time, to solutions to the desired problem. We shall, however, only indicate one possibility. Observe that both X and P depend upon ϵ so that in fact

$$X = X(t;\epsilon), \qquad P = P(t;\epsilon).$$

Expand both X and P in power series in ϵ to find

$$\begin{cases} X(t;\epsilon) = X_0(t) + X_1(t)\epsilon + \dots \\ P(t;\epsilon) = P_0(t) + P_1(t)\epsilon + \dots \end{cases} \tag{4.12}$$

where

$$X_k(t) = \frac{1}{k!}\left.\frac{\partial^k X(t;\epsilon)}{\partial\epsilon^k}\right|_{\epsilon=0}, \qquad P_k(t) = \frac{1}{k!}\left.\frac{\partial^k P(t;\epsilon)}{\partial\epsilon^k}\right|_{\epsilon=0}.$$

Now at $t = 0$, $X_0(0) = X^0$, $X_1(0) = 0, \ldots$, and $P_0(0) = P^0$, $P_1(0) = 0$, $\ldots$. X_0 and P_0 satisfy the system

$$\dot{X}_0 = 1$$
$$\dot{P}_0 = 0$$

so that

$$\begin{aligned} X_0(t) &= t + X^0 \\ P_0(t) &= P^0. \end{aligned} \tag{4.13}$$

Next, differentiate (4.9) with respect to ϵ and set $\epsilon = 0$ to conclude that X_1 and P_1 satisfy

$$\begin{aligned} \dot{X}_1 &= P_0(t)\sin^2\big(\omega X_0(t)\big) = P^0\sin^2\big(\omega(t+X^0)\big) \\ \dot{P}_1 &= \omega P_0^2(t)\sin\big(\omega X_0(t)\big)\cos\big(\omega X_0(t)\big) \\ &= \omega {P^0}^2\sin\big(\omega(t+X^0)\big)\cos\big(\omega(t+X^0)\big), \end{aligned}$$

so that

$$\begin{cases} X_1(t) = \dfrac{p^0 t}{2} - \dfrac{\sin\big(2\omega(t+X^0)\big)}{4\omega} + \dfrac{\sin(2\omega X^0)}{4\omega} \\ P_1(t) = -\dfrac{{P^0}^2}{4}\left[\cos\big(2\omega(t+X^0)\big) - \cos(2\omega X^0)\right]. \end{cases} \tag{4.14}$$

One proceeds in this way to generate as many terms in (4.12) as desired, although the calculations become increasingly laborious. The solution is given by

$$\begin{cases} X(t;\epsilon) = (t+X^0) + \left(\dfrac{P^0 t}{2} - \dfrac{\sin\big(2\omega(t+X^0)\big)}{4\omega} + \dfrac{\sin(2\omega X^0)}{4\omega}\right)\epsilon + \ldots \\ P(t;\epsilon) = P^0 + \dfrac{{P^0}^2}{4}\left(\cos(2\omega X^0) - \cos\big(2\omega(t+X^0)\big)\right)\epsilon + \ldots \end{cases} \tag{4.15}$$

and now the solution itself to the original canonical system

$$\begin{cases} \dot{x} = \dfrac{p}{m} \\ \dot{p} = -(kx + lx^3), \end{cases} \tag{4.16}$$

is given by inserting the expressions (4.15), into (4.16). An approximation is obtained by truncating the series in (4.15), say neglecting all terms containing powers of ϵ greater than or equal to two. Thereby achieving a suitable approximation.

Bibliography

1. E. Kamke, *Differentialgleichungen, Lösungsmethoden und Lösungen*; I, *Gewöhnliche Differentialgleichungen*, Akademische Verlagsgesellschaft, Leipzig, 1951; II, *Partielle Differentialgleichungen erster Ordnung für eine gesuchte Funktion*, Akademische Verlagsgesellschaft, Leipzig, 1948.

2. E.A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, Robert E. Krieger, Malabar, Florida, 1984.

3. S. Lie and G. Scheffers, *Geometrie der Berührungstransformationen*, Teubner, Leipzig, 1896.

4. S. Lie and F. Engel, *Theorie der Transformationsgruppe*; 3 Bände, Teubner, Leipzig, 1888; insbes Band II, *Berührungstransformationen*, Ein Nachdruck ist 1930 erschienen.

5. S. Lie, *Gesammelte Abhandlungen*, Bd. III Differentialgleichungen, vol. 1, Teubner and Kristiania, Leipzig and Aschehoug, 1922; Bd. IV, *Differentialgliechungen*, vol. 2, Teubner and Oslo, Leipzig and Aschehoug, 1929.

6. C. Carathéodory, *Variationsrechnung und Partielle Differentialgleichungen Erster Ordnung*, B.G. Teubner Verlagsgesellschaft, Leipzig.

7. Felix Klein, *Vorlesungen uber Höhere Geometrie*, J. Springer, Berlin, 1926, Die Grundlehren der Mathematischen Wissenschaft Bd. XXII.

8. Gustav Herglotz, *Vorlesungen über die Mechanik der Kontinua*, BSB B.G. Teubner Verlagsgesellschaft, Leipzig, 1985.

9. ______, *Gesammelte Schriften*, Vandenhoeck & Ruprecht, Göttingen.