5 APPLICATIONS IN PHYSICS

A Thermodynamics

5.1 Simple systems

We confine our attention at first to a one-component fluid, for which the equation of conservation of energy is

$$\delta Q = P\mathrm{d}V + \mathrm{d}U, \tag{5.1}$$

where U is the internal energy of the fluid and δQ is the heat absorbed as the fluid does work $P\mathrm{d}V$ and changes its energy. We shall interpret this equation as a relation among various one-forms in the two-dimensional manifold whose coordinates are (V, U), on which the function $P(V, U)$ is defined (called the equation of state). Then since $\tilde{\mathrm{d}}V$ and $\tilde{\mathrm{d}}U$ are one-forms, so is $\tilde{\delta}Q$. But is $\tilde{\delta}Q$ an *exact* one-form? That is, can one find a function $Q(V, U)$ such that $\tilde{\delta}Q = \tilde{\mathrm{d}}Q$? If this were true, then one would have $\widetilde{\mathrm{d}\mathrm{d}}Q = 0$, which would mean

$$0 = \tilde{\mathrm{d}}P\wedge\tilde{\mathrm{d}}V = \left[\left(\frac{\partial P}{\partial V}\right)_U \tilde{\mathrm{d}}V + \left(\frac{\partial P}{\partial U}\right)_V \tilde{\mathrm{d}}U\right]\wedge\tilde{\mathrm{d}}V$$
$$= \left(\frac{\partial P}{\partial U}\right)_V \tilde{\mathrm{d}}U\wedge\tilde{\mathrm{d}}V.$$

(Subscripts on derivatives indicate which variable is fixed during differentiation.) Thus, a function Q can exist only if $(\partial P/\partial U)_V$ vanishes everywhere: this would be a strange fluid indeed!

Since $\tilde{\delta}Q$ is a one-form in a two-space, its ideal is automatically *closed*, so by Frobenius' theorem (§4.26) there must exist functions $T(U, V)$ and $S(U, V)$ such that $\tilde{\delta}Q = T\tilde{\mathrm{d}}S$. Thus, we define the temperature and entropy functions for the single-component gas in thermodynamic equilibrium simply as a representation of the one-form in equation (5.1):

♦ $$T\tilde{\mathrm{d}}S = P\tilde{\mathrm{d}}V + \tilde{\mathrm{d}}U. \tag{5.2}$$

It is important to understand that this is a purely mathematical definition of T and S, and it has no relation to the second law of thermodynamics, which we will consider in a moment. No mathematical *identity* of this sort would hold for a multi-component fluid. (We shall see that the second law of thermodynamics is equivalent to requiring $\tilde{\delta}Q = T\tilde{\mathrm{d}}S$ for composite systems. Because this is not an automatic identity, the second law *is* a physical law: it restricts the possible mathematical nature of physical systems.)

5.2 Maxwell and other mathematical identities

Taking the exterior derivative of (5.2) gives

$$\tilde{\mathrm{d}}T \wedge \tilde{\mathrm{d}}S = \tilde{\mathrm{d}}P \wedge \tilde{\mathrm{d}}V. \tag{5.3}$$

Suppose we write $T = T(S, V), P = P(S, V)$. Then (5.3) gives (since $\tilde{\mathrm{d}}S \wedge \tilde{\mathrm{d}}S \equiv 0$, $\tilde{\mathrm{d}}V \wedge \tilde{\mathrm{d}}V \equiv 0$):

$$\left(\frac{\partial T}{\partial V}\right)_S \tilde{\mathrm{d}}V \wedge \tilde{\mathrm{d}}S = \left(\frac{\partial P}{\partial S}\right)_V \tilde{\mathrm{d}}S \wedge \tilde{\mathrm{d}}V = -\left(\frac{\partial P}{\partial S}\right)_V \tilde{\mathrm{d}}V \wedge \tilde{\mathrm{d}}S.$$

From this we conclude

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V, \tag{5.4}$$

which is known as one of the Maxwell identities. Similarly, by writing $S = S(T, V), P = P(T, V)$, we can deduce

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V, \tag{5.5}$$

another Maxwell identity. By dividing (5.2) by T and then taking the exterior derivative we get

$$\frac{1}{T}\,\tilde{\mathrm{d}}P \wedge \tilde{\mathrm{d}}V - \frac{P}{T^2}\,\tilde{\mathrm{d}}T \wedge \tilde{\mathrm{d}}V - \frac{1}{T^2}\,\tilde{\mathrm{d}}T \wedge \tilde{\mathrm{d}}U = 0.$$

By writing $U = U(T, V), P = P(T, V)$, we get

$$\frac{1}{T}\left(\frac{\partial P}{\partial T}\right)_V \tilde{\mathrm{d}}T \wedge \tilde{\mathrm{d}}V - \frac{P}{T^2}\,\tilde{\mathrm{d}}T \wedge \tilde{\mathrm{d}}V - \frac{1}{T^2}\left(\frac{\partial U}{\partial V}\right)_T \tilde{\mathrm{d}}T \wedge \tilde{\mathrm{d}}V = 0,$$

or

$$T\left(\frac{\partial P}{\partial T}\right)_V - P = \left(\frac{\partial U}{\partial V}\right)_T. \tag{5.6}$$

Exercise 5.1

Derive the identity

$$T\left(\frac{\partial P}{\partial T}\right)_S - P = \left(\frac{\partial P}{\partial T}\right)_S \left(\frac{\partial U}{\partial S}\right)_T - \left(\frac{\partial P}{\partial S}\right)_T \left(\frac{\partial U}{\partial T}\right)_S \tag{5.7}$$

by multiplying (5.2) by $1/P$ and differentiating.

Another important relation which follows easily from the use of forms is

$$\left(\frac{\partial T}{\partial P}\right)_S \left(\frac{\partial S}{\partial T}\right)_P \left(\frac{\partial P}{\partial S}\right)_T = -1, \tag{5.8}$$

which is equally true of any set of three of (P, V, U, T, S). We prove this by writing

$$T = T(P,S), S = S(T,P), P = P(T,S), \tag{5.9}$$

which is possible since the manifold is two-dimensional. Then we have the successive identities:

$$\begin{aligned}\tilde{d}T \wedge \tilde{d}S &= \left(\frac{\partial T}{\partial P}\right)_S \tilde{d}P \wedge \tilde{d}S \\ &= \left(\frac{\partial T}{\partial P}\right)_S \left(\frac{\partial S}{\partial T}\right)_P \tilde{d}P \wedge \tilde{d}T \\ &= \left(\frac{\partial T}{\partial P}\right)_S \left(\frac{\partial S}{\partial T}\right)_P \left(\frac{\partial P}{\partial S}\right)_T \tilde{d}S \wedge \tilde{d}T,\end{aligned}$$

from which follows (5.8). Notice that the derivation here relies only on the ability to write (5.9), so that it is really an identity among any three functions on a two-dimensional manifold.

The ease with which the Maxwell identities and (5.8) can be derived using forms is an illustration of the natural way in which they fit into thermodynamics: the one-forms $\tilde{d}P$, $\tilde{d}S$, etc. are the mathematically precise substitutes for the physicists' rather fuzzier concept of the infinitesimals dP, dS, etc.

5.3 Composite thermodynamic systems: Caratheodory's theorem

We now consider composite thermodynamic systems, the parts of which may exchange energy with each other and with the outside world. In this case the law of conservation of energy is (for a system with N parts)

$$\begin{aligned}\tilde{\delta}Q &= P_1\tilde{d}V_1 + \tilde{d}U_1 + P_2\tilde{d}V_2 + \tilde{d}U_2 + \ldots \\ &= \sum_{i=1}^{N} (P_i\tilde{d}V_i + \tilde{d}U_i).\end{aligned} \tag{5.10}$$

We regard this as a relation among one-forms on a $2N$-dimensional manifold whose coordinates are $(V_i, U_i; i = 1, \ldots, N)$, and we assume that each P_i can be expressed as a function of these coordinates. The question arises of whether one can define an entropy and temperature for the system as a whole, i.e. whether T and S exist such that

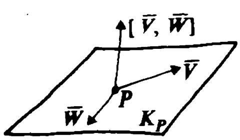
♦ $$\tilde{\delta}Q = T\tilde{d}S. \tag{5.11}$$

This equation is just the statement that $\tilde{\delta}Q$ is *integrable* (in the sense of the Frobenius theorem). Now the Frobenius theorem tells us that the necessary and sufficient condition for this to be true is $\tilde{d}\tilde{\delta Q} \wedge \tilde{\delta Q} = 0$. It is easy to see from (5.6) that this will not generally be true, so we can conclude that for a *general* interacting system there is no global temperature or entropy function. But the situation can be different for an *equilibrium* system, because the conditions for mechanical and thermodynamic equilibrium among the constituent parts restrict the problem (we assume) to a submanifold of the $2N$-dimensional one. We shall from now on let the world 'manifold' refer to this *equilibrium submanifold*, and examine the possibility that $\tilde{\delta Q}$ is integrable in it from the point of view of *Caratheodory*.

If $\tilde{\delta Q}$ is integrable, then every point of the manifold is on one and only one integral submanifold; these submanifolds are defined by $S = \text{const}$. None of these surfaces intersect. Therefore, starting at one point, it is *not* possible to reach an arbitrary point of the manifold along a curve on which δQ is everywhere zero. In other words, if an entropy function exists it is not possible to reach every equilibrium state of the system along an adiabatic path of equilibria. The physically interesting question is whether the *converse* is true: if we know that not every state is reachable along a path for which $\delta Q = 0$, can we say that $\tilde{\delta Q}$ is integrable? This is interesting because one version of the second law of thermodynamics asserts that it is impossible in a closed system to transfer heat from a colder to a hotter body without making other changes as well. By a closed system we mean one for which $\delta Q = 0$, so that the second law tells us that not every state can be achieved with $\delta Q = 0$. So does the second law imply the existence of an entropy function? Caratheodory's theorem says it does.

What we shall prove is that if $\tilde{\delta Q}$ is not integrable then all points in the neighborhood of some initial point P are reachable from P on a curve which annuls $\tilde{\delta Q}$. Since $\tilde{\delta Q}$ is not integrable, the version of Frobenius' theorem given in §4.26 shows us that there are at least two vector fields $\bar{V}$ and $\bar{W}$ for which $\tilde{\delta Q}(\bar{V}) = \tilde{\delta Q}(\bar{W}) = 0$ in a neighborhood of any point P, but $\tilde{\delta Q}([\bar{V}, \bar{W}]) \neq 0$ at P. That is, the one-form $\tilde{\delta Q}$ defines at each point P a subspace K_P of T_P, the vectors of which annul $\tilde{\delta Q}$; the nonintegrability of $\tilde{\delta Q}$ means that vector fields everywhere in K_P do not form a hypersurface: at least one of their Lie brackets does not lie in K_P (see figure 5.1). Because annulling $\tilde{\delta Q}$ is only one equation, K_P has

Fig. 5.1. The tangent hyperplane K_P contains the vectors annulling $\tilde{\delta Q}$ but not all of their Lie brackets at P.



dimension $n - 1$, where n is the dimension of the equilibrium manifold. Now, recall the exponentiation notation for the Taylor series introduced in §2.13. If we take any vector field U which is in K_P at all points P, and we move along it a parameter distance ϵ from P, we reach the point whose coordinates are $x^i = \exp(\epsilon\bar{U})x^i|_P$, where we use $\bar{U}$ as a derivative operator on the function x^i along the curve. The set of all points in a small neighborhood of P reachable in this way may be called $\exp(\epsilon K_P)$: it is the representation in the manifold of the vector space K_P. This set of points is locally like a piece of an $(n-1)$-dimensional hypersurface. We shall show that, by following the curves of $\bar{V}$ and $\bar{W}$ defined above, we can reach points 'above' or 'below' this 'hypersurface' – i.e. that we can reach *all* points near P. The trip we make is the following: we move first a distance ϵ along $\bar{V}$, then ϵ along $\bar{W}$, then $-\epsilon$ along $\bar{V}$, and finally $-\epsilon$ along $\bar{W}$. This takes us to (cf. equation (2.6))

$$\begin{aligned} x^i &= e^{-\epsilon\bar{W}}e^{-\epsilon\bar{V}}e^{\epsilon\bar{W}}e^{\epsilon\bar{V}}x^i|_P \\ &= (1 + \epsilon^2[\bar{W}, \bar{V}] + O(\epsilon^3))x^i|_P. \end{aligned} \tag{5.12}$$

This means that we wind up almost back at P, but a parameter distance ϵ^2 away from it along $[\bar{V}, \bar{W}]$. This point is not in $\exp(\epsilon K_P)$, since $[\bar{V}, \bar{W}]$ is not in K_P. It is on one side of $\exp(\epsilon K_P)$; to finish on the other side we would have travelled first on $\bar{W}$, then on $\bar{V}$. Now, our path was along $\bar{V}$ or $\bar{W}$ everywhere, so it was adiabatic: $\delta Q = 0$ everywhere. It is clear, therefore, that if $\tilde{\delta Q}$ is not integrable, all states of the system will be reachable along adiabatic paths. This proves that the second law requires integrability of $\tilde{\delta Q}$ in the equilibrium manifold and the existence of an entropy function for composite systems in equilibrium.

