Chapter 1

Manifolds, Flows, Lie Groups and Lie Algebras

In geometrical theory of dynamical systems, fundamental notions and tools are manifolds, diffeomorphisms, flows, exterior algebras and Lie algebras.

1.1. Dynamical Systems

In mechanics, we deal with physical systems whose state at a time t is specified by the values of n real variables,

$$x^1, x^2, \dots, x^n,$$

and furthermore the system is such that its time evolution is *completely determined* by the values of the n variables. In other words, the rate of change of these variables, i.e. $\mathrm{d}x^1/\mathrm{d}t, \dots, \mathrm{d}x^n/\mathrm{d}t$, depends on the values of the variables themselves, so that the equations of motion can be expressed by means of n differential equations of the first order,

$$\frac{\mathrm{d}x^i}{\mathrm{d}t} = X^i(x^1, x^2, \dots, x^n), \quad (i = 1, 2, \dots, n). \tag{1.1}$$

A system of time evolution of variables, such as $(x^1(t), \dots, x^n(t))$ described by (1.1), is termed a *dynamical system* [Birk27]. A simplest example would be the rectlinear motion of a point mass m located at x under a restoring force $-kx$ of a spring:

$$\mathrm{d}x/\mathrm{d}t = y, \quad \mathrm{d}y/\mathrm{d}t = -kx,$$

where k is a spring constant. A system of N point masses under self-interaction governed by Newton's equations of motion is another example.

However, the notion of the dynamical system is more general, and not restricted to such Newtonian dynamical system.

The space where the n-tuple of real numbers $(x^1, \ldots, x^n)$ reside is called a n-dimensional manifold M^n which will be detailed in the following sections. The space is also called the *configuration space* of the system, while the *physical state* of the system is determined by the $2n$ variables: the coordinates $(x^1, \ldots, x^n)$ and the velocities $(\dot{x}^1, \ldots, \dot{x}^n)$ where $\dot{x}^i = \mathrm{d}x^i/\mathrm{d}t$. Such a system is said to have n *degrees of freedom*. It is of fundamental importance how the differential equations are determined from basic principles, and in fact this is the subject of the present monograph.

Study of dynamical systems may be said to have started with the work of Henri Poincaré at the turn of the 19th to 20th century. Existence of very complicated orbits was disclosed in the problem of interacting three celestial bodies. After Poincaré, Birkhoff studied an exceedingly complex structure of orbits arising when an integrable system is perturbed [Birk27; Ott93]. Later, the basic question of how prevalent integrability is, was given a mathematical answer by Kolmogorov (1954), Arnold (1963) and Moser (1973), which is now called the KAM theorem and regarded as a fundamental theorem of chaos in Hamiltonian systems (e.g. [Ott93]).

The present approach to the dynamical systems is based on a geometrical point of view.[1] The geometrical frameworks concerned here were founded earlier in the 19th century by Gauss, Riemann, Jacobi and others. However in the 20th century, stimulated by the success of the theory of general relativity, the gauge theory (a geometrical theory) has been developed in theoretical physics. It has now become possible to formulate a geometrical theory of dynamical systems, mainly due to the work of Arnold [Arn66].

1.2. Manifolds and Diffeomorphisms

A fundamental object in the theory of dynamical systems is a manifold. A *manifold* $\mathbf{M}^n$ is an n-dimensional space that is locally an n-dimensional euclidean space $\mathbb{R}^n$ in the sense described just below, but is not necessarily $\mathbb{R}^n$ itself.[2] A unit n-sphere S^n in $(n+1)$-dimensional euclidean space $\mathbb{R}^{n+1}$ is a typical example of the n-dimensional manifold M^n. Consider

[1] In this context, the following textbooks may be useful: [Fra97; AK98; AM78].

[2] The euclidean space $\mathbb{R}^n$ is endowed with a global coordinate system $(x^1, \ldots, x^n)$ and is basically an important manifold. Henceforth the lower case e is used as "euclidean" because of its frequent occurrence.

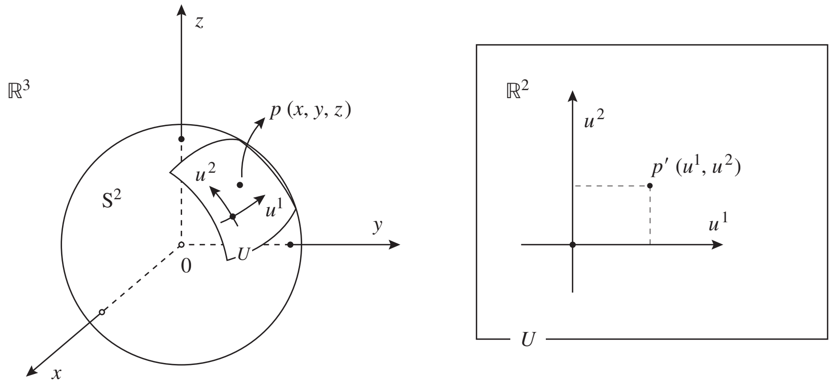

Fig. 1.1. Two-sphere S^2 and local coordinates.

a unit two-sphere $\mathbf{S}^2$ which is a two-dimensional object imbedded in three-dimensional space $\mathbf{R}^3$ (Fig. 1.1). Denoting a point in $\mathbb{R}^3$ by $p = (x, y, z)$, the two-sphere S^2 is defined by all points p satisfying $\|p\|^2 = x^2 + y^2 + z^2 = 1$, where $\|\cdot\|$ is the euclidean norm. The two-sphere S^2 is not a part of the euclidean space $\mathbb{R}^2$. However, an observer on S^2 would see that the immediate neighborhood is described by two coordinates and cannot be distinguished from a small domain of $\mathbb{R}^2$. A point p' in a patch U (an open subset of S^2) is represented by (u^1, u^2).

In general, an n-dimensional manifold M^n is a topological space (Appendix A.1), which is covered with a collection of open subsets $U_1, U_2, \ldots$ such that each point of M^n lies in at least one of them (Fig. 1.2). Using a map F_U, called a *homeomorphism* (Appendix A.2), each open

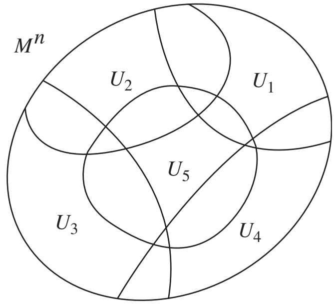

Fig. 1.2. Atlas.

subset U is in one-to-one correspondence with an open subset $F_U(U)$ of $\mathbb{R}^n$. Each pair (U, F_U), called a *chart*, defines a *coordinate patch* on M. To each point p $(\in U \subset M)$, we may assign the n coordinates of the point $F_U(p)$ in $\mathbb{R}^n$. For this reason, we call F_U a *coordinate map* with the jth component written as x_U^j. This is often described in the following way. On the patch U, a point p is represented by a *local coordinate*, $p = (x_p^1, \ldots, x_p^n)$. The whole system of charts is called an *atlas*.

The unit circle in the plane $\mathbb{R}^2$ is a manifold of one-sphere S^1. The S^1 has a local coordinate $\theta \in [0,1]$ (with the ends 0 and 1 identified) $\subset \mathbb{R}^1$. Consider a map by a complex function $f(\theta)$,

$$f(\theta) = e^{i2\pi\theta}, \quad f : \theta \in [0,1] \subset \mathbb{R}^1 \to p(x,y) \in S^1 \subset \mathbb{R}^2 \tag{1.2}$$

where $e^{i2\pi\theta} = x + iy$ $(i = \sqrt{-1}, x^2 + y^2 = 1)$. The map is one-to-one and onto if we identify the endpoints by $f(0) = f(1) \to (1,0) \in \mathbb{R}^2$ (Fig. 1.3). Choosing a patch (open subset) $U \subset S^1$, a homeomorphism map F_U is given by $f^{-1}(U)$.

It is readily seen that the unit circle $\mathbf{S}^1$ (a connected space[3]) is *covered* by the real axis $\mathbf{R}^1$ (another connected space) an infinite number of times by the map $f : \mathbb{R}^1 \to S^1$. Corresponding to an open subset $U \subset S^1$, the preimage $f^{-1}(U)$ consists of infinite number of disjoint open subsets $\{U_\alpha\}$ of $\mathbb{R}^1$, each U_α being diffeomorphic with U under $f : U_\alpha \to U$. It is said that the $\mathbb{R}^1$ is an *infinite-fold cover* of S^1.