B Hamiltonian mechanics

5.4 Hamiltonian vector fields

The Hamiltonian version of a dynamical system of equations begins with the Lagrangian $\mathscr{L}(q, q_{,t})$ for some dynamical variable $q(t)$. The momentum p is defined as

$$p = \partial\mathscr{L}/\partial(q_{,t}), \tag{5.13}$$

and the Hamiltonian H as

$$H = pq_{,t} - \mathscr{L} = H(p, q). \tag{5.14}$$

The dynamical equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathscr{L}}{\partial q_{,t}} - \frac{\partial\mathscr{L}}{\partial q} = 0, \tag{5.15}$$

and the definition of p can be written, respectively, as

$$\frac{\partial H}{\partial q} = -\frac{\mathrm{d}p}{\mathrm{d}t}, \quad \text{and} \quad \frac{\partial H}{\partial p} = \frac{\mathrm{d}q}{\mathrm{d}t}. \tag{5.16}$$

We now make a geometric picture of Hamiltonian dynamics by defining a manifold M called 'phase space', whose coordinates are p and q. On M we define the two-form

♦ $$\tilde{\omega} \equiv \tilde{\mathrm{d}}q \wedge \tilde{\mathrm{d}}p. \tag{5.17}$$

Consider a curve $\{q = f(t), p = g(t)\}$ on M which is a solution of (5.16). Its tangent vector, $\bar{U} = \mathrm{d}/\mathrm{d}t = f_{,t}\, \partial/\partial q + g_{,t}\, \partial/\partial p$, has the property

♦ $$£_{\bar{U}}\tilde{\omega} = 0, \tag{5.18}$$

as we shall now prove. Since $\tilde{\mathrm{d}}\tilde{\omega} = 0$, we have from (4.67)

$$£_{\bar{U}}\tilde{\omega} = \tilde{\mathrm{d}}[\tilde{\omega}(\bar{U})]. \tag{5.19}$$

But since $\tilde{\omega} = \tilde{\mathrm{d}}q \otimes \tilde{\mathrm{d}}p - \tilde{\mathrm{d}}p \otimes \tilde{\mathrm{d}}q$, we have

$$\begin{aligned} \tilde{\omega}(\bar{U}) &= \langle \tilde{\mathrm{d}}q, \bar{U}\rangle\, \tilde{\mathrm{d}}p - \langle \tilde{\mathrm{d}}p, \bar{U}\rangle\, \tilde{\mathrm{d}}q \\ &= \frac{\mathrm{d}f}{\mathrm{d}t}\tilde{\mathrm{d}}p - \frac{\mathrm{d}g}{\mathrm{d}t}\tilde{\mathrm{d}}q. \end{aligned} \tag{5.20}$$

On the other hand, since f and g satisfy (5.16), we have

$$\tilde{\omega}(\bar{U}) = \frac{\partial H}{\partial p}\tilde{\mathrm{d}}p + \frac{\partial H}{\partial q}\tilde{\mathrm{d}}q = \tilde{\mathrm{d}}H. \tag{5.21}$$

Therefore $\tilde{\mathrm{d}}[\tilde{\omega}(\bar{U})]$ vanishes, establishing (5.18). A vector field $\bar{U}$ that satisfies (5.18) is called a *Hamiltonian vector field*.

Exercise 5.2

(a) Prove that if $\bar{U}$ is a Hamiltonian vector field, there exists some $H(p, q)$ such that equations (5.16) are satisfied along the integral curves of $\bar{U}$.

(b) Prove that Hamiltonian vector fields form a Lie algebra.

By exercise 5.2(a), we interpret $\bar{U}$ as a tangent to the solution curves in phase space if $\bar{U}$ is Hamiltonian. Notice that the system is *conservative*, since (5.16) implies

$$£_{\bar{U}}H \equiv \frac{\mathrm{d}H}{\mathrm{d}t} = 0. \tag{5.22}$$

5.5 Canonical transformation

Now the coordinates p and q are not unique. We define a *canonical transformation* as one which leaves $\tilde{\omega}$ in the same form. That is, new coordinates $P = P(q, p)$ and $Q = Q(q, p)$ are called canonical if

$$\tilde{d}q \wedge \tilde{d}p = \tilde{d}Q \wedge \tilde{d}P. \tag{5.23}$$

The necessary and sufficient condition for this is

$$\left(\frac{\partial Q}{\partial q}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q}\right) = 1. \tag{5.24}$$

One such transformation is $Q = p, P = -q$. A less trivial one is found if we follow a procedure similar to the one we used to deduce the Maxwell identities in thermodynamics: we write $p = p(q, Q), P = P(q, Q)$ and find from (5.23) that

$$\partial p/\partial Q = -\partial P/\partial q. \tag{5.25}$$

So if we take an arbitrary function $F(q, Q)$ and define

$$p = \partial F/\partial q, \qquad P = -\partial F/\partial Q,$$

then (5.25) is satisfied identically. Thus, $F(q, Q)$ is said to generate a canonical transformation. Since we could have chosen, instead of (q, Q), the pairs (q, P), (p, Q), or (p, P) to be independent in (5.23), there are clearly four types of such generating functions for canonical transformations. They are explored more fully in Goldstein (1950) (see bibliography).

5.6 Map between vectors and one-forms provided by $\tilde{\omega}$

One of the most important features of this geometrical point of view on Hamiltonian dynamics is that $\tilde{\omega}$ can be cast in a role similar to that which a metric plays on Riemannian manifolds: it provides an invertible 1–1 mapping between vectors and one-forms. If $\bar{V}$ is a vector field on M, we define a one-form field

$$\tilde{V} \equiv \tilde{\omega}(\bar{V}), \tag{5.26}$$

with components

$$(\tilde{V})_i = \omega_{ij}V^j. \tag{5.27}$$

Similarly, given a one-form field $\tilde{\alpha}$ we define a vector field $\bar{\alpha}$ as the (unique) vector such that

$$\tilde{\alpha} = \tilde{\omega}(\bar{\alpha}). \tag{5.28}$$

Exercise 5.3

Prove that $\langle\tilde{V}, \bar{V}\rangle = 0$, so that $\tilde{\omega}$ is not suitable as a metric.

Exercise 5.4

Prove that if $\tilde{\alpha} = f\,\tilde{d}q + g\,\tilde{d}p$, then

$$\bar{\alpha} = g\frac{\partial}{\partial q} - f\frac{\partial}{\partial p}. \tag{5.29}$$

Exercise 5.5

Prove that $\bar{X}$ is a Hamiltonian vector field on M if and only if $\tilde{X}$ is an exact one-form, i.e. if and only if there exists some function H such that $\tilde{X} = \tilde{d}H$, or $\bar{X} = \overline{dH}$.

5.7 Poisson bracket

Suppose there are two functions f and g on the manifold, and we define the vector fields $\bar{X}_f \equiv \overline{df}$ and $\bar{X}_g \equiv \overline{dg}$. Then consider the scalar

$$\{f,g\} \equiv \tilde{\omega}(\bar{X}_f, \bar{X}_g) = \langle \tilde{d}f, \bar{X}_g \rangle. \tag{5.30}$$

Since $\tilde{\omega} = \tilde{d}q \otimes \tilde{d}p - \tilde{d}p \otimes \tilde{d}q$, we have

$$\bar{X}_g = \frac{\partial g}{\partial q}\frac{\partial}{\partial p} - \frac{\partial g}{\partial p}\frac{\partial}{\partial q}, \tag{5.31}$$

which can be established by verifying that $\tilde{\omega}(\bar{X}_g) = \tilde{d}g$. Therefore we have

$$\{f,g\} = \langle df, \bar{X}_g \rangle = \frac{\partial g}{\partial q}\frac{\partial f}{\partial p} - \frac{\partial g}{\partial p}\frac{\partial f}{\partial q}.$$

This is what is usually called the *Poisson bracket* of the functions f and g. The definition (5.30) gives it a geometrical significance, and shows that the Poisson bracket is actually independent of the coordinates. It depends only on $\tilde{\omega}$.

Exercise 5.6

(a) Defining $\bar{X}_H \equiv \overline{dH}$, show that for any function K,

$$\{K,H\} = \bar{X}_H(K) = dK/dt, \tag{5.32}$$

where t is the parameter such that $\bar{X}_H = d/dt$. Thus, the Poisson bracket of a function with the Hamiltonian gives the time-derivative of that function along a solution curve. In particular, constants of the motion have vanishing Poisson bracket with H.

(b) Show that the Poisson brackets satisfy the Jacobi identity

$$\{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} = 0 \tag{5.33}$$

for any C^2 functions f, g, h.

(c) Show from this that

$$[\bar{X}_f, \bar{X}_g] = -\bar{X}_{\{f,g\}}, \tag{5.34}$$

so that the Hamiltonian vector fields form a Lie algebra.

5.8 Many-particle systems: symplectic forms

In general one deals with systems which have more than one degree of freedom, so there are more than one q and p. A particle in three dimensions has 3 qs and 3 ps, so phase space is 6-dimensional. A system containing N such particles has a $6N$-dimensional phase space. If we consider now a general system with n degrees of freedom, then phase space is $2n$-dimensional, and all the above results still hold if we take the two-form $\tilde{\omega}$ to be

$$\tilde{\omega} = \sum_{A=1}^{n} \tilde{d}q^A \wedge \tilde{d}p_A. \tag{5.35}$$

Such an $\tilde{\omega}$ is called a *symplectic form*, and then phase space is a *symplectic manifold*.

Exercise 5.7

(a) Show that f is a constant of the motion if $\bar{X}_f = (\overline{\tilde{d}f})$ is an invariant of H, i.e.

$$£_{\bar{X}_f} H = 0. \tag{5.36}$$

(Refer to exercise 5.6.)

(b) Define a volume-form $\tilde{\sigma}$ for phase space by

$$\tilde{\sigma} = \underbrace{\tilde{\omega} \wedge \ldots \wedge \tilde{\omega}}_{n \text{ times}}, \tag{5.37}$$

where $2n$ is the dimension of the space, Show that $\tilde{\sigma} \neq 0$ and that a Hamiltonian vector field $\bar{U}$ is divergence-free in this volume measure. Said another way, this volume in phase space is preserved by the time-evolution of the system. This is known as *Liouville's theorem*.

Exercise 5.8

We now prove the remarks made in §3.12 about the relation between Killing vectors and conserved quantities. For particle motion the coordinates of phase space are $\{q^A, p_A\} = \{x^i, p_i = mv_i\}$ and the Hamiltonian is $H = (1/2m)g^{ij}p_ip_j + \Phi(x^i)$. Prove that if $\bar{U}$ is a Killing vector and if Φ is constant along $\bar{U}$, then its conjugate momentum, $p_{\bar{U}} \equiv U^i p_i$, is a conserved quantity. Hint: using exercise 5.7, define $\bar{X}_f$ as the vector field in phase space whose space components equal $\bar{U}$ and whose momentum components vanish. Show that

$$£_{\bar{X}_f} H = 0,$$

and find f from equation (5.31).