Suppose that a patch U with its local coordinates $p = x = (x^1, \ldots, x^n)$ overlap with another patch V with local coordinates $p = y = (y^1, \ldots, y^n)$.

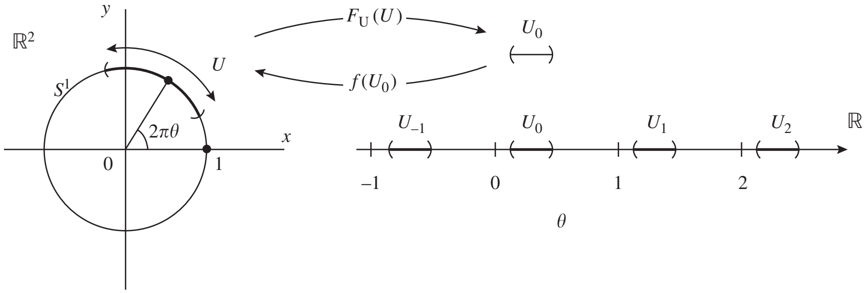

Fig. 1.3. Manifold S^1.

[3]A manifold M is said to be (*path-*)*connected* if any two points in M can be joined by a (piecewise smooth) curve belonging to M.

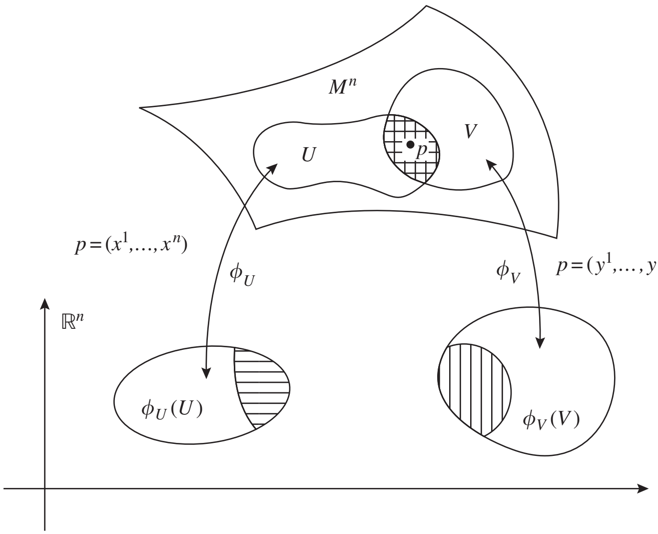

Fig. 1.4. Coordinate maps.

Then, a point p lying in the overlapping domain can be represented by both systems of x and y (Fig. 1.4). In particular, y^i is expressed in terms of x as

$$y^i = y^i(x^1, \ldots, x^n), \quad (i = 1, \ldots, n). \tag{1.3}$$

We require that these functions are smooth and differentiable, and that the Jacobian determinant

$$|J| = \frac{\partial(y)}{\partial(x)} = \frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)}. \tag{1.4}$$

does not vanish at any point $p \in U \cap V$ [Fla63, Ch. V].

Let $F : M^n \to W^r$ be a smooth map from a manifold M^n to another W^r. In local coordinates $x = (x^1, \ldots, x^n)$ in the neighborhood of the point $p \in M^n$ and $z = (z^1, \ldots, z^r)$ in the neighborhood of $F(p)$ on W^r, the map F is described by r functions $F^i(x), (i = 1, \ldots, r)$ of n variables, abbreviated to $z = F(x)$ or $z = z(x)$, where F^i are differentiable functions of x^j $(j = 1, \ldots, n)$.

When $n = r$, we say that the map F is a *diffeomorphism*, provided F is differentiable (thus continuous), one-to-one, onto, and in addition F^{-1} is differentiable (Fig. 1.5). Such an F is a *differentiable homeomorphism*. If the inverse F^{-1} does exist and the Jacobian determinant does not vanish, then the inverse function theorem would assure us that the

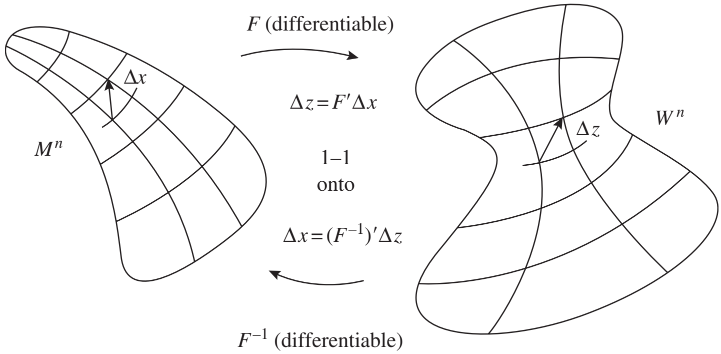

Fig. 1.5. Diffeomorphism.

inverse is differentiable. In the next section, the fluid flow is described to be a smooth sequence of diffeomorphisms of particle configuration (of infinite dimension).

1.3. Flows and Vector Fields

The vector field we are going to consider is not an object residing in a flat euclidean space and is different from a field of simple n-tuple of real numbers.

1.3.1. *A steady flow and its velocity field*

Given a steady flow[4] of a fluid in $\mathbb{R}^3$, one can construct a one-parameter family of maps: $\phi_t : \mathbb{R}^3 \to \mathbb{R}^3$, where ϕ_t takes a fluid particle located at p when $t = 0$ to the position $\phi_t(p)$ of the same particle at a later time $t > 0$ (Fig. 1.6). The family of maps are the so-called *Lagrangian* representation of motion of fluid particles. In terms of local coordinates, the jth coordinate of the particle is written as $x^j \circ \phi_t(p) = x^j_t(p)$, where "$x^j \circ$" denotes a *projection map* to take the jth component.

Associated with any such flow, we have a velocity at p,

$$v(p) := \frac{\mathrm{d}}{\mathrm{d}t}\phi_t(p)\bigg|_{t=0}.$$

In terms of the coordinates, we have $v^j(p) = (\mathrm{d}x^j_t(p)/\mathrm{d}t)|_{t=0}$. Taking a smooth function $f(x) = f(x^1, x^2, x^3)$, i.e. $f : \mathbb{R}^3 \to \mathbb{R}$ and differentiating

[4] *Steady* velocity field does not depend on time t by definition.

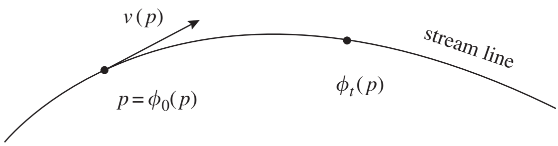

Fig. 1.6. Map ϕ_t.

$f(\phi_t(p))$ with respect to t, we have[5]

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\phi_t(p))\bigg|_{t=0} = \sum_j \frac{\mathrm{d}x_t^j(p)}{\mathrm{d}t} \frac{\partial f}{\partial x^j} = \sum_j v^j(p) \frac{\partial}{\partial x^j} f \tag{1.5}$$

$$=: X(p)f, \quad X(p) := \sum_j v^j(p) \frac{\partial}{\partial x^j}. \tag{1.6}$$

This is also written in the following way by bearing in mind that f is a map, $f : \mathbb{R}^3 \to \mathbb{R}$:

$$Xf = \frac{\mathrm{d}}{\mathrm{d}t} f(\phi_t) := \frac{\mathrm{d}}{\mathrm{d}t} f \circ \phi_t. \tag{1.7}$$

The differential operator X is written also as v by the reason described in the next subsection.

Conversely, to each vector field $v(x) = (v^j)$ in $\mathbb{R}^3$, one may associate a flow $\{\phi_t\}$ having v as its velocity field. The map $\phi_t(p)$ with t as an *integration parameter* can be found by solving the system of ordinary differential equations,

$$\frac{\mathrm{d}x^j}{\mathrm{d}t} = v^j(x^1(t), x^2(t), x^3(t))$$

with the initial condition, $x(0) = p$. Thus one finds an integral curve (called a *stream line*) in a neighborhood of $t = 0$, which is a one-parameter family of maps $\phi_t(p)$ for any $p \in \mathbb{R}^3$, called a flow *generated* by the vector field v, where $v = \dot{\phi}_t$ (Fig. 1.7). The map ϕ_t is a diffeomorphism, because $\phi_t(p)$ is differentiable, one-to-one, onto and ϕ_t^{-1} is differentiable, with respect to every point $p \in \mathbb{R}^3$.[6] This is assured in flows of a fluid by its physical property that each fluid particle is a physical entity which keeps its identity

[5]We use the symbol := to define the left side by the right side, and =: to define the right side by the left side.
[6]The flow $\{\phi_t\}$ is considered to be diffeomorphisms of Sobolev class H^s in Chapter 8 ($s > n/2 + 1$ in $\mathbb{R}^n$, Appendix F).

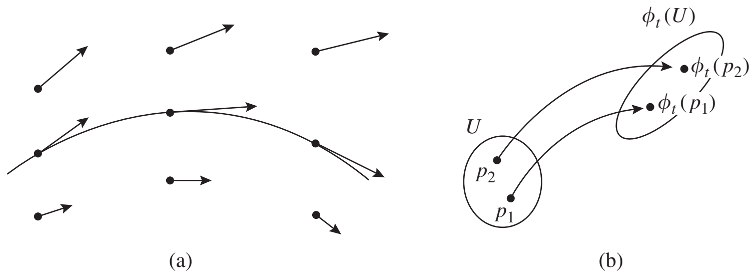

Fig. 1.7. (a) An integral curve and (b) a flow ϕ_t.

during the motion, as long as two particles do not come to occupying an identical point simultaneously.[7]

Remark. Continuous distribution of fluid particles in a three-dimensional euclidean space has infinite degrees of freedom. Therefore, the velocity field of all the particles as a whole is regarded to be of infinite dimensions. In this context, the set of diffeomorphisms ϕ_t forms an infinite dimensional manifold $D^{(\infty)}$ and a point $\eta = \phi_t \in D^{(\infty)}$ represents a configuration (as a whole) of all particles composing the fluid at a given time t.

1.3.2. *Tangent vector and differential operator*

The vector fields we are going to consider on M^n are not an object residing in a flat euclidean space. We need a sophisticated means to represent vectors which are different from a simple n-tuple of real numbers. In general, on a manifold M^n, one can define a vector v tangent to the parameterized curve ϕ_t at any point x on the curve. We motivate the definition of vector as follows.

A *flow* $\phi_t(p) = (x_t^j(p))$ on an n-dimensional manifold M^n is described by the system of ordinary differential equations,

$$\frac{\mathrm{d}x_t^j}{\mathrm{d}t} = v^j(x_t^1, \ldots, x_t^n), \quad (j = 1, \ldots, n), \tag{1.8}$$

with the initial condition, $\phi_0 = p$. The one-to-one correspondence between the tangent vector $v = (v^j)$ to M^n at x and the first order differential

[7]The present formulation is relevant to the time before a spontaneous formation of singularity (if any).

operator $\sum_j v^j(x)\partial/\partial x^j$, mediated by (1.8) and the n-dimensional version of (1.6), implies the following representation,

$$v(x) := \sum_j v^j(x)\frac{\partial}{\partial x^j}, \tag{1.9}$$

which defines the *vector field* $v(x)$ as a differential operator $v^j(x)\partial/\partial x^j$.

In fact, with a local coordinate patch (U, x_U) in the neighborhood of a point p, a curve will be described by n differentiable functions $(x_U^1(t), \ldots, x_U^n(t))$. The tangent vector at p is described by $v_U = (\dot{x}_U^1(0), \ldots, \dot{x}_U^n(0))$ where $\dot{x}(0) = \mathrm{d}x/\mathrm{d}t|_{t=0}$. If p also lies in the coordinate patch (V, x_V), then the same tangent vector is described by another n-tuple $v_V = (\dot{x}_V^1(0), \ldots, \dot{x}_V^n(0))$. In terms of the transformation function (1.3) on the overlapping domain which is now represented by $x_V^i = x_V^i(x_U^j)$, the two sets of tangent vectors are related by the chain rule,

$$v_V^i = \left.\frac{\mathrm{d}x_V^i}{\mathrm{d}t}\right|_{t=0} = \sum_j \left(\frac{\partial x_V^i}{\partial x_U^j}\right)\left.\frac{\mathrm{d}x_U^j}{\mathrm{d}t}\right|_{t=0} = \sum_j \left(\frac{\partial x_V^i}{\partial x_U^j}\right) v_U^j. \tag{1.10}$$

This suggests a transformation law of a tangent vector. Owing to this transformation, the definition (1.9) of a vector v is frame-independent, i.e. independent of local coordinate basis. In fact, by the transformation $x_V^i = x_V^i(x_U^j)$, we obtain

$$\sum_j v_U^j \frac{\partial}{\partial x_U^j} = \sum_j v_U^j(x) \sum_i \left(\frac{\partial x_V^i}{\partial x_U^j}\right)\frac{\partial}{\partial x_V^i} = \sum_i v_V^i \frac{\partial}{\partial x_V^i}. \tag{1.11}$$

It is not difficult to see that the properties of the linear vector space are satisfied by the representation (1.9).[8] Usually, in the differential geometry, no distinction is made between a vector and its associated differential operator. The vector $v(x)$ thus defined at a point $x \in M^n$ is called a *tangent* vector.

1.3.3. *Tangent space*

Each one of the n operators $\partial/\partial x^\alpha$ $(\alpha = 1, \ldots, n)$ defines a vector. The αth vector $\partial/\partial x^\alpha$ ($v^\alpha = 1$ and $v^i = 0$ for $i \neq \alpha$) is the tangent vector to the αth coordinate curve parameterized by x^α. This curve is described by

[8] It is evident from (1.9) that the sum of two vectors at a point is again a vector at that point, and that the product of a vector by a real number is a vector.

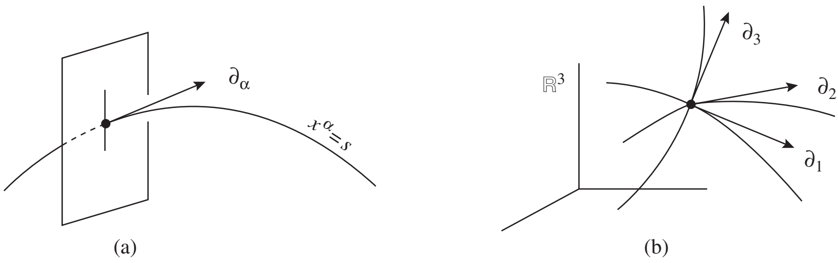

Fig. 1.8. (a) αth coordinate curve, (b) coordinate basis in $\mathbb{R}^3$.

$x^\alpha = s$ and $x^i = \text{const}$ for $i \neq \alpha$. Then the tangent vector $\partial/\partial x^\alpha$ for the αth curve has components $\mathrm{d}x^\alpha/\mathrm{d}s = 1$ and $\mathrm{d}x^i/\mathrm{d}s = 0$ for $i \neq \alpha$ (Fig. 1.8(a)). The n vectors $\partial/\partial x^1, \ldots, \partial/\partial x^n$ form a basis of a vector space, and this base is called a *coordinate* basis (Fig. 1.8(b)). The basis vector $\partial/\partial x^\alpha$ is simply written as ∂_α. A tangent vector X is written in general as[9]

$$X = X^j \partial_j, \quad \text{or} \quad X_x = X^j(x)\partial_j.$$

If $\boldsymbol{r} = (r^1, \ldots, r^N)$ is a position vector in the euclidean space $\mathbb{R}^N$ and M^n is a submanifold of $\mathbb{R}^N$: $M^n \subset \mathbb{R}^N$ $(n \leq N)$, the vector $\partial/\partial x^\alpha$ is understood as $\partial_\alpha \equiv \partial/\partial x^\alpha = \partial \boldsymbol{r}/\partial x^\alpha = (\partial/\partial x^\alpha)\ (r^1, \ldots, r^N)$, where $r^i = r^i(x^1, \ldots, x^n)$.[10]

The *tangent space* is defined by a vector space consisting of all tangent vectors to M^n at x and is written as $\mathbf{T}_x\mathbf{M}^n$.[11] When the coefficients X^j are smooth functions $X^j(x)$ for $x \in M^n$, the $X(x)$ is called a *vector field.*

1.3.4. *Time-dependent (unsteady) velocity field*

In most dynamical systems, a parameter t called the *time* plays a special role, and the tangent vector $v = (v^j)$ is called the *velocity.* A velocity field is said to be *time-dependent*, or *unsteady*, when v^j depends on t (an integration parameter) as well as space coordinates (Fig. 1.9). In the unsteady problem, an additional coordinate x^0 is introduced, and the n equations of (1.8) for

[9]The summation convention is used hereafter, i.e. the summation with respect to j is understood for the pair of double indices like j without the summation symbol $\sum$.
[10]The parameters $(x^1, \ldots, x^n)$ form a *curvilinear* coordinate system.
[11]It is useful in later sections to keep in mind that the tangent space T_xM^n is the usual n-dimensional *affine* subspace of $\mathbb{R}^N$.