5.9. Linear dynamical systems: the symplectic inner product and conserved quantities

Even more strikingly simple ways of formulating conservation laws are possible for linear systems, by which we mean dynamical systems whose Hamiltonian has the form

$$H = \sum_{A,B=1}^{n} (T^{AB}p_A p_B + V_{AB}q^A q^B), \tag{5.38}$$

where T^{AB} and V_{AB} are independent of the p_As and q^As. This system is called linear because the equations of motion are linear in $\{q^A, p_A\}$:

$$\frac{\mathrm{d}p_A}{\mathrm{d}t} = -\frac{\partial H}{\partial q^A} = -\sum_B V_{AB}q^B, \tag{5.39}$$

$$\frac{\mathrm{d}q^A}{\mathrm{d}t} = \frac{\partial H}{\partial p_A} = \sum_B T^{AB}p_B. \tag{5.40}$$

Notice that we can take $T^{AB} = T^{BA}$ and $V_{AB} = V_{BA}$, since the antisymmetric part of, say, T^{AB} would make no contribution to H when contracted with the symmetric expression $p_A p_B$.

The linearity of the system ensures that if $\{q^A_{(1)}, p_{(1)A}\}$ and $\{q^A_{(2)}, p_{(2)A}\}$ are solutions then so is $\{\alpha q^A_{(1)} + \beta q^A_{(2)}, \alpha p_{(1)A} + \beta p_{(2)A}\}$ for arbitrary constants α and β. Thus, this phase space is not just a manifold; it has a natural vector-space structure as well. A vector space is, of course, a kind of manifold, since it has a map into R^n, but it is a manifold which can be *identified* with its tangent space at every point. That is, since a curve in a vector space is a sequence of vectors, the tangent to the curve is just the derivative of the vectors along the curve, which is another vector, i.e. another element of the vector space. A vector space is its own tangent space. More than this, *all* the tangent spaces T_P have a natural identification with each other: we are able to speak about vectors in different T_Ps as being equal or not, simply by whether or not their components are equal. (This means a vector space is a *flat* manifold: see chapter 6.)

Since a point in phase space is a vector, we can use the symplectic form $\tilde{\omega}$ to define an inner product between elements of phase space. If $\bar{Y}_{(1)}$ is the vector whose components are $\{q^A_{(1)}, p_{(1)A}, A = 1, \ldots, N\}$ and if $\bar{Y}_{(2)}$ similarly has components $\{q^A_{(2)}, p_{(2)A}\}$, then their *symplectic inner product* is defined as

$$\tilde{\omega}(\bar{Y}_{(1)}, \bar{Y}_{(2)}) = \sum_A (q^A_{(1)}p_{(2)A} - q^A_{(2)}p_{(1)A}). \tag{5.41}$$

If $\bar{Y}_{(1)}(t)$ and $\bar{Y}_{(2)}(t)$ are solution curves, then their symplectic inner product is independent of time t. To prove this, we simply substitute the equations of motion into the expression for $\mathrm{d}\tilde{\omega}(\bar{Y}_{(1)}, \bar{Y}_{(2)})/\mathrm{d}t$ (sum on repeated indices here):

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\omega}(\bar{Y}_{(1)}, \bar{Y}_{(2)}) = \frac{\mathrm{d}}{\mathrm{d}t}(q^A_{(1)})p_{(2)A} + q^A_{(1)}\frac{\mathrm{d}}{\mathrm{d}t}p_{(2)A}$$

$$-\frac{\mathrm{d}}{\mathrm{d}t}(q^A_{(2)})p_{(1)A} - q^A_{(2)}\frac{\mathrm{d}}{\mathrm{d}t}p_{(1)A}$$

$$= T^{AB}p_{(1)B}p_{(2)A} + V_{AB}q^A_{(1)}q^B_{(2)} - T^{AB}p_{(1)A}p_{(2)B} - V_{AB}q^A_{(2)}q^B_{(1)}.$$

From the symmetry of T^{AB} and V_{AB} we conclude:

$$\frac{d}{dt}\tilde{\omega}(\bar{Y}_{(1)}, \bar{Y}_{(2)}) = 0 \tag{5.42}$$

if $\bar{Y}_{(1)}(t)$ and $\bar{Y}_{(2)}(t)$ are solutions.

The symplectic inner produce enables us to define in an elegant way certain *conserved quantities* associated with solutions. At first sight this may not be obvious: although the symplectic inner product is conserved, the symplectic inner product of a solution with itself vanishes identically. The trick is to use an invariance of the system (i.e. of T^{AB} and V_{AB}) to generate from one solution $\bar{Y}$ another closely related one. For example, suppose T^{AB} and V_{AB} are independent of time. Then the equations of motion tell us that if $\bar{Y}(t)$ is a solution, so is $d\bar{Y}/dt$. We define the canonical energy E_c of the solution $\bar{Y}$ to be

♦ $$E_c(\bar{Y}) = \tilde{\omega}\left(\frac{d\bar{Y}}{dt}, \bar{Y}\right). \tag{5.43}$$

It is easy to verify that $E_c(\bar{Y})$ is just the value of the Hamiltonian on the solution $\bar{Y}$.

Other conserved quantities are just as easy to derive. It usually happens that T^{AB} and V_{AB} depend on the coordinates $\{x^i\}$ of the manifold in which the dynamical system is defined (Euclidean space for nonrelativistic dynamics). If, as in exercise 5.8, there is some vector field $\bar{U}$ for which

$$£_{\bar{U}}T^{AB} = 0 = £_{\bar{U}}V_{AB}, \tag{5.44}$$

then there is a conserved quantity associated with $\bar{U}$. (In computing $£_{\bar{U}}T^{AB}$ it is important to distinguish between indices A, B which refer to coordinates in phase space and the tensorial character of T^{AB} on the original manifold. The quantities T^{AB} may be scalars, or tensors on the original manifold, depending upon whether the quantities q^A are scalars or tensors of higher order. The indices A and B are *labels*; they do not imply that T^{AB} should be treated as a tensor of type $\binom{2}{0}$ when computing the Lie derivative with respect to $\bar{U}$, because $\bar{U}$ is a vector field in the original manifold, not in phase space.) As before, if $\bar{Y}$ is a solution, then so is $£_{\bar{U}}\bar{Y}$. (Again the same remark applies: this is a derivative in the original manifold, not in a phase space.) We therefore define the (conserved) *canonical $\bar{U}$-momentum*

♦ $$P_{\bar{U}}(\bar{Y}) = \tilde{\omega}(£_{\bar{U}}\bar{Y}, \bar{Y}). \tag{5.45}$$

The reader is invited to try a simple example, such as the one given in exercise 5.8, to verify that the usual conserved quantity does indeed appear.

Although our discussion has been confined to systems with a finite number

(N) of degrees of freedom, the formalism generalizes in a straightforward way to continuous systems, such as wave equations. Readers familiar with the Klein–Gordon equation may recognize the symplectic inner product: the integral of the conserved Klein–Gordon current density $\psi^*\dot{\psi} - \psi\dot{\psi}^*$ is just (to within constant factors) $\tilde{\omega}(\psi^*, \psi)$. A discussion of the canonical conserved quantities for waves in fluids, with application to questions of stability, can be found in Friedman & Schutz (1978) (see bibliography).

5.10 *Fiber bundle structure of the Hamiltonian equations*

Our original statement in §5.4 that we defined phase space to be the manifold whose coordinates are p and q, hid a lot of interesting and important structure. Suppose a dynamical system has the N coordinates $\{q^i\}$ corresponding to its N degrees of freedom. These define a manifold called *configuration space* M, and the evolution of the dynamical system in time is described by a curve $q^i(t)$ in M. The Lagrangian $\mathscr{L}$ is a function of q^i and $\mathrm{d}q^i/\mathrm{d}t$, and so is a function on TM, the tangent bundle of M. We now show that the momentum

$$p_i = \partial\mathscr{L}/\partial(q^i{}_{,t}), \tag{5.46}$$

is a one-form field on M, a cross-section of the cotangent bundle T^*M. We show this by its transformation properties. Let us define new coordinates for M

$$Q^{j'} = Q^{j'}(q^i). \tag{5.47}$$

Then the new momenta are

$$P_{j'} = \frac{\partial\mathscr{L}}{\partial Q^{j'}{}_{,t}} = \frac{\partial\mathscr{L}}{\partial q^k{}_{,t}}\frac{\partial q^k{}_{,t}}{\partial Q^{j'}{}_{,t}}. \tag{5.48}$$

Now, both $q^k{}_{,t}$ and $Q^{j'}{}_{,t}$ are elements of the fiber over any point P, and coordinates on this fiber undergo a natural change induced by (5.47). That is, if $\bar{V}$ is any vector at P its components change by

$$V^{j'} = \Lambda^{j'}{}_k V^k, \quad V^k = \Lambda^k{}_{j'} V^{j'}.$$

This applies as well to the velocity vector $q^k{}_{,t}$:

$$q^k{}_{,t} = \Lambda^k{}_{j'} Q^{j'}{}_{,t} \Longrightarrow \frac{\partial q^k{}_{,t}}{\partial Q^{j'}{}_{,t}} = \Lambda^k{}_{j'}.$$

Using this in (5.48) gives

$$P_{j'} = \Lambda^k{}_{j'} p_k, \tag{5.49}$$

so that the momentum is indeed a one-form.

It follows that phase space, whose coordinates are $\{q^i, p_i\}$, is nothing but the *cotangent bundle* T^*M, and the Hamiltonian is a function on this bundle. What is more, the symplectic form,

$$\tilde{\omega} = \tilde{\mathrm{d}}q^i \wedge \tilde{\mathrm{d}}p_i,$$

(summation convention employed) is independent of the coordinates in M. The transformation for it is

$$\begin{aligned} Q^{j'} &= Q^{j'}(q^i) \Longrightarrow \tilde{d}Q^{j'} = \Lambda^{j'}{}_i \tilde{d}q^i, \\ P_{j'} &= \Lambda^k{}_{j'} p_k \Longrightarrow \tilde{d}P_{j'} = \Lambda^k{}_{j',l} p_k \tilde{d}q^l + \Lambda^k{}_{j'} \tilde{d}p_k. \end{aligned} \tag{5.50}$$

(Remember that this $\tilde{d}$ operator acts in T^*M, not in M, and that the functions $\Lambda^k{}_{j'}$ are functions only of the coordinates of M). Then we find

$$\tilde{d}Q^{j'} \wedge \tilde{d}P_{j'} = \Lambda^{j'}{}_i \Lambda^k{}_{j',l} p_k \, \tilde{d}q^i \wedge \tilde{d}q^l + \Lambda^{j'}{}_i \Lambda^k{}_{j'} \tilde{d}q^i \wedge \tilde{d}p_k. \tag{5.51}$$

Now we also have

$$\Lambda^{j'}{}_i \Lambda^k{}_{j'} = \delta_i{}^k \Longrightarrow \Lambda^{j'}{}_i \Lambda^k{}_{j',l} = -\Lambda^{j'}{}_{i,l} \Lambda^k{}_{j'}.$$

So (5.51) becomes

$$\tilde{d}Q^{j'} \wedge \tilde{d}P_{j'} = -\Lambda^{j'}{}_{i,l} \Lambda^k{}_{j'} p_k \tilde{d}q^i \wedge \tilde{d}q^l + \tilde{d}q^i \wedge \tilde{d}p_i.$$

The first term on the right-hand side vanishes because

$$\Lambda^{j'}{}_{i,l} = \frac{\partial^2 Q^{j'}}{\partial q^i \partial q^l}$$

is symmetric in i and l and is contracted with the antisymmetric form $\tilde{d}q^i \wedge \tilde{d}q^l$. Therefore $\tilde{\omega}$ is independent of the coordinates of M and is a *natural* structure on the cotangent bundle T^*M. Moreover, T^*M is always orientable, since the volume-form $\tilde{\sigma}$ defined in exercise 5.7(b) is nowhere zero.