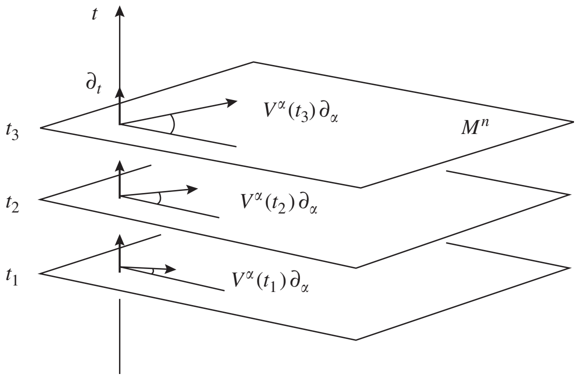

Fig. 1.9. Time-dependent velocity field.

$v^j \in \mathbb{R}^n$ are replaced by the following $(n+1)$ equations,

$$\frac{\mathrm{d}x^j}{\mathrm{d}t} = v^j(x^0(t), x^1(t), \ldots, x^n(t)), \quad \text{with } v^0 = 1, \tag{1.12}$$

for $j = 0, 1, \ldots, n$. It is readily seen that the newly added equation reduces to $x^0 = t$. Correspondingly, the tangent vector in the time-dependent case is written as, using the *tilde* symbol,

$$\tilde{v} := \tilde{v}^i \partial_i = v^0 \partial_0 + v^\alpha \partial_\alpha = \partial_t + v^\alpha \partial_\alpha, \tag{1.13}$$

where the index α denotes the spatial components $1, \ldots, n$.[12]

1.4. Dynamical Trajectory

A fundamental space of the theory of dynamical systems is a fiber bundle. How is the phase space of Hamiltonian associated with it?

1.4.1. *Fiber bundle (tangent bundle)*

In mechanics, a Lagrangian function L of a dynamical system of n degrees of freedom is usually defined in terms of generalized coordinates $q = (q^1, \ldots, q^n)$ and generalized velocities $\dot{q} = (\dot{q}^1, \ldots, \dot{q}^n)$ such as $L(q, \dot{q})$, while a Hamiltonian function is usually represented as $H(q, p)$, where

[12] *Nonvelocity* tangent vector such as the Jacobi vector $\tilde{J}$ is written simply as $\tilde{J} = J^\alpha \partial_\alpha$ (see the footnote of §8.3.4).

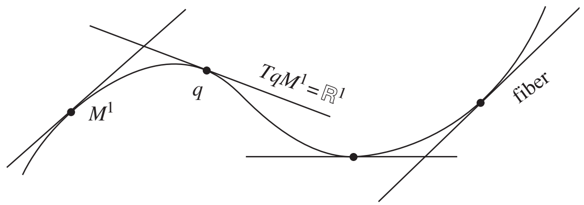

Fig. 1.10. A tangent bundle TM^1 for M^1 (a curve).

$p = (p_1, \ldots, p_n)$ are generalized momenta. Is there any significant difference between the pairs of independent variables?

Suppose that $q = (q^1, \ldots, q^n)$ is a point in an n-dimensional manifold U^n, which is a coordinate patch of a manifold M^n and a portion of $\mathbb{R}^n$, and that $\dot{q} = (\dot{q}^1, \ldots, \dot{q}^n)$ is a tangent vector to M^n at q. The pair $(q, \dot{q})$ is an element of a tangent bundle, $\mathbf{TM}^n$. Namely, a *tangent bundle* TM^n is defined as the collection of all tangent vectors at all points of M^n, called a *base manifold*.[13] A schematic diagram of a tangent bundle TM is drawn in Fig. 1.10 (see also Fig. 1.23 for a tangent bundle TS^1).

Associated with any bundle space TM, a projection map $\pi : TM \to M$ is defined by $\boldsymbol{\pi}(Q) = q$, where $Q \in TM, q \in M$. On the other hand, the inverse map $\pi^{-1}(q)$ represents all vectors v tangent to $M = M^n$ (base manifold) at q, i.e. a vector space $\mathbf{T}_q\mathbf{M} = \mathbb{R}^n$. It is called the *fiber* over q. In this regard, the tangent bundle is also called a *fiber bundle*, or a *vector bundle*[14] (Fig. 1.11). Since $\pi^{-1}(U^n)$ is topologically $U^n \otimes \mathbb{R}^n$, the tangent bundle is locally a product. However, this is not so in general (see [Fla63, Ch. 2; Sch80, Ch. 2; NS83, Ch. 7]).

1.4.2. *Lagrangian and Hamiltonian*

The space of generalized coordinates $q = (q^1, \ldots, q^n)$ is called the *configuration space* in mechanics (also called a base space), whereas the space $(q, \dot{q})$ is called the tangent bundle, a *mathematical* term. The Lagrangian $L(q, \dot{q})$ is a function on the tangent bundle to M^n, namely $L : TM^n \to \mathbb{R}$.

[13] If a point of TM^n is represented globally as $(q, \dot{q})$, i.e. a global product bundle $q \otimes \dot{q}$, the tangent bundle is called a *trivial* bundle. Note that the first n coordinates $(q^1, \ldots, q^n)$ take their values in a portion $U^n \in \mathbb{R}^n$, whereas the second set $(\dot{q}^1, \ldots, \dot{q}^n)$ take any value in $\mathbb{R}^n$. Thus, the patch is of the form, $U^n \otimes \mathbb{R}^n$.

[14] A fiber is not necessarily a simple vector. It takes, for example, even an element of Lie algebra. See Chapter 9.

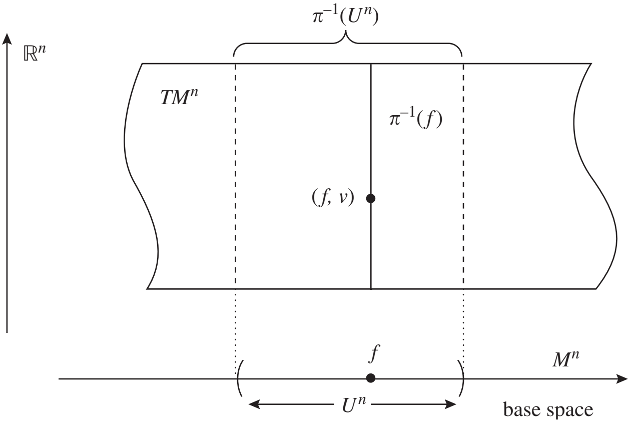

Fig. 1.11. A fiber bundle TM^n.

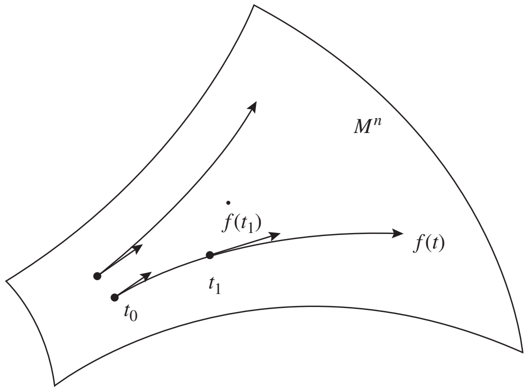

Fig. 1.12. Dynamical trajectories.

If we consider a specific trajectory $q(t)$ in the configuration space with t as the time parameter, then we have $\dot{q} = \mathrm{d}q/\mathrm{d}t$. Thus, the pair $(q, \dot{q})$ has a certain *geometrical* significance (Fig. 1.12).