Clearly, although our examples treated the fiber structure as trivial (i.e. as a product of the q-space and p-space), it is possible to have nontrivial manifolds M and fiber bundles T^*M, in which all the coordinate-dependent formulae above are valid only in local coordinate patches. Even an example as simple as that of a bead constrained to move on the surface of a sphere has a nontrivial bundle structure for phase space, as we pointed out in §2.11.

C Electromagnetism

5.11 Rewriting Maxwell's equations using differential forms

Maxwell's equations, written in conventional form but with units where $c = \mu_0 = \epsilon_0 = 1$, are

$$\nabla \times \mathbf{B} - \frac{\partial}{\partial t}\mathbf{E} = 4\pi \mathbf{J}, \tag{5.52a}$$

$$\nabla \times \mathbf{E} + \frac{\partial}{\partial t}\mathbf{B} = 0, \tag{5.52b}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{5.52c}$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho. \tag{5.52d}$$

In writing these equations we have, of course, used the curl and divergence operations of ordinary flat three-space.

What we shall show below is that there exists a way of writing these equations using only the concepts of the metric and the exterior derivative. First we rewrite the equations in their relativistically invariant form† by first defining the *Faraday two-form* $\tilde{F}$, whose components are

$$\blacklozenge \qquad (F_{\mu\nu}) \equiv \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{5.53}$$

(Here, as in §2.31, Greek indices run over t, x, y, z.)

Exercise 5.9

Prove that under a spatial rotation $F_{\mu\nu}$ transforms in such a way that both **E** and **B** transform as three-vectors.

In terms of the Faraday tensor, Maxwell's equations take a particularly simple form. For instance, the four equations (5.52b, c) are just

$$F_{[\mu\nu,\gamma]} = 0 \Leftrightarrow \tilde{d}\tilde{F} = 0, \tag{5.54}$$

where we have used the square-bracket notation to denote antisymmetrization.

Exercise 5.10

(a) Prove that (5.54) constitutes four linearly independent equations.

(b) Evaluate (5.54) for the components of $\tilde{F}$ given by (5.53) and prove their equality to (5.52b, c).

As for the rest of the equations, if we introduce the special-relativistic metric whose components in this coordinate system are

$$(g_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.55}$$

† For readers to whom this is unfamiliar, recall that Maxwell's equations are the correct theory for light and that special relativity was invented to explain certain properties of light, so the theory is *already* relativistically correct. All we do here is to find a convenient form for the equations.

then we can define an antisymmetric $\binom{2}{0}$ tensor $\mathbf{F}$ whose components are

$$F^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta},$$

$$(F^{\mu\nu}) = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{5.56}$$

Exercise 5.11

Prove equation (5.56).

Then the remaining equations are

$$F^{\mu\nu}{}_{,\nu} = 4\pi J^\mu, \tag{5.57}$$

where we have defined the current four-vector to have components $\{J^t = \rho, J^i = (\mathbf{J})^i \text{ for } i = x, y, z\}$.

Exercise 5.12

Prove that the four equations (5.57) are just the same as (5.52a–d).

So far we have stuck to Lorentz coordinates because, while (5.54) is coordinate-independent, (5.57) is not a valid tensor equation in every coordinate system (recall exercise 4.15). On the other hand, we saw in exercise 4.23 how to define the divergence of an antisymmetric $\binom{2}{0}$ tensor (two-vector) if we have a volume-form. Because we have a metric, and because $\{\partial/\partial t, \partial/\partial x, \partial/\partial y, \partial/\partial z\}$ form an orthonormal basis in this metric, the preferred volume-form is

$$\tilde{\omega} = \tilde{d}t \wedge \tilde{d}x \wedge \tilde{d}y \wedge \tilde{d}z.$$

The following exercise develops the argument.

Exercise 5.13

(a) Define the two-form $^*\tilde{F}$ to be the contraction

$$^*\tilde{F} \equiv \tfrac{1}{2}\tilde{\omega}(\mathbf{F}), \tag{5.58}$$

i.e.

$$(^*\tilde{F})_{\mu\nu} = \tfrac{1}{2}\omega_{\alpha\beta\mu\nu}F^{\alpha\beta}.$$

This is, of course, the dual of $\mathbf{F}$ introduced in chapter 4. Find the components $(^*\tilde{F})_{\mu\nu}$ in terms of $\mathbf{E}$ and $\mathbf{B}$.

(b) Define the three-form $^*\tilde{J}$ by the contraction

$$^*\tilde{J} \equiv \tilde{\omega}(\bar{J}), \tag{5.59}$$

and show that (5.57) is equivalent to

$$\tilde{d}(^*\tilde{F}) = 4\pi{}^*\tilde{J}. \tag{5.60}$$

By exercise 4.23 this is also

$$\operatorname{div}_\omega \mathbf{F} = 4\pi\bar{J}. \tag{5.61}$$

Note the great formal similarity between the two halves of our new form for Maxwell's equations:

♦ $$\tilde{d}\tilde{F} = 0, \tag{5.54}$$

♦ $$\tilde{d}^*\tilde{F} = 4\pi{}^*\tilde{J}. \tag{5.60}$$

Note also that they now are completely coordinate-free, so they have this form in *any* manifold with metric (because the metric was needed to obtain $^*\tilde{F}$ from $\tilde{F}$). The similarity between (5.54) and (5.60) is deep in Maxwell's equations. Note that the * operation on $\tilde{F}$ simply results in an exchange of **E** and **B** (cf. exercise 5.13(a)), and recall also that $\bar{J}$ was the *electrical* current density. If there were magnetic monopoles we would have two current densities, $\bar{J}_\mathrm{e}$ and $\bar{J}_\mathrm{m}$, and Maxwell's equations would take the symmetric form

$$\tilde{d}\tilde{F} = 4\pi{}^*\tilde{J}_\mathrm{m}, \quad \tilde{d}^*\tilde{F} = 4\pi{}^*\tilde{J}_\mathrm{e}. \tag{5.62}$$

Exercise 5.14

(a) Prove (5.62).

(b) Prove by exterior differentiation that equation (5.60) guarantees conservation of charge, i.e. that

$$\operatorname{div}(\bar{J}) = 0. \tag{5.63}$$

Exercise 5.15

Establish the integral theorem for charge in the following way.

(a) Choose *any* oriented three-dimensional hypersurface $\mathscr{H}$ and restrict (5.60) to it. Prove that restriction commutes with exterior differentiation, i.e. that

$$\tilde{d}[(^*\tilde{F})|_{\mathscr{H}}] = (\tilde{d}^*\tilde{F})|_{\mathscr{H}}.$$

(b) Choose a region $\mathscr{D}$ of $\mathscr{H}$, with boundary $\partial\mathscr{H}$. Integrate the restriction of (5.60) over $\mathscr{D}$ and apply Stokes' theorem to find (appropriate restrictions implied)

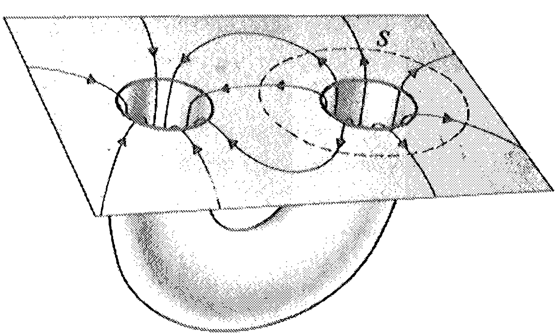
$$\int_{\mathscr{D}} {}^*\tilde{J} = \frac{1}{4\pi}\oint_{\partial\mathscr{D}} {}^*\tilde{F}.$$

(c) In the case where $\mathscr{H}$ is a hypersurface $t = \text{const}$ in Minkowski spacetime and $\partial\mathscr{D}$ is a sphere, show that this gives the total charge in $\mathscr{D}$ as an integral of the normal component of the electric field over $\partial\mathscr{D}$.

5.12 Charge and topology

Since we can now formulate Maxwell's equations on any manifold with a metric, we can mention two attempts which have been made to resolve the puzzling question 'what is charge?' by answering 'charge is topology'. The first explanation, due to J. A. Wheeler (1962), is extremely simple. Consider figure 5.2, in which a hypersurface $t = \text{const}$ of some hypothetical spacetime is depicted. The lines drawn are integral curves of **E**. There is no charge density anywhere, and these integral curves are either closed (threading through the handle, out one hole, and down the other) or infinite (though they pass through the handle). Consider what an experimenter who measures **E** on the sphere S surrounding one hole will deduce: the integral $\int_S \tilde{F}|_S$ will certainly not vanish (**E** is outward-pointing all over S), and he will say the hole has positive charge. Likewise, a sphere around the other hole would give it negative charge, of exactly the same magnitude. (The calculation of exercise 5.15 fails because S does not divide the manifold into an inside and outside, cf. figure 4.10.) So this is a model for 'charge without charge', which has the bonus of explaining why negative charges equal positive charges. It has two drawbacks: first, no-one pretends to have a solution to, say, Einstein's equations which gives a geometry for spacetime that looks like this; and second, it is perhaps philosphically displeasing to think of

Fig. 5.2. A 'wormhole' or handle attached to a three-dimensional manifold with one dimension suppressed. Lines of force can thread through the handle, come out, and go backdown again to give each 'mouth' the appearance of charge in a charge-free space.



two charges, which may be separated by huge distances, linked together by their own special 'handle'.

The second explanation is more sophisticated, using a manifold made non-orientable by a special construction of the handle. This is due to Sorkin (1977) (reference in the bibliography of chapter 4). In this model, both holes have the *same* charge and so may be assumed to be close together, forming what to an outside observer looks like a single charge of twice the strength of each hole. Here the breakdown in exercise 5.15 occurs because the manifold is nonorientable. This model overcomes the second objection to Wheeler's picture, but not the first. And neither model explains why two unrelated charges should be equal. Nevertheless they illustrate a maxim which is becoming more convincing all the time: there is more to theoretical physics than just its local differential equations!

5.13 The vector potential

The existence of a 'vector potential' for Maxwell's equations follows naturally from (5.54). Since $\tilde{F}$ is a closed two-form, there is a one-form $\tilde{A}$ such that

$$\blacklozenge \qquad \tilde{F} = \tilde{d}\tilde{A} \tag{5.64}$$

in some neighborhood of any point. This one-form can be mapped into a vector by the metric, and this is called the vector potential. A more natural concept is, of course, the one-form potential. Note that $\tilde{A}$ is not uniquely defined: $\tilde{A}' = \tilde{A} + \tilde{d}f$, for an arbitrary function f, also gives $\tilde{F}$ in (5.64). This is a *gauge transformation*. Note also that if magnetic monopoles exist, then $\tilde{d}\tilde{F}$ does not vanish everywhere. By our discussion of exact forms in chapter 4, it will be possible to define $\tilde{A}$ only in simple regions which contain no magnetic monopoles. In particular, in a region of spacetime containing the world-line of a magnetic monopole, the one-form potential *cannot* be consistently defined everywhere.

Exercise 5.16

(a) Show that, if a one-form potential $\tilde{A}$ exists, then in nonrelativistic language it is related to the scalar potential ϕ and the vector potential A^i by $\phi = A_0, A^i$ (vector potential) $= -A_i$ (one-form), where indices refer to the coordinates of (5.52).

(b) Show how ϕ and A^i defined in (a) change under a gauge transformation.

(c) To illustrate the problems caused to the one-form potential $\tilde{A}$ by magnetic monopoles, consider a situation with charges and *no* monopoles, but in which one defines a one-form potential $\tilde{\alpha}$ for $^*\tilde{F}$ by the equation $^*\tilde{F} = \tilde{d}\tilde{\alpha}$.

(By the duality between electric and magnetic fields under the

*-operation, $\tilde{\alpha}$ should have the same problems with electric charge as $\tilde{A}$ has with magnetic.) Write down Maxwell's equations in terms of $\tilde{\alpha}$ and show that $\tilde{\alpha}$ exists in regions that contain no charge and that can be shrunk to zero. Show this by finding an explicit solution for $\tilde{\alpha}$ in the case of a single isolated static charge q.

5.14 Plane waves: a simple example

Plane electromagnetic waves, as is well-known, travel at the speed of light. Consider a particular Faraday tensor $F^{\alpha\beta}$, all of whose components are functions only of $u \equiv t - x$ (recall that we are using units in which $c = 1$):

$$F^{\alpha\beta} = A^{\alpha\beta}(t-x) = A^{\alpha\beta}(u). \tag{5.65}$$

What are the conditions that this satisfy the empty-space equations $\mathrm{d}\tilde{F} = 0$, $\mathrm{d}{}^*\tilde{F} = 0$? From (5.65) we have

$$\begin{aligned}\mathrm{d}\tilde{F} &= \tilde{\mathrm{d}}(\tfrac{1}{2}F_{\mu\nu}\tilde{\mathrm{d}}x^{\mu} \wedge \tilde{\mathrm{d}}x^{\nu}) = \tfrac{1}{2}\tilde{\mathrm{d}}(F_{\mu\nu}) \wedge \tilde{\mathrm{d}}x^{\mu} \wedge \tilde{\mathrm{d}}x^{\nu} \\ &= \tfrac{1}{2}(\mathrm{d}A_{\mu\nu}/\mathrm{d}u)\tilde{\mathrm{d}}u \wedge \tilde{\mathrm{d}}x^{\mu} \wedge \tilde{\mathrm{d}}x^{\nu}.\end{aligned}$$

From (5.53) it is easy to deduce

$$\begin{aligned}\tilde{\mathrm{d}}\tilde{F} = \Big[&\frac{\mathrm{d}}{\mathrm{d}u}(B_z - E_y)\tilde{\mathrm{d}}t \wedge \tilde{\mathrm{d}}x \wedge \tilde{\mathrm{d}}y + \frac{\mathrm{d}}{\mathrm{d}u}(B_x)\tilde{\mathrm{d}}t \wedge \tilde{\mathrm{d}}y \wedge \tilde{\mathrm{d}}z \\ &+ \frac{\mathrm{d}}{\mathrm{d}u}(-B_x)\tilde{\mathrm{d}}x \wedge \tilde{\mathrm{d}}y \wedge \tilde{\mathrm{d}}z + \frac{\mathrm{d}}{\mathrm{d}u}(-B_y - E_z)\tilde{\mathrm{d}}t \wedge \tilde{\mathrm{d}}x \wedge \tilde{\mathrm{d}}z\Big],\end{aligned}$$

the vanishing of which implies (ignoring any static fields)

$$B_z = E_y, \qquad B_y = -E_z, \qquad B_x = 0. \tag{5.66}$$

Exercise 5.17

Show that the equation $\tilde{\mathrm{d}}{}^*\tilde{F} = 0$ implies

$$B_z = E_y, \qquad B_y = -E_z, \qquad E_x = 0. \tag{5.67}$$

By this exercise we see that a plane electromagnetic wave has transverse electric and magnetic fields (i.e. perpendicular to its direction of propagation), and that these are determined by two independent functions, $E_y(u)$ and $E_z(u)$, corresponding to the two independent polarizations of the wave.

D Dynamics of a perfect fluid

5.15 Role of Lie derivatives

By a 'perfect' fluid we mean one which has no viscosity and moves adiabatically, i.e. with no heat conduction. It is well-known that such a fluid obeys certain local conservation laws: during its motion any fluid element has a constant mass, entropy, and – in some sense – vorticity. These conservation laws are usually derived using ordinary vector calculus, and can seem rather complicated. From the geometric point of view, the existence of a flow suggests immediately the use of the Lie derivative, and we now show that the local conservation laws become much more transparent when framed with Lie derivatives.

5.16 The comoving time-derivative

We have seen in exercise 4.22 that the equation of continuity, whose conventional form is

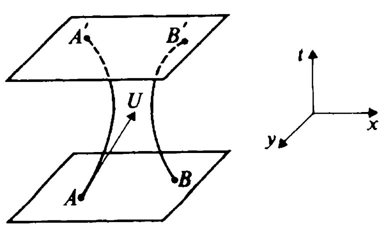
$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \bar{V}) = 0$$

takes the form

$$\blacklozenge \qquad \left(\frac{\partial}{\partial t} + £_{\bar{V}}\right)(\rho\tilde{\omega}) = 0, \tag{5.68}$$

where $\tilde{\omega} = \tilde{d}x \wedge \tilde{d}y \wedge \tilde{d}z$ is the volume three-form of Euclidean space. The operator $(\partial/\partial t + £_{\bar{V}})$ is a natural time-derivative operator following a particular fluid element. To see this, think not of space but of the four-dimensional manifold called Galilean spacetime, whose coordinates are (x, y, z, t) (see §2.10). Any hypersurface $t = \text{const}$ is in fact Euclidean space. Then the motion of a fluid element describes a curve on spacetime, called the world-line of the element. In figure 5.3, two such world-lines (AA' and BB') are drawn. For an infinitesimal change in time $\mathrm{d}t$, a point on this curve moves from the point with coordinates (x, y, z, t) to the one with coordinates $(x + V^x\mathrm{d}t, y + V^y\mathrm{d}t, z + V^z\mathrm{d}t, t + \mathrm{d}t)$. If we call $\bar{U}$ the tangent to the world line in the four-dimensional manifold, then it clearly has components $(V^x, V^y, V^z, 1)$. The time-derivative following a fluid element is simply $£_{\bar{U}}$, the natural derivative along the world-line of the element.

Fig. 5.3. Two moments of Galilean time and the world lines AA' and BB' of two particles. The vector $\bar{U}$ is the tangent to AA' parameterized by time t.



Exercise 5.18

Using equation (2.7) show that

$$\pounds_{\bar{U}}\bar{W} = \left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right)\bar{W}, \tag{5.69}$$

where $\bar{W}$ is any vector field in the hypersurface $t =$ const, i.e. any purely spatial vector field ($W^t \equiv 0$).

Equation (5.69) clearly holds if $\bar{W}$ is replaced by *any* $\binom{n}{0}$ tensor which is entirely in the three-space $t =$ const. It might seem that the notion of a tensor being purely spatial is not invariant under coordinate changes in the four-dimensional manifold, since it simply says that all the t-components of the tensor vanish. This is acceptable here, however, because of the rigid distinction made in non-relativistic physics between space and time.

Exercise 5.19

The most general kind of coordinate transformation which remains 'natural' to the fiber-bundle structure of Galilean spacetime (§2.10) is

$$t' = g(t); \qquad x^{i'} = f^{i'}(x^i, t), i = 1, 2, 3. \tag{5.70}$$

Show that under this transformation a $\binom{n}{0}$ tensor **A** with no time-components ($\mathbf{A}(\ldots, \tilde{\omega}^t, \ldots) = 0$) remains one with no time-components, and a $\binom{0}{n}$ tensor **B** with no spatial components (i.e. only $B_{t\ldots t}$ is nonzero) remains one with no spatial components.

5.17 Equation of motion

The condition that the flow be adiabatic means that the total entropy of a fluid element must be conserved. It is convenient to work with S, the *specific* entropy (entropy per unit mass). This must clearly be *constant* during the flow:

$$\blacklozenge \qquad \left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right)S = 0. \tag{5.71}$$

The Euler equation of motion for a fluid whose pressure is p and which moves in a gravitational field whose potential is Φ can be written in Cartesian coordinates as

$$\frac{\partial}{\partial t}V^i + V^j\frac{\partial}{\partial x^j}V^i + \frac{1}{\rho}\frac{\partial}{\partial x^i}p + \frac{\partial}{\partial x^i}\Phi = 0. \tag{5.72}$$

There are two reasons that this equation is valid only in Cartesian coordinates: first, some indices i are up and some are down, and only in an orthonormal basis does this make no difference; second, the term $\partial V^i/\partial x^j$ transforms like a $\binom{1}{1}$ tensor *only* if the transformation matrix $\Lambda^{i'}{}_j$ is independent of position (exercise 4.5), which is true for a transformation from one Cartesian frame to another. The usual way to adapt it to arbitrary coordinates is to introduce the covariant derivative, which is defined in the chapter on Riemannian geometry. Here we show that there is a different, and very instructive, approach. First, note that the first two terms of (5.72) can be written as

$$\frac{\partial V_i}{\partial t} + V^j \frac{\partial V_i}{\partial x^j},$$

since there is no difference between V^i and V_i in Cartesian coordinates. (We use here, of course, the fact that the three-dimensional space has a metric tensor.) Next, replace the derivative $V^j\partial/\partial x^j$ with the Lie derivative (equation (3.14)) of the *one-form* $\tilde{V} = g(\bar{V},\ )$:

$$\begin{aligned}(\pounds_{\bar{V}}\tilde{V})_i &= V^{j'}\frac{\partial}{\partial x^j} V_i + V_j \frac{\partial}{\partial x^i} V^j \\ &= V^j \frac{\partial}{\partial x^j} V_i + \frac{1}{2}\frac{\partial}{\partial x^i}(V_j V^j),\end{aligned}$$

where in obtaining the final expression we again used the fact that $V_j = V^j$. Therefore we find

$$V^j \frac{\partial}{\partial x^j} V_i = (\pounds_{\bar{V}}\tilde{V})_i - \frac{\partial}{\partial x^i}(\tfrac{1}{2}V^2). \qquad (5.73)$$

Both terms on the right-hand side are tensors in any coordinate system! Therefore (5.72) becomes the frame-independent expression

$$\blacklozenge \qquad \left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right)\tilde{V} + \frac{1}{\rho}\,\tilde{d}p + \tilde{d}(\Phi - \tfrac{1}{2}V^2) = 0. \qquad (5.74)$$

In this the role of the metric is crucial but hidden: it is required to form $\tilde{V}$ from $\bar{V}$, and hence to form $V^2 = \tilde{V}(\bar{V})$.

5.18 Conservation of vorticity

Now we are in a position to consider conservation of vorticity. In conventional terms, the vorticity is the curl of the velocity, $\nabla \times \bar{V}$. As we saw in chapter 4, this is properly the exterior derivative $\tilde{d}\tilde{V}$. Now, exterior differentiation and Lie differentiation commute (and of course $\tilde{d}$ and $\partial/\partial t$ commute since $\tilde{d}$ only involves spatial derivatives), so we find from (5.74)

$$\left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right) \mathrm{d}V = \frac{1}{\rho^2}\, \mathrm{d}\rho \wedge \mathrm{d}p. \qquad (5.75)$$

(We have dropped tildes over symbols for clarity.) There are two cases to be considered. The easier is when the fluid obeys an equation of state $p = p(\rho)$. Then $\mathrm{d}p \wedge \mathrm{d}\rho \equiv 0$ and we find that the vorticity two-form $\mathrm{d}V$ obeys the local (or convective) conservation law

$$\left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right) \mathrm{d}V = 0. \qquad (5.76)$$

This is the *Helmholtz circulation theorem*, written in its most natural form. A different result holds, however, if the more general equation of state $p = p(\rho, S)$ obtains. Then the right-hand side of (5.75) does not vanish, but its wedge product with $\mathrm{d}S$ does:

$$\mathrm{d}S \wedge \mathrm{d}\rho \wedge \mathrm{d}p = 0. \qquad (5.77)$$

Exercise 5.20
Prove (5.77).

The exterior derivative of (5.71) gives

$$\left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right) \mathrm{d}S = 0. \qquad (5.78)$$

Therefore we can wedge $\mathrm{d}S$ with (5.75) to get

$$\mathrm{d}S \wedge \left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right) \mathrm{d}V = 0,$$

or

♦ $$\left(\frac{\partial}{\partial t} + \pounds_{\bar{V}}\right) \mathrm{d}S \wedge \mathrm{d}V = 0. \qquad (5.79)$$

This equation is the most general vorticity conservation law. It is called *Ertel's theorem*.