Dynamical trajectory of the point $q(t)$ is determined by the following Lagrange's equation of motion (see §7.2.3):

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q} = 0. \tag{1.14}$$

The Hamiltonian function $H(q,p)$ is defined by

$$H(q,p) = \sum_i p_i \dot{q}^i - L(q,\dot{q}), \tag{1.15}$$

where p_i is an ith component of the generalized momentum defined by

$$p_i(q,\dot{q}) := \frac{\partial}{\partial \dot{q}^i} L(q,\dot{q}). \tag{1.16}$$

Change of variables from $(q,\dot{q})$ for the Lagrangian $L(q,\dot{q})$ to (q,p) for the Hamiltonian $H(q,p)$ has a certain significance more than a mere change of coordinates. Consider a coordinate transformation from q_U to q_V by $q_V = q_V(q_U)$. Correspondingly, the change of velocity, $\dot{q}_U \to \dot{q}_V$, is represented by

$$\dot{q}_V^i = \sum_k \frac{\partial q_V^i}{\partial q_U^k} \dot{q}_U^k. \tag{1.17}$$

On the other hand, the generalized momentum is transformed as follows,

$$\begin{aligned} (p_V)_i &= \frac{\partial}{\partial \dot{q}_V^i} L(q_U, \dot{q}_U) = \sum_k \frac{\partial \dot{q}_U^k}{\partial \dot{q}_V^i} \frac{\partial L}{\partial \dot{q}_U^k} \\ &= \sum_k \frac{\partial \dot{q}_U^k}{\partial \dot{q}_V^i} (p_U)_k = \sum_k \frac{\partial q_U^k}{\partial q_V^i} (p_U)_k, \end{aligned} \tag{1.18}$$

since $q_U = q_U(q_V)$ and therefore $\partial q_U^k / \partial \dot{q}_V^i = 0$ in the second equality, and (1.17) is used to obtain the last equality since $\partial \dot{q}_U^k / \partial \dot{q}_V^i = \partial q_U^i / \partial q_V^k$. Thus, it is found that the transformation matrix for p is the inverse of that of $\dot{q}$.

The expression (1.17) represents the transformation law of vectors and characterizes the tangent bundle, while the expression (1.18) characterizes the transformation law of covectors (see §1.5.2). The two transformation laws imply that the product $\sum_i p_i \dot{q}^i$ would be a scalar, an invariant under a coordinate transformation, since $\sum_i (p_V)_i \dot{q}_V^i = \sum_i (p_U)_i \dot{q}_U^i$ can be shown. A covector and a vector are associated with each other by means of a metric tensor (see §1.4.2).

1.4.3. *Legendre transformation*

Mathematically, the change $\dot{q} \to p$ is interpreted as a *Legendre transformation* (e.g. [Arn78, §14]). Consider a function $l(x)$ of a single variable x, where $l''(x) > 0$, i.e. $l(x)$ is *convex*. Let p be a given real number and define the function $h(x,p) = px - l(x)$. The function $h(x,p)$ has a maximum with respect to x at a point $x_*(p)$. The point x_* is determined uniquely by the

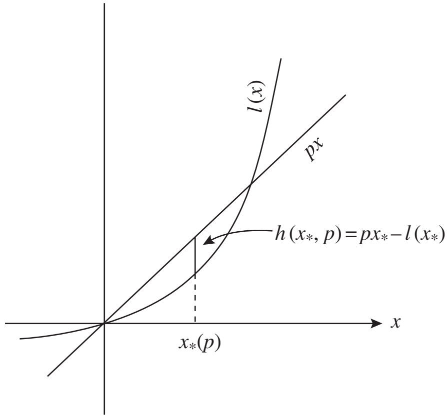

Fig. 1.13. Legendre transformation.

condition, $\partial h/\partial x = p - l'(x_*) = 0$, since $l'(x)$ is a monotonically increasing function by the convexity (Fig. 1.13). Thus, $p = l'(x_*)$. If we write $x = \dot{q}$, the relation $p = l'(x)$ is equivalent to (1.16) as far as the variable $\dot{q}$ is concerned.

By the Legendre transformation, the Lagrangian $L(q, \dot{q})$ on a vector space is transformed to the Hamiltonian $H(q, p)$ on the dual space, defined by (1.15) and (1.16). In mechanics, the space (q, p) is called the *phase space*. The equations of motion in the phase space are derived as follows:

$$\mathrm{d}H(q,p) = \sum_i \left(\frac{\partial H}{\partial q^i}\,\mathrm{d}q^i + \frac{\partial H}{\partial p_i}\,\mathrm{d}p_i \right). \tag{1.19}$$

On the other hand, taking the differential of the right-hand side of (1.15) and using (1.16), we obtain

$$\begin{aligned} \mathrm{d}\left(\sum_i p_i \dot{q}^i - L(q,\dot{q}) \right) &= \sum_i \left(p_i \mathrm{d}\dot{q}^i + \dot{q}^i \mathrm{d}p_i - \frac{\partial L}{\partial q^i}\,\mathrm{d}q^i - \frac{\partial L}{\partial \dot{q}^i}\,\mathrm{d}\dot{q}^i \right) \\ &= \sum_i \left(-\frac{\partial L}{\partial q^i}\,\mathrm{d}q^i + \dot{q}^i \mathrm{d}p_i \right). \end{aligned} \tag{1.20}$$

Equating the right sides of the above two equations, we obtain the following Hamilton's equations of motion,

$$\frac{\mathrm{d}q^i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i}, \quad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q^i}, \tag{1.21}$$

since $\mathrm{d}p_i/\mathrm{d}t = \partial L/\partial q^i$ by using (1.16) and Lagrange's equation of motion (1.14).

1.5. Differential and Inner Product

A basic tool of a dynamical system is a metric. How are vectors and covectors related to it?

1.5.1. *Covector (1-form)*

Differential $\mathrm{d}f$ of a function f on M^n is *defined* by $\mathrm{d}f[v] := vf$ and is regarded as a linear functional $T_xM^n \to \mathbb{R}$ for any vector $v \in E = T_xM^n$. In local coordinates, we have $v = v^j\partial_j$. Using (1.9), we obtain

$$\mathrm{d}f[v] = \mathrm{d}f[v^j\partial_j] = vf = \sum_j v^j(x)\frac{\partial f}{\partial x^j}. \tag{1.22}$$

This is a basis-independent definition (see (1.11)). The differential $\mathrm{d}f[v^j\partial_j]$ is linear with respect to the scalar coefficient v^j. In particular, if f is the coordinate function x^i, we obtain

$$\mathrm{d}x^i[v] = \mathrm{d}x^i[v^j\partial_j] = v^j\,\mathrm{d}x^i\left[\frac{\partial}{\partial x^j}\right] = v^j\frac{\partial x^i}{\partial x^j} = v^j\delta^i_j = v^i \tag{1.23}$$

by replacing f with x^i. Namely the operator $\mathrm{d}x^i$ reads off the ith component of any vector v (Fig. 1.14). It is seen that[15]

$$\mathrm{d}x^i[\partial_j] = \delta^i_j.$$

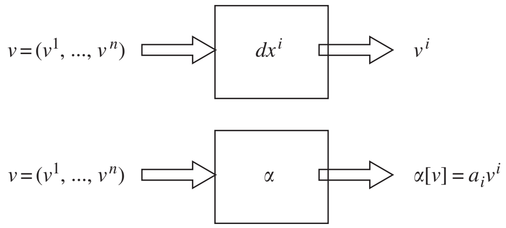

Fig. 1.14. 1-forms: $\mathrm{d}x^i$ and $\alpha = a_i\mathrm{d}x^i$.

[15]The symbols δ_{ij}, δ^{ij} and δ^i_j are identity tensors of rank 2, i.e. second order *covariant*, second order *contravariant* and *mixed* (first order covariant and first order contravariant) unit tensor, respectively (see §1.10).

Thus, the n functionals $\mathrm{d}x^i$ $(i = 1, \ldots, n)$ yield the dual bases corresponding to the coordinate bases $(\partial_1, \ldots, \partial_n)$ of a vector space $T_x M^n$, in the sense described below. The dual bases $(\mathrm{d}x^1, \ldots, \mathrm{d}x^n)$ form a dual space $(T_x M^n)^*$. The most general linear functional, $\alpha : T_x M^n \to \mathbb{R}$, is expressed in coordinates as

$$\alpha := a_1 \mathrm{d}x^1 + \cdots + a_n \mathrm{d}x^n. \tag{1.24}$$

The α is called a *covector*, or a *covariant vector*, or a differential *one-form* (1-form),[16] and is an element of the *cotangent* space $E^* = (T_x M^n)^*$. Corresponding to the covariant vector α, the vector v is also called a *contravariant vector*. Given a contravariant vector $v = v^j \partial_j$, the 1-form $\alpha \in E^*$ takes the value,

$$\alpha[v] = \sum_i a_i \, \mathrm{d}x^i [v^j \partial_j] = a_i v^i. \tag{1.25}$$

Correspondingly, a contravariant vector $v \in E$ can be considered as a linear functional on the covariant vectors with the definition of the same value as (1.25)[17]:

$$v[\alpha] \equiv \alpha[v] = a_i v^i. \tag{1.26}$$

When the coefficients a_i are smooth functions $a_i(x)$, the α is a 1-form *field* and an element of the *cotangent bundle* $(TM^n)^*$.

Appendix B describes exterior forms, products and differentials in some detail. A function $f(x)$ on $x \in M^n$ is a zero-form. Differential of a function $f(x)$ is a typical example of the covector (1-form):

$$\mathrm{d}f = \frac{\partial f}{\partial x^i} \mathrm{d}x^i = \partial_i f \, \mathrm{d}x^i, \quad \partial_i f = \frac{\partial f}{\partial x^i}, \tag{1.27}$$

where $\mathrm{d}x^i$ is a basis covector and $\partial f / \partial x^i$ is its component. This form holds in any manifold. In the next subsection, a vector grad f is defined as one corresponding to the covector $\mathrm{d}f$.

---

[16] The differential one-form is called also *Pfaff form*, and the equation $a_1 \mathrm{d}x^1 + \cdots + a_n \mathrm{d}x^n = 0$ is called *Pfaffian equation* on M^n.

[17] In Eqs. (1.25) and (1.26), the Einstein's *summation convention* is used, i.e. a summation is implied over a pair of double indices (i in the above cases) appearing in a lower (covariant) and an upper (contravariant) index in a single term, and is used henceforth.

1.5.2. *Inner (scalar) product*

Let the vector space $T_x M^n$ be endowed with an inner (scalar) product $\langle \cdot, \cdot \rangle$. For each pair of vectors $X, Y \in T_x M^n$, the inner product $\langle X, Y \rangle$ is a real number, and it is bilinear and symmetric with respect to X and Y. Furthermore, the $\langle X, Y \rangle$ is *nondegenerate* in the sense that

$$\langle X, Y \rangle = 0 \quad \text{for } {}^{\forall} Y \in T_x M^n, \;\; \text{only if } X = 0.$$

Writing $X = X^i \partial_i$ and $Y = Y^j \partial_j$, the inner product is given by

$$\langle X, Y \rangle := g_{ij} X^i Y^j, \tag{1.28}$$

where

$$g_{ij} := \langle \partial_i, \partial_j \rangle = g_{ji} \tag{1.29}$$

is the *metric tensor*. If it happens that the tensor is the unit matrix,

$$g_{ij} = \delta_{ij}, \quad \text{i.e.} \;\; g = (\delta_{ij}) = I, \tag{1.30}$$

we say that the metric tensor is the *euclidean metric*, where δ_{ij} is the Kronecker's delta: $\delta_{ij} = 1$ (if $i = j$), 0 (if $i \neq j$).

By definition, the inner product $\langle A, X \rangle$ is linear with respect to X when the vector A is fixed. Then the following α-operation on X, $\alpha[X] = \langle A, X \rangle$, is a linear functional: $\alpha = \langle A, \cdot \rangle$. In other words, to each vector $A = A^j \partial_j$, one may associate a covector α. By definition, $\alpha[X] = g_{ij} A^j X^i = (g_{ij} A^j) X^i$. On the other hand, for a covector of the form (1.24), one has $\alpha[X] = a_i \, \mathrm{d}x^i[X] = a_i X^i$, in terms of the basis $\mathrm{d}x^i$. Thus one obtains

$$a_i = g_{ij} A^j = g_{ji} A^j =: A_i, \tag{1.31}$$

which defines a covector A_i, and the component a_i is given by $g_{ij} A^j$ and written as A_i using the same letter A. The covector $\alpha = A_i \mathrm{d}x^i = (g_{ij} A^j) \mathrm{d}x^i$ is called the covariant version of the vector $A = A^j \partial_j$. In tensor analysis, Eq. (1.31) is understood as indicating that the upper index j is lowered by means of the metric tensor g_{ij}. In other words, *a covector A_i is obtained by lowering the upper index of a vector A^j by means of g_{ij}*. In summary, the inner product is represented as

$$\langle X, Y \rangle = g_{ij} X^i Y^j = X^i Y_i = X_j Y^j. \tag{1.32}$$

On the other hand, a vector A^j is obtained by raising the lower index of the covector A_i as

$$A^j = g^{ji} A_i, \tag{1.33}$$

which is equivalent to solving Eq. (1.31) to obtain A^j. This is verified by the property that the metric tensor $g = (g_{ij})$ is assumed nondegenerate, therefore the inverse matrix g^{-1} must exist and is symmetric. The inverse is written as $g^{-1} =: (g^{ji})$ in Eq. (1.33) using the same letter g. As an example, we obtain the expression of the vector grad f as

$$(\mathrm{grad}\ f)^j = g^{ji} \frac{\partial f}{\partial x^i}. \tag{1.34}$$

1.6. Mapping of Vectors and Covectors

Dynamical development is a smooth sequence of maps from one state to another with respect to a parameter "time". Here we consider general rules of mappings (transformations).

1.6.1. *Push-forward transformation*

Let $\phi : M^n \to V^r$ be a smooth map. In addition, let us define the differential of the map ϕ by $\boldsymbol{\phi_*} : T_x M^n \to T_y V^r$. In local coordinates, the map ϕ is represented by a function $F(x)$ as $y = \phi(x) = F(x)$, where $x \in M^n$ and $y \in V^r$. Let $p(t)$ be a curve on M^n with $p(0) = p$ and $\dot{p}(0) = X$ (a *tangent* vector), where $X \in T_p M^n$. The differential map ϕ_* at p is defined by

$$Y = \phi_* X (= F_* X) := \frac{\mathrm{d}}{\mathrm{d}t} (F(p(t))|_{t=0}. \tag{1.35}$$

This is called a *push-forward* transformation (Fig. 1.15) of the velocity vector X to the vector Y (the velocity vector of the image curve at $F(p)$).

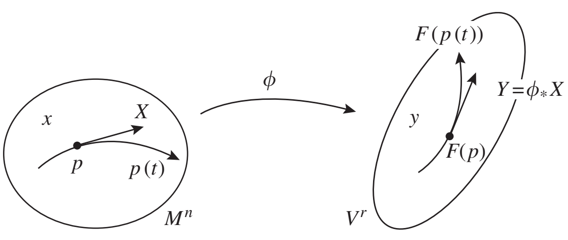

Fig. 1.15. Push-forward transformation ϕ by a function $y = F(x)$.

(a) Let us consider the case $n = r$. Suppose that the transformation is given by $x \mapsto y = (F^k(x))$ within the same reference frame ∂_k, and that the tangent vector $X = X^j \partial_j$ is mapped to $Y = Y^k \partial_k$. Then the components are transformed as (see §4.2.1)

$$Y^k = (\phi_* X)^k = \left(\frac{\partial F^k}{\partial x^j}\right) X^j. \tag{1.36}$$

(b) Next, consider a transformation between two basis vectors for $n = r$ again. The transformation ϕ_* applies to the basis vectors $\partial/\partial x^j$, and we have

$$Y = \phi_* X = \phi_* \left[X^j \frac{\partial}{\partial x^j}\right] = X^j \phi_* \left[\frac{\partial}{\partial x^j}\right] = X^j \frac{\partial y^k}{\partial x^j} \frac{\partial}{\partial y^k} = Y^k \frac{\partial}{\partial y^k}. \tag{1.37}$$

The components of Y are given by

$$Y^k = \frac{\partial y^k}{\partial x^j} X^j = J^k_j X^j, \quad J^k_j := \frac{\partial y^k}{\partial x^j}. \tag{1.38}$$

This is also written as $Y = JX$, where $J = (J^k_j)$.