The meaning of the three-form $\mathrm{d}S \wedge \mathrm{d}V$ may not be immediately apparent, but it is possible to convert (5.79) into a conservation law for a scalar. The reason is that there is another conserved three-form, $\rho\omega$, and any two three-forms in a three-dimensional space are proportional. Therefore there is a scalar function α such that

$$\mathrm{d}S \wedge \mathrm{d}V = \alpha\rho\omega, \qquad (5.80)$$

and (5.68) and (5.79) then give the scalar equation

$$\left(\frac{\partial}{\partial t} + £_{\bar{V}}\right)\alpha = 0.$$

It can be shown that, in conventional vector notation,

$$\alpha = \frac{1}{\rho} \nabla S \cdot \nabla \times V. \tag{5.81}$$

Exercise 5.21

Prove (5.81). (Hint: express both sides of (5.80) in terms of $dx \wedge dy \wedge dz$.)

In the notation introduced in chapter 4 we have

$$\alpha = \frac{1}{\rho} \epsilon^{ijk} S_{,i} V_{k,j}. \tag{5.82}$$

Therefore α is the *dual* of $dS \wedge dV$ with respect to $\rho\omega$. The conservation of α is then a natural consequence of the conservation of $dS \wedge dV$: the fact that $\rho\omega$ is conserved means that forming duals with respect to it is an operation which is also conserved, i.e. which commutes with the operator $\partial/\partial t + £_{\bar{V}}$.

Exercise 5.22

The shear of a velocity field $\bar{V}$ is defined in Cartesian coordinates by the equation

$$\sigma_{ij} = V_{i,j} + V_{j,i} - \tfrac{1}{3}\delta_{ij}\theta, \tag{5.83}$$

where θ is the *expansion*

$$\theta = \nabla \cdot \bar{V}. \tag{5.84}$$

Show that in an arbitrary coordinate system

$$\theta = \tfrac{1}{2} g^{ij} £_{\bar{V}} g_{ij}, \tag{5.85}$$

$$\sigma_{ij} = £_{\bar{V}} g_{ij} - \tfrac{1}{3}\theta g_{ij}. \tag{5.86}$$

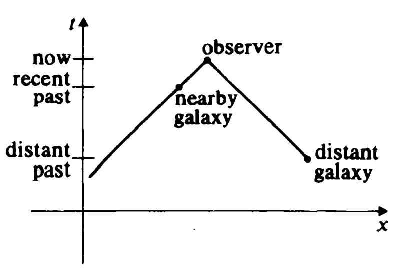
E Cosmology

5.19 The cosmological principle

Most physicists are aware that Einstein's theory of general relativity has given modern physics a consistent and fruitful framework in which to study cosmology, the large-scale structure of our universe. Most are also aware that, at least at the simplest level, there are only three basic cosmological models: the 'closed', 'flat', and 'open' universes. What is probably less well known is that this simplicity of having only three models is not at all a prediction or consequence of Einstein's equations. Rather, it is simply a consequence of assuming that the universe is homogeneous and isotropic in its large-scale properties. (Homogeneity and isotropy will be defined precisely below.) General relativity, like all the fundamental theories of physics, is a dynamical theory: given initial conditions, it will predict their future evolution and past history. The uniformity of the universe is part of the initial conditions we put in to construct the simplest models. The important contribution of general relativity is that it permits us to choose the *geometry* of space – its metric tensor field – as a part of the initial conditions. This is not possible in Newtonian gravity, of course. Once we decide to choose the most uniform initial conditions, it is differential geometry that tells us that only three metric tensor fields are possible. Our aim in the next few sections is to find these metrics. We shall use the mathematics of symmetry and invariance developed in chapter 3, but we will not need to know anything about general relativity nor even about Riemannian geometry.

We begin with the physical problem: the universe. On a small scale the universe is certainly lumpy. On nearly any length scale from the nuclear (10^{-15} m) to the interstellar (10^{17} m), our world is characterized by clumping of matter into small regions with sharp demarcations between different kinds of matter or between matter and the vacuum. The stars themselves group into more or less isolated galaxies, galaxies congregate into clusters of several tens to thousands, and even clusters may associate in loose superclusters. But modern astronomy can see well beyond the supercluster length scale, and we find that in all directions the tendency is for greater and greater homogeneity in the properties of the universe when they are averaged over larger and larger length scales. Since it is these large-scale averaged properties (particularly the mean density and

Fig. 5.4. A slice of spacetime showing all the events labelled by coordinates t (time) and x, with $y = z = 0$. Because electromagnetic radiation travels at a finite speed, distant objects are seen at an earlier time in their own histories than nearby objects.



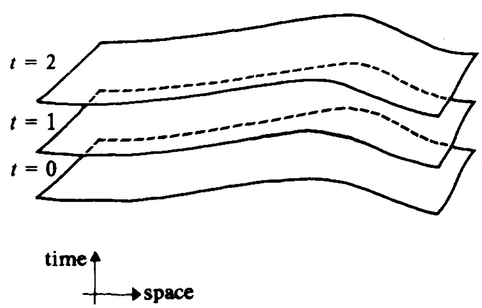
velocity) that are important for the dynamics of the universe, the cosmologist would like to incorporate this homogeneity into at least the simplest models. But what does homogeneity really mean? After all, in a dynamical universe, the more distant regions should look different from those nearby if only because they are seen at an earlier time in their history, as illustrated in figure (5.4). Indeed this is the case: the number of quasars, for instance, is much higher in distant regions than locally. The homogeneity one 'observes' is really an extrapolation to the present time of the condition of distant regions. Yet in relativity even 'the present time' is not an absolute concept. We cannot give a full discussion of these problems here, but we can say how they are resolved.

The basic idea is to split spacetime up into a family of three-dimensional spacelike submanifolds filling it up (a foliation). These are called hypersurfaces of constant time (see figure 5.5). This really amounts just to a choice of time-coordinate. The metric tensor g of spacetime has, like any $\binom{0}{m}$ tensor, a natural restriction to each hypersurface, and the hypersurface is space-like if g is positive-definite on all vectors tangent to it. The 'uniformity' of the cosmology depends on the Killing vectors or isometries of these hypersurfaces.

Let G be the Lie group of isometries of some manifold S with metric tensor field g. The Lie algebra of G is that of the Killing vector fields of g. Elements of G are mappings of S onto itself (diffeomorphisms). The action of G on S is said to be *transitive* on S if, for any two points P and Q of S, there is some element g of G for which $g(P) = Q$, i.e. which maps P to Q. The manifold S is said to be *homogeneous* if its isometry group acts transitively on it (see figure 5.6). What this means is just that the geometry is the same *everywhere* in S.

Suppose there are elements of G which leave some point P of S fixed. Then the product of any two also leaves P fixed, and since the identity e is one of them, they form a subgroup H_P of G called the *isotropy group* of P. These are, of course, the familiar rotations about an axis through P. The isotropy group of

Fig. 5.5. Slicing spacetime into spaces of constant time t.



P keeps P fixed and therefore maps any curve through P to another curve through P (see figure 5.7). It consequently induces a map of tangent vectors at P to others at P: a map $T_P \to T_P$. This group of mappings is the linear isotropy group of P. (Recall the similar discussion of the adjoint representation of a Lie group, §3.17.) A manifold S of dimension m is said to be *isotropic about P* if its isotropy group H_P is just $SO(m)$, the group of rotations about arbitrary axes through P. If S is isotropic about every point P it is said to be *isotropic*.

A cosmological model M is said to be a *homogeneous cosmology* if it has a foliation of space-like hypersurfaces, each of which is homogeneous; and similarly for an isotropic cosmology. As discussed above, the evidence is strong that our universe is homogeneous, at least on large scales in our observable neighborhood. We also see no systematic variations in its structure in different directions in the sky. This suggests the universe is isotropic about us. But modern science does not like to assume that we live in a particularly favorable location in the universe. This is often elevated to the status of a principle, variously known as the *cosmological principle*, the Copernican principle, or the principle of mediocrity: the properties of the universe we see near us would be seen, on average, by any observer anywhere else in the universe. This principle enables cosmologists, in the absence of information to the contrary, to extend our local

Fig. 5.6. Some neighborhood U of P is mapped by g onto a neighborhood V of $Q = g(P)$ isometrically: there is no difference in the geometry near P from that near Q.

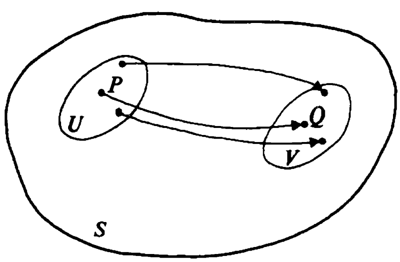
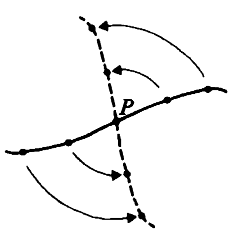


Fig. 5.7. The isotropy group of P maps $T_P \to T_P$ by mapping curves through P to other curves.



homogeneity and isotropy to the whole universe. This is not *necessary*, of course, and much current research is devoted to exploring inhomogeneous and/or anisotropic cosmologies. But the three basic models are the only three which have homogeneous, isotropic three-spaces. This is what we shall now prove.

Exercise 5.23

As we know from §3.9, the Killing vectors of the sphere S^2 are the vectors $\bar{l}_x, \bar{l}_y, \bar{l}_z$. These form a basis for the Lie algebra of the group of isometries of S^2, $SO(3)$. Prove that S^2 is a homogeneous and isotropic manifold.

5.20 Lie algebra of maximal symmetry

We shall begin by studying the Killing vector fields of a three-dimensional manifold S. If $\bar{\xi}$ is a Killing vector, its components in any coordinate system satisfy the equations

$$(\pounds_{\bar{\xi}}\mathsf{g})_{ij} = \xi^k g_{ij,k} + \xi^k{}_{,i} g_{kj} + \xi^k{}_{,j} g_{ik} = 0. \tag{5.87}$$

It will be more convenient to use the components of the one-form $\mathsf{g}(\bar{\xi}, \ )$,

$$\xi_k = g_{kl}\xi^l. \tag{5.88}$$

These satisfy the equivalent equations

$$\xi_{i,j} + \xi_{j,i} - 2\xi_l \Gamma^l{}_{ij} = 0, \tag{5.89}$$

with the definition

$$\Gamma^l{}_{ij} = \tfrac{1}{2}g^{lm}(g_{mi,j} + g_{mj,i} - g_{ij,m}). \tag{5.90}$$

(The definition of $\Gamma^k{}_{ij}$, including its factor of $\frac{1}{2}$, is conventional and would make more sense after a reading of chapter 6. For us equation (5.90) simply defines a convenient shorthand notation.)