In particular, setting $X^i = 1$ (for an integer i) and others as zero in (1.37), it is found that the bases $(\partial/\partial x^i)$ are transformed as

$$\phi_* \left[\frac{\partial}{\partial x^i}\right] = \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k}. \tag{1.39}$$

If we write this in the form,

$$\frac{\partial}{\partial y^k} = B^i_k \frac{\partial}{\partial x^i}, \tag{1.40}$$

the matrix $B^i_k = \partial x^i/\partial y^k$ is the inverse of J since

$$BJ = \frac{\partial x^i}{\partial y^k} \frac{\partial y^k}{\partial x^j} = \frac{\partial x^i}{\partial x^j} = \delta^i_j = I. \tag{1.41}$$

A physical example of the transformations (a) and (b) is seen in §4.2.1 for rotations of a rigid body. Equation (1.39) is also written as

$$\phi_* \left[\frac{\partial}{\partial x^i}\right] f = \frac{\partial y^k}{\partial x^i} \frac{\partial f}{\partial y^k} = \frac{\partial}{\partial x^i} f(\phi(x)) \equiv \frac{\partial}{\partial x^i} f \circ \phi(x). \tag{1.42}$$

Writing as $X = X^j(x)\partial/\partial x^j$,

$$\phi_* X[f] = X[f \circ \phi]. \tag{1.43}$$

(c) A manifold M^n is called a *submanifold* of a manifold V^r (where $n < r$) provided that there is a one-to-one smooth mapping $\phi : M^n \to V^r$ in which the matrix J has (maximal) rank n at each point. We refer to ϕ as an *imbedding* or an *injection*. This appears often when $V = \mathbb{R}^r$ so that we consider submanifolds of an euclidean space $\mathbb{R}^r$.

1.6.2. *Pull-back transformation*

Corresponding to the *push-forward* ϕ_*, one can define the *pull-back* ϕ^*, which is the linear transformation taking a covector at y back to a covector at x, i.e. $\phi^* : (T_y V)^* \to (T_x M)^*$. Suppose that a vector X at $x \in M$ is transformed to $Y = \phi_*(X)$ at $y = \phi(x) \in V$, then the pull-back ϕ^* of a covector α (one-form) is defined, using the push-forward $\phi_*(X)$, by

$$(\phi^*\alpha)[X] := \alpha[\phi_*(X)], \tag{1.44}$$

for any one-form $\alpha = A_i \mathrm{d}y^i$. This defines an invariance of the pull-back transformation. Namely, the value of the covector $\alpha = A_i \mathrm{d}y^i$ at the vector $Y = \phi_* X$ (in V) is equal to the value of the pull-back covector $\phi^*\alpha$ at the original vector X (in M).

Note that, owing to $\mathrm{d}x^i[\partial_j] = \delta^i_j$, one has

$$\alpha\left[\frac{\partial}{\partial y^k}\right] = A_i \mathrm{d}y^i \left[\frac{\partial}{\partial y^k}\right] = A_k. \tag{1.45}$$

Writing

$$\phi^*\alpha = a_i\, \mathrm{d}x^i, \tag{1.46}$$

one obtains $a_i = \phi^*\alpha[\partial/\partial x^i]$, and furthermore one can derive the following transformation of the components of covectors by using (1.39) and (1.45):

$$\begin{aligned} a_i = \phi^*\alpha\left[\frac{\partial}{\partial x^i}\right] &= \alpha\left[\phi_*\frac{\partial}{\partial x^i}\right] = \alpha\left[\frac{\partial y^k}{\partial x^i}\frac{\partial}{\partial y^k}\right] \\ &= \frac{\partial y^k}{\partial x^i}\alpha\left[\frac{\partial}{\partial y^k}\right] = A_k \frac{\partial y^k}{\partial x^i}. \end{aligned} \tag{1.47}$$

Thus, using J^k_i of (1.38), we have the transformation law,

$$a_i = A_k J^k_i. \tag{1.48}$$

Substituting the expression $A_k \mathrm{d}y^k$ for α in (1.46) and using (1.47), we have

$$\phi^*(A_k \mathrm{d}y^k) = A_k \frac{\partial y^k}{\partial x^j} \mathrm{d}x^j. \tag{1.49}$$

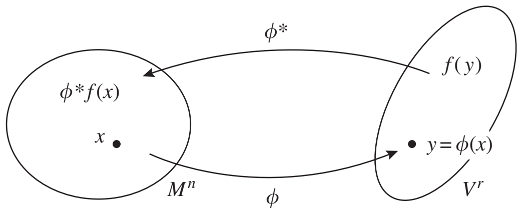

Fig. 1.16. Pull-back of a function $f(y)$ to $(\phi^* f)(x)$.

Setting A_i (only) $= 1$ (the other components being zero) as before (for an integer k), it is found that the bases $(\mathrm{d}y^i)$ are transformed as

$$\phi^*[\mathrm{d}y^i] = \frac{\partial y^i}{\partial x^j}\mathrm{d}x^j. \tag{1.50}$$

The *pull-back of a function* $f(y)$ (Fig. 1.16) is given by

$$(\phi^* f)(x) = f(\phi(x)),$$

where a scalar function $f(y)$ is a zero-form. If one sets $A_i = \partial f/\partial y^i$ in (1.44), Eq. (1.44) expresses invariance of the differential:

$$\phi^*(\mathrm{d}f)_y = \phi^*\left[\left(\frac{\partial f}{\partial y^i}\right)\mathrm{d}y^i\right] = \left(\frac{\partial f}{\partial y^i}\right)\left(\frac{\partial y^i}{\partial x^j}\right)\mathrm{d}x^j$$
$$= \left(\frac{\partial f}{\partial x^j}\right)\mathrm{d}x^j = (\mathrm{d}f)_x.$$

Based on this invariance, the general pull-back formula is defined for the integral of a form (covector) α over a curve σ as

$$\int_{\phi(\sigma)} \alpha = \int_\sigma \phi^*\alpha, \tag{1.51}$$

where $\phi : \sigma \subset M \to \phi(\sigma) \subset V$. Namely, the integral of a form α over the image $\phi(\sigma)$ is the integral of the pull-back $\phi^*\alpha$ over the original σ. See the next section 1.6.3 for $M = V$, and Appendix B.7 for an integral of a general form α.

1.6.3. *Coordinate transformation*

Change of coordinate frame can be regarded as a mapping $y = y(x) : x \in U^n \to y \in V^n$, where (U^n, x) and (V^n, y) are two identical coordinate patches. A same vector is denoted by $X = (X^j)$ in U^n and by $Y = (Y^k)$

in V^n. Transformation of the components of the same vector $X = Y$ is described by Eq. (1.38) (using W in place of J):

$$Y^k = W^k_j X^j, \quad W^k_j = \frac{\partial y^k}{\partial x^j}, \tag{1.52}$$

which is equivalent to (1.17). Correspondingly, transformation of bases is described by (1.39):

$$\frac{\partial}{\partial x^j} = W^k_j \frac{\partial}{\partial y^k}, \quad \text{or} \quad \frac{\partial}{\partial y^k} = (W^{-1})^j_k \frac{\partial}{\partial x^j}, \tag{1.53}$$

where $W = (W^k_j)$. It is easy to see the identity: $Y^k\partial/\partial y^k = X^j\partial/\partial x^j$.

The transformation (1.18) corresponds to Eq. (1.48) which describes the transformation of components of a covector. Solving (1.48) for A_k, we obtain

$$A_k = a_i \ (W^{-1})^i_k. \tag{1.54}$$

Thus, we find the invariance of inner product:

$$A_k Y^k = a_i (W^{-1})^i_k W^k_j X^j = a_i \delta^i_j X^j = a_i X^i. \tag{1.55}$$

1.7. Lie Group and Invariant Vector Fields

Dynamical evolution of a physical system is described by a trajectory over a manifold, which is often represented by a space of Lie group, a symmetry group of the system. This and the following section are a concise account of some aspects of the theory of Lie group and Lie algebra related to the present subject.

We consider various Lie groups G associated with various physical systems below. In abstract terms, a group **G** of smooth transformations (maps) of a manifold M into itself is called a group, provided that (i) with two maps $g, h \in G$, the product $gh = g \circ h$ belongs to $G: G \times G \to G$, (ii) for every $g \in G$, there is an inverse map $g^{-1} \in G$. From (i) and (ii), it follows that the group contains an *identity* map *id*, which is often called *unity* denoted by e. Thus, $gg^{-1} = g^{-1}g = e$.