Equation (5.89) is symmetric under exchange of i and j, so it represents in n dimensions $\frac{1}{2}n(n+1)$ independent differential equations, six for $n = 3$. Since there are only three components of $\bar{\xi}$ to solve for, the system is overdetermined: a general metric tensor g has *no* Killing vectors. Our object is to find what form g must take in order that it allow the maximum number of Killing vectors. To see what this maximum number is, we differentiate (5.89) to get

$$\xi_{i,jk} + \xi_{j,ik} = 2(\xi_l \Gamma^l{}_{ij})_{,k}. \tag{5.91}$$

By adding (5.91) to itself with the index permutation $(i \to k, j \to i, k \to j)$ and subtracting the permutation $(i \to j, j \to k, k \to i)$ we arrive at the equation

$$\xi_{i,jk} = H_{ijk}{}^l \xi_l + K_{ijk}{}^{lm} \xi_{l,m}, \tag{5.92}$$

where $H_{ijk}{}^{l}$ is a complicated function of g_{ij} and its first and second derivatives, and $K_{ijk}{}^{lm}$ similarly depends on g_{ij} and its first derivatives. The key point about (5.92) is that if we know ξ_i and $\xi_{i,j}$ at any point P and if we know g_{ij} everywhere, then we can determine $\xi_{i,jk}$ at P from (5.92), and similarly all its higher derivatives at P by successively differentiating (5.92). On an analytic manifold (which we shall assume) this suffices to determine the vector field $\bar{\xi}$ everywhere. Moreover, we know that ξ_i at P determines the *symmetric part* of $\xi_{i,j}$ at P by equation (5.89). If follows that every Killing vector field on S is determined completely by giving the values of

$$\eta_i \equiv \xi_i(P) \text{ and } A_{ij} \equiv \xi_{[i,j]}(P) \tag{5.93}$$

at *any* point P of S. It is important that a choice of $\{\eta_i, A_{ij}\}$ at P does not necessarily determine a Killing vector, because it may happen that (5.92) has no solutions: its right-hand side may not be symmetric under exchange of j and k. But the argument does show that there cannot be more Killing vectors than the number of independent choices of $\{\eta_i, A_{ij}\}$, which in m dimensions is

$$m + \tfrac{1}{2}m(m-1) = \tfrac{1}{2}m(m+1), \tag{5.94}$$

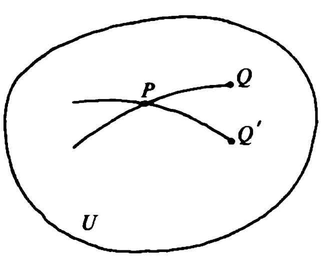
by virtue of (5.93). A manifold is said to be *maximally symmetric* if it has the maximum number of Killing vector fields.

It is easy to show that a maximally symmetric connected manifold S is homogeneous. At any point P we can choose a Killing vector field having any tangent at P. The one-parameter subgroups associated with these Killing vectors can therefore map P to any point Q in some neighborhood U of P (see figure 5.8). By a succession of such maps we can clearly map P to any point in S whatever. It follows that the isometry group maps P to any point, and S is homogeneous.

Next we take a look at the isotropy group of P. Such transformations leave P fixed, so the associated Killing vector fields vanish at P. The Lie bracket of *any* two Killing fields $\bar{V}$ and $\bar{W}$ is

$$[\bar{V}, \bar{W}]^i = V^i{}_{,j}W^j - W^i{}_{,j}V^j,$$

Fig. 5.8. By choosing the appropriate one-parameter subgroup of the isometry group one can map P to any point Q or Q' in a neighborhood U.



or

$$[\bar{V}, \bar{W}]_i = V_{i,j}W^j - W_{i,j}V^j - g_{ik,j}(V^k W^j - W^k V^j). \quad (5.95)$$

If $\bar{V}$ and $\bar{W}$ both vanish at P, then so does $[\bar{V}, \bar{W}]$. But $[\bar{V}, \bar{W}]$ is a linear combination of Killing vector fields, so for it to vanish at P it must be a linear combination only of those fields which also vanish at P. So these fields form a Lie subalgebra, clearly the algebra of the isotropy group at P. The next exercise shows that the isotropy group is $SO(m)$ if S is space-like, i.e. that a maximally symmetric space-like manifold is isotropic.

Exercise 5.24

Choose at P the sort of coordinate system permitted by exercise 2.14, in which for a *space-like* manifold $g_{ij}(P) = \delta_{ij}$ and $g_{ij,k}(P) = 0$.

(a) Show that near P an isotropy Killing vector field is given by

$$V^i = A^i{}_j x^j + O(x^2), \quad (5.96)$$

where $A^i{}_j$ is an arbitrary antisymmetric matrix

$$A^i{}_j = -A^j{}_i. \quad (5.97)$$

(b) Let $\bar{W}$ be another isotropy Killing vector field,

$$W^i = B^i{}_j x^j + O(x^2),$$

and show that

$$[\bar{V}, \bar{W}]^i = [A, B]^i{}_j x^j + O(x^2), \quad (5.98)$$

where $[A, B]^i{}_j$ denotes the elements of the matrix commutator of $A^i{}_j$ and $B^i{}_j$. This shows that the Lie algebra of the isotropy group is the same as the Lie algebra of $SO(m)$.

(c) Argue from this that the isotropy group of P is $SO(m)$.

(d) Show that if $\mathsf{g}|$ is *not* positive-definite (or negative-definite) then the isotropy group is not $SO(m)$. In particular show that the isotropy group of a point P in four-dimensional Minkowski space is the Lorentz group $L(4)$.

5.21 The metric of a spherically symmetric three-space

Now we restrict our attention to space-like three-manifolds. The isotropy group is $SO(3)$ and we say the manifold is spherically symmetric about any point. In this section we construct a convenient coordinate system for the rest of our calculation. We know that the Killing vectors of $SO(3)$ define spheres S^2 by their integral curves. Since every point is on one such sphere, they must foliate the manifold S. We will adopt spherical coordinates, with the usual θ and ϕ on each sphere and a third 'radial' coordinate labelling spheres. There is a particularly convenient choice for the radial coordinate. The metric of S induces a metric tensor on each sphere, which in turn defines a volume two-form and a total area (integral of the volume two-form). We *define* the radial coordinate r of a sphere by the equation

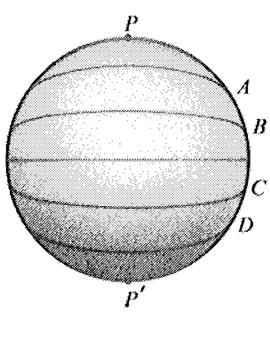
$$\text{area} = 4\pi r^2, \qquad r = (\text{area}/4\pi)^{1/2}. \tag{5.99}$$

This intrinsically defined coordinate need not be monotonically increasing everywhere, as figure 5.9 shows. But at least in some neighborhood of P it is guaranteed to be good by the local flatness theorem, exercise 2.14. (It is singular at $r = 0$, of course, but we know how to handle that.)

In addition to the radial coordinate we have to define θ and ϕ more precisely. We have placed θ and ϕ on each sphere but we have not said how the pole $\theta = 0$ of one sphere is related to that of another. That is, we are free to slide the coordinates of a sphere around as we move from one to another. We fix the pole in the following manner. At every point Q there is a vector $\bar{n}$ orthogonal to the sphere at that point ($\mathsf{g}(\bar{n}, \bar{V}) = 0$ for any $\bar{V}$ in $T_Q(S^2)$), normalized to unity ($\mathsf{g}(\bar{n}, \bar{n}) = 1$), and pointing away from P (which is well defined near P and extends to all of S by continuity). This vector field is called the unit normal vector field, and is C^∞ except at P. Choose the pole of any particular S^2 arbitrarily and then fix the poles of all the others by demanding they lie on the integral curve of $\bar{n}$ through the original pole. This is illustrated in figure 5.10. This clearly will imply that *any* integral curve of $\bar{n}$ is a curve of constant θ and ϕ, or in other words a coordinate line of the radial coordinate. Since $\partial/\partial\theta$ and $\partial/\partial\phi$ are tangent to the spheres this construction implies

$$g_{r\theta} = \mathsf{g}(\partial/\partial r, \partial/\partial\theta) = 0, \tag{5.100a}$$

Fig. 5.9. A radial coordinate labelling circles on a sphere, defined as the circumference ÷ 2π. This is the two-dimensional analogue of the situation described in the text. The radial coordinate increases away from P at first (say from A to B) but begins decreasing (from C to D) and becomes zero at P'.



$$g_{r\phi} = \mathsf{g}|(\partial/\partial r, \partial/\partial\phi) = 0. \quad (5.100b)$$

Moreover, on each sphere the metric is that of the unit sphere times r^2, the appropriate factor to make the area be $4\pi r^2$:

$$g_{\theta\theta} = r^2, \quad g_{\theta\phi} = 0, \quad g_{\phi\phi} = r^2 \sin^2\theta. \quad (5.100c)$$

We therefore have only one unknown metric component, g_{rr}.

Exercise 5.25

(a) Define the radial distance from P to a sphere with coordinate r to be the integral

$$\int_0^r (g_{rr})^{1/2}\,\mathrm{d}r \quad (5.101)$$

along a line $\theta = \text{const}$, $\phi = \text{const}$. Argue that g_{rr} must be independent of θ and ϕ.

(b) Show from exercise 2.14 that as one approaches P,

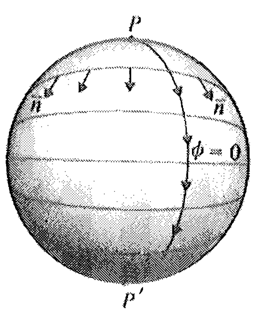
$$\lim_{r\to 0} g_{rr} = 1. \quad (5.102)$$

By exercise 5.25(a) we write $g_{rr} = f(r)$ and have the metric

$$(\mathsf{g}|) = \begin{pmatrix} f(r) & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \quad (5.103)$$

As we have used only the isotropy group of P to get this, we should not expect to be able to determine $f(r)$. For that we must use the rest of the isometries of S.

Fig. 5.10. Establishing the pole of each circle of constant r in figure 5.9 by requiring them all to lie on a single integral curve of the unit normal field $\bar{n}$.



5.22 Construction of the six Killing vectors

There are a number of methods we could use to find the form of $f(r)$ that guarantees the homogeneity of S. The method we shall use is to construct all the Killing vector fields of S by using the vector spherical harmonics of §4.29.

Any vector field $\bar{V}$ on S can be written in the form

$$\bar{V} = \xi_{lm}(r)Y_{lm}\frac{\partial}{\partial r} + \eta_{lm}(r)\bar{Y}^{+}_{lm} + \zeta_{lm}(r)\bar{Y}^{-}_{lm}, \tag{5.104}$$

with an implied summation on l and m here and wherever they are repeated in the same term. We shall need the components of this equation. It is easy to deduce from equation (4.101) that

$$(\bar{Y}^{+}_{lm})^{\theta} = Y_{lm,\theta}; \quad (\bar{Y}^{+}_{lm})^{\phi} = \frac{1}{\sin^2\theta}Y_{lm,\phi}; \tag{5.105a}$$

$$(\bar{Y}^{-}_{lm})^{\theta} = \frac{1}{\sin\theta}Y_{lm,\phi}; \quad (\bar{Y}^{-}_{lm})^{\phi} = -\frac{1}{\sin\theta}Y_{lm,\theta}. \tag{5.105b}$$

It follows that

$$V^r = \xi_{lm}Y_{lm}, \tag{5.106a}$$

$$V^{\theta} = \eta_{lm}Y_{lm,\theta} + \zeta_{lm}Y_{lm,\phi}/\sin\theta, \tag{5.106b}$$

$$V^{\phi} = \eta_{lm}Y_{lm,\phi}/\sin^2\theta - \zeta_{lm}Y_{lm,\theta}/\sin\theta. \tag{5.106c}$$

These components have to satisfy Killing's equation

$$K_{ij} \equiv V^k g_{ij,k} + V^k{}_{,i}g_{kj} + V^k{}_{,j}g_{ik} = 0, \tag{5.107}$$

with g_{ij} from (5.103).

The three equations $\{K_{\theta\theta} = 0, K_{\theta\phi} = 0, K_{\phi\phi} = 0\}$ do not involve derivatives of ξ_{lm}, η_{lm}, or ζ_{lm}, so we shall tackle them first. First consider the combination (indices raised with (5.103))

$$0 = K^{\theta}{}_{\theta} + K^{\phi}{}_{\phi} = \frac{4}{r}\xi_{lm}Y_{lm} + 2\eta_{lm}L^2(Y_{lm}),$$

where L^2 is the operator defined by equation (3.33). Using (3.33) we get

$$[(2/r)\xi_{lm} - l(l+1)\eta_{lm}]Y_{lm} = 0.$$

By the linear independence of the spherical harmonics we have

$$\frac{2}{r}\xi_{lm} - l(l+1)\eta_{lm} = 0. \tag{5.108}$$

Next consider the combinations

$$0 = \tfrac{1}{2}(K^{\theta}{}_{\theta} - K^{\phi}{}_{\phi}) = F_{lm}\eta_{lm} + G_{lm}\zeta_{lm}, \tag{5.109a}$$

$$0 = -\frac{1}{r^2\sin\theta}K_{\theta\phi} = -G_{lm}\eta_{lm} + F_{lm}\zeta_{lm}, \tag{5.109b}$$

where F_{lm} and G_{lm} are abbreviations for the expressions

$$F_{lm} = Y_{lm,\theta\theta} - \cot\theta\, Y_{lm,\theta} - Y_{lm,\phi\phi}/\sin^2\theta,$$

$$G_{lm} = 2Y_{lm,\theta\phi}/\sin\theta - 2\cot\theta\, Y_{lm,\phi}/\sin\theta.$$

Equations (5.109) have the solution $\zeta_{lm} = \eta_{lm} = 0$ unless the determinant of their coefficients vanishes. But this is $(F_{lm})^2 + (G_{lm})^2$, so it vanishes only if both F_{lm} and G_{lm} vanish. It is easy to work out that this happens for $l = 0$ and $l = 1$ (any m) but not for $l \geqslant 2$. Moreover, it is obvious from (5.106) that $l = 0$ does not have a contribution from η or ζ (the fixed-point theorem for S^2 again!) so that we can conclude

$$\begin{aligned} &l = 1: \eta_{1m}, \zeta_{1m} \text{ arbitrary;} \\ &l \geqslant 2: \eta_{lm} = \zeta_{lm} = 0. \end{aligned} \tag{5.110}$$

Then (5.108) gives us

$$\begin{aligned} &l = 0: \ \xi_{00} = 0, \\ &l = 1: \ \xi_{1m} = r\eta_{1m}, \\ &l \geqslant 2: \ \xi_{lm} = 0. \end{aligned} \tag{5.111}$$

Now we turn to the other three equations in (5.107). The first is a scalar with respect to rotations:

$$0 = K_{rr} = (2f\xi_{lm,r} + f_{,r}\xi_{lm})Y_{lm},$$

which implies

$$f\xi_{lm,r} + \tfrac{1}{2}f_{,r}\xi_{lm} = 0. \tag{5.112}$$

The remaining two equations, $K_{r\theta} = K_{r\phi} = 0$, transform as a vector under rotations. The divergence of this vector (with respect to the volume of S^2) is

$$0 = (\sin\theta K_r{}^{\theta})_{,\theta} + (\sin\theta K_r{}^{\phi})_{,\phi} = \left(\eta_{lm,r} + \frac{1}{r^2}f\xi_{lm}\right)\sin\theta L^2(Y_{lm}),$$

which again implies (for $l > 0$)

$$\eta_{lm,r} + \frac{1}{r^2}f\xi_{lm} = 0. \tag{5.113}$$

The remaining equation can be taken to be the divergence of the dual of the vector in S^2,

$$0 = K_{r\theta,\phi} - K_{r\phi,\theta} = r^2\zeta_{lm,r}\sin\theta L^2(Y_{lm}),$$

which of course implies

$$\zeta_{lm,r} = 0. \tag{5.114}$$

We may conclude that $\{\zeta_{1m}, m = -1, 0, 1\}$ are three arbitrary constants, the only contribution from $\bar{Y}_{lm}$. The three equations (5.111) for the unknowns ξ_{1m}, η_{1m}, and f have the following solution in terms of the arbitrary constants K and V_m:

$$f = (1 - Kr^2)^{-1}, \tag{5.115}$$

$$\xi_{1m} = V_m(1 - Kr^2)^{1/2}, \tag{5.116}$$

$$\eta_{1m} = \frac{1}{r} V_m(1 - Kr^2)^{1/2}. \tag{5.117}$$

Exercise 5.26
Verify equations (5.105), (5.108), (5.109), (5.112), (5.113), (5.114), and (5.115–17).

Exercise 5.27
Show that the Killing vectors with $V_m = 0$ are those corresponding to the isotropy group of the origin $r = 0$.

Exercise 5.28
Show that the apparent singularity in η_{1m} as $r \to 0$ is a coordinate effect: the vector field is well-behaved at the origin.

Exercise 5.29
Set $K = 0$ in (5.115–17) and show that S is just E^3, Euclidean space. Find the constants V_m that define the Killing vectors $\{\partial/\partial x, \partial/\partial y, \partial/\partial z\}$, where the Cartesian coordinates are obtained from our polars in the usual way.

5.23 Open, closed, and flat universes

We now have a complete description of the geometry of the homogeneous and isotropic spaces of the cosmological model: they have the metric tensor

$$(g_{ij}) = \begin{pmatrix} (1 - Kr^2)^{-1} & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \tag{5.118}$$

It only remains to try to get a picture of this geometry. The following coordinate transformations are a help.

Exercise 5.30
Find a coordinate transformation from r to χ which produces the following metric components

for $K > 0$:

$$(g_{ij}) = \frac{1}{K} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin^2\chi & 0 \\ 0 & 0 & \sin^2\chi \sin^2\theta \end{pmatrix} \tag{5.119a}$$

for $K<0$:

$$(g_{ij}) = \frac{1}{|K|}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sinh^2\chi & 0 \\ 0 & 0 & \sinh^2\chi\sin^2\theta \end{pmatrix}. \qquad (5.119b)$$

This shows that the geometry really depends only on the sign of K. Its magnitude serves only as an overall scale factor.

In the case $K>0$, the sphere of radial coordinate χ has area $4\pi\sin^2\chi/K$, which increases away from $\chi=0$ to a maximum at $\chi=\pi/2$ and then decreases to zero at $\chi=\pi$. This is reminiscent of S^2 (figure 5.9). In fact, this is the metric of the sphere S^3 of radius $K^{-1/2}$. Because the space is finite, the universe is said to be *closed*.

Exercise 5.31

Find a coordinate transformation of E^4 from Cartesian coordinates $\{x^i\} = \{w, x, y, z\}$ to spherical coordinates $\{x^{i'}\} = \{r, \chi, \theta, \phi\}$ in which the metric $g_{ij} = \delta_{ij}$ has the components $g_{i'j'}$ given by (5.119a) when restricted to the sphere S^3, $w^2+x^2+y^2+z^2=K^{-1}$.

The case $K=0$ has been considered in exercise 5.29. It is the *flat* universe.

The case $K<0$ is the *open* universe, and it is the hardest to visualize. The surface area of a sphere of radial coordinate χ is $4\pi\sinh^2\chi/|K|$, and increases ever more rapidly with χ. This universe is unbounded.

Exercise 5.32

(a) By considering the relation between the areas of spheres $\chi=$ const and the distance of the sphere from the origin $\chi=0$, equation (5.101), prove that the metric (5.119b) is not the restriction of the Euclidean metric to *any* submanifold of any E^n.

(b) Find a submanifold of Minkowski space whose metric is that of (5.119b).

When Einstein's equations are supplied with initial data which are homogeneous and isotropic (and this includes not only the geometry but the matter variables as well), then the subsequent evolution of the universe maintains the symmetry. It follows that the only aspect of the geometry which can change with time is the scale factor K: the universe gets 'larger' or 'smaller' as time goes on. One must be careful, however, not to make coordinate-dependent statements. For the closed universe, whose total volume is finite, the change in K does cause a change in the total volume. But the flat and open universes are both infinite, so it is not meaningful to talk about their total volume. What general relativity tells us is that the coordinates of equation (5.119) are 'comoving': the local mean rest frame of the galaxies in any small region of the universe stays at constant $\{\chi, \theta, \phi\}$ as time evolves. It follows then that a change in K produces a change in the distance between galaxies, and this is what is meant by an expanding universe. In the 'standard model' of the universe, which assumes homogeneity and isotropy and a few other things, all three kinds of universe begin with zero 'volume' ($K = \infty$) and expand away from this 'big bang'. The closed universe expands to a maximum and recollapses, the flat universe expands at a rate which goes asymptotically to zero, and the open universe expands at a rate which goes asymptotically to a nonzero limit. All of these things are consequences of Einstein's equations. To understand these equations it is necessary to add one more level of structure to our manifolds: the affine connection. This is the subject of chapter 6.

5.24 Bibliography

A concise and well-written introduction to thermodynamics is E. Fermi, *Thermodynamics* (Dover, New York 1956). Caratheodory's theorem is discussed by S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (Dover, New York, 1958), and at a more advanced level in R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic Press, New York, 1968).

Our discussion of Hamiltonian mechanics follows the spirit of R. Abraham & J. E. Marsden, *Foundations of Mechanics*, 2nd edn. (Benjamin/Cummings, Reading, Mass., 1978). See also W. Thirring, *A Course in Mathematical Physics I: Classical Dynamical Systems* (Springer Verlag, Vienna, 1978) and V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer Verlag, Berlin, 1978). An introduction to the same ideas without the geometrical point of view may be found in H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950) or L. D. Landau & E. M. Lifshitz, *Mechanics* (Pergamon Press, London, 1959). The use of the canonical conserved quantities makes certain fluid instabilities easier to understand; see J. L. Friedman & B. F. Schutz, *Astrophys. J.* 221, 937–57 (1978), and 222, 281–96 (1978). A number of useful articles on Hamiltonian mechanics from a geometrical viewpoint may be found in *Topics in Nonlinear Dynamics – A Tribute to Sir Edward Bullard*, ed. S. Jorna (American Institute of Physics, 1978: A.I.P. Conference Proceeding no. 46). An interesting use of differential forms is the proof of the necessary and sufficient conditions that a set of dynamical equations possess a Hamiltonian (i.e. symplectic) structure, in R. M. Santilli, *Foundations of Theoretical*

Mechanics I – The Inverse Problem in Newtonian Mechanics (Springer, Berlin, 1978).

For an introduction to electromagnetic theory which includes a discussion of its relativistic version see J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1976). A discussion which extends our own is C. W. Misner, K. S. Thorne & J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973). Another is W. Thirring, *A Course in Mathematical Physics II: Classical Field Theory* (Springer Verlag, Vienna, 1979). An advanced discussion of relativistic wave equations is F. G. Friedlander, *The Wave Equation on a Curved Space–Time* (Cambridge University Press, 1976). Wheeler's 'charge without charge' is discussed in articles reprinted in J. A. Wheeler, *Geometrodynamics* (Academic Press, New York, 1962). Sorkin's nonorientable charge model is described in R. Sorkin, *J. Phys. A.* **12**, 403–21 (1979).

Introduction to fluid dynamics which include discussions of the Helmholtz theorem are *Fluid Mechanics* by L. D. Landau & E. M. Lifshitz (Pergamon Press, London, 1959) and *Hydrodynamics* by H. Lamb (Dover, New York, 1975). What we have called Ertel's theorem is really a special case of a more general result derived by H. Ertel, *Meteorologische Zeitschrift* **59**, 277 (1942). We touched on the properties of the manifold called 'Galilean spacetime', which is the arena for pre-relativistic physics. Discussions of its structure involve the concept of an affine connection, developed in the next chapter. See C. W. Misner, K. S. Thorne & J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973) for a lucid discussion, or R. Hermann, *Topics in General Relativity* (Math-Sci Press, Brookline, Mass., 1973) for a more technical treatment. The use of Lie derivatives in continuum mechanics is very fruitful in elasticity theory. Indeed, it is very difficult to formulate the general-relativistic theory of elasticity without such techniques. See the treatment by B. Carter & H. Quintana, *Proc. Roy. Soc. London*, **A331**, 57 (1972), or B. Carter, *Proc. Roy. Soc. London, A*, to be published.

Cosmology is treated in most textbooks on general relativity. Our approach draws elements from both S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), and Misner *et al.*, op. cit. For an easier introduction see M. Berry, *Principles of Cosmology and Gravitation* (Cambridge University Press, 1977). The astrophysical and observational sides of cosmology are dealt with in Weinberg and in P. J. E. Peebles, *Physical Cosmology* (Princeton University Press, 1971). Homogeneous but not necessarily isotropic cosmologies are developed using the techniques of group theory in M. P. Ryan & L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton University Press, 1975).