A *Lie group* is a group which is a differentiable manifold, for which the operations (i) and (ii) are differentiable. Some lists of typical Lie groups are given in Appendix C. A Lie group always has two families of diffeomorphisms, the left and right translations. Namely, with a fixed element $h \in G$,

$$L_h(g) = hg \quad (\text{or } R_h(g) = gh), \quad \text{for any } g \in G,$$

where $\mathbf{L}_h$ (or $\mathbf{R}_h$) denotes the *left*- (or *right*-) translation of the group onto itself, respectively. Note that $L_g(h) = R_h(g) = gh$. The operation inverse to L_h (or R_h) is simply $L_{h^{-1}}$ (or $R_{h^{-1}}$), respectively.

Suppose that g_t is a curve on G described in terms of a parameter t. The left translation of g_t by $g_{\Delta t}$ for an infinitesimal Δt is given by $g_{\Delta t} \circ g_t$. Hence the t-derivative is expressed as

$$\dot{g}_t = \lim_{\Delta t \to 0} \frac{g_{t+\Delta t} - g_t}{\Delta t} = X \circ g_t, \quad X = \lim_{\Delta t \to 0} \frac{g_{\Delta t} - id}{\Delta t}, \tag{1.56}$$

where id denotes the identity map. Thus, the *left*-translation leads to the *right*-invariant vector field [AzIz95] in the sense defined just below. Similarly, the *right*-translation leads to the *left*-invariant vector field. The $\dot{g}_t$ is said to be a tangent vector at a point g_t.

A vector field X^{L}, or X^{R} on G is *left-invariant*, or *right-invariant*, if it is invariant under all left-translations, or right-translations respectively, namely for all $g, h \in G$, if

$$(\mathbf{L}_h)_* X_g^{\mathrm{L}} = X_{hg}^{\mathrm{L}}, \quad \text{or} \quad (\mathbf{R}_h)_* X_g^{\mathrm{R}} = X_{gh}^{\mathrm{R}}, \tag{1.57}$$

respectively. Given a tangent vector X to G at e, one may left-translate or right-translate X to every point $g \in G$ as

$$X_g^{\mathrm{L}} = (L_g)_* X = g \circ X = gX, \tag{1.58}$$

$$X_g^{\mathrm{R}} = (R_g)_* X = X \circ g = Xg, \tag{1.59}$$

respectively. It is readily seen from (1.59) that $(R_h)_* X_g^{\mathrm{R}} = X_{gh}^{\mathrm{R}}$, hence the transformation (1.59) gives a right-invariant field generated by X. Similarly, the transformation (1.58) gives a left-invariant field.

Consider a curve $\xi_t : t \in \mathbb{R} \to G$ with the tangent $\dot{\xi}_0 = X$ at $t = 0$. The left-invariant field is given by $X_s^L = (\mathrm{d}/\mathrm{d}t)(g_s \circ \xi_t)$ for $g_s \in G$ (s: a parameter), whereas the right-invariant vector field is represented by $X_s^R = (\mathrm{d}/\mathrm{d}t)(\xi_t \circ g_s)$. Examples of such invariant fields are given by (3.86) and (3.87) in §3.7.3.

The left-translation $(L_g)_* X$ is understood as a transformation of a vector X located at x under the push-forward to gX at $g(x)$ (Fig. 1.17(a)). On the other hand, the right-translation $(R_g)_* X = X \circ g\ (x)$ is understood as follows: first let the map g act on the point x and then the vector X is taken at the point $g(x)$ (Fig. 1.17(b)). This is something like a change of variables when g is an element of a transformation group.

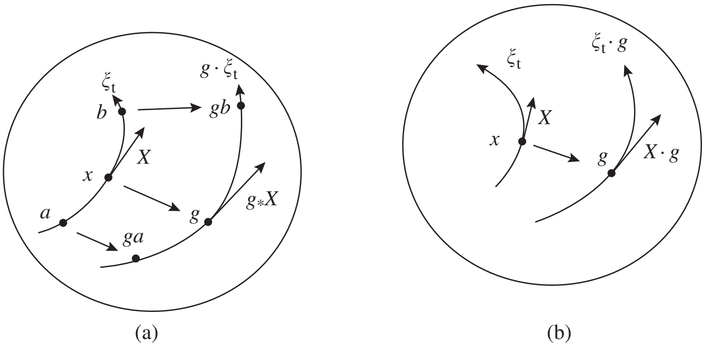

Fig. 1.17. (a) Left- and (b) Right-translation of a vector field X.

For two right-invariant tangent vectors $X_g^{\rm R}$ and $Y_g^{\rm R}$, the metric (1.28) is called *right-invariant* if

$$\langle X_g^{\rm R}, Y_g^{\rm R}\rangle = \langle X_e, Y_e\rangle.$$

Similarly, the metric is *left-invariant* if $\langle X_g^{\rm L}, Y_g^{\rm L}\rangle = \langle X_e, Y_e\rangle$ for left-invariant vectors, $X_g^{\rm L}, Y_g^{\rm L}$. Examples of the left-invariant field are given in Chapters 4 and 9.

1.8. Lie Algebra and Lie Derivative

1.8.1. *Lie algebra, adjoint operator and Lie bracket*

Every pair of vector fields defines a new vector field called the Lie bracket $[\cdot,\cdot]$. More precisely, the tangent space $\mathbf{T}_e\mathbf{G}$ at the identity e of a Lie group G is called the *Lie algebra* g of the group G. The *Lie algebra* g $(= T_eG)$ is equipped with the bracket operation $[\cdot,\cdot]$ of bilinear skew-symmetric pairing, $[\cdot,\cdot] : \mathsf{g}\times\mathsf{g} \to \mathsf{g}$, defined below. The bracket satisfies the *Jacobi identity*,

$$[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0, \tag{1.60}$$

for any triplet of $X, Y, Z \in \mathsf{g}$.

Any element of the Lie algebra $X \in \mathsf{g}$ defines a one-parameter subgroup (Appendix C.2, Eq. (C.4)):

$$\xi_t = \exp[tX] = e + tX + \frac{1}{2!}t^2X^2 + O(t^3), \quad X \in \mathsf{g} \tag{1.61}$$

where ξ_t is a curve $t \to G$ with the tangent $\dot{\xi}_0 = X \in \mathsf{g}$ at $t = 0$. In this sense, the element X is called an (infinitesimal) *generator* of the subgroup.

A Lie group G acts as a group of linear transformations on its own Lie algebra g. Namely for $^{\forall}g \in G$, there is an operator Ad_g, such that

$$Ad_g Y := (L_g)_* \circ (R_{g^{-1}})_* Y = gYg^{-1}, \tag{1.62}$$

for $^{\forall}Y \in \mathsf{g}$.[18] The operator Ad_g transforms $Y \in \mathsf{g}$ into $Ad_g Y \in \mathsf{g}$ linearly (Fig. 1.18). The set of all such Ad_g, i.e. $Ad(G)$, is called the *adjoint* representation of G, an *adjoint group*. Setting g with the inverse $\xi_t^{-1} := (\xi_t)^{-1}$, the adjoint transformation $Ad_{\xi_t^{-1}} Y$ is a function of t. Its derivative with respect to t is a linear transformation from Y to $ad_X Y$ defined by

$$ad_X Y = \frac{\mathrm{d}}{\mathrm{d}t} \xi_t^{-1} Y \xi_t \bigg|_{t=0} := [X, Y]. \tag{1.63}$$

This defines the Lie bracket $[X, Y]$.[19] Its explicit expression depends on each group or each dynamical system considered. It can be shown that the bracket $[X, Y]$ thus introduced satisfies all the properties required for the Lie bracket in each example considered below. The bracket operation is usually called the *commutator*. The ad_X is a linear transformation, $\mathsf{g} \to \mathsf{g}$,

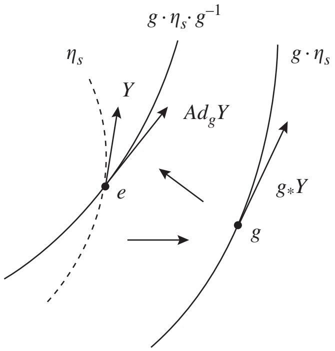

Fig. 1.18. Adjoint transformation $Ad_g Y$, where $Y = \mathrm{d}\eta_s/\mathrm{d}s|_{s=0}$ and $Ad_g Y = (\mathrm{d}/\mathrm{d}s)(g \circ \eta_s \circ g^{-1})|_{s=0}$.

[18] g is the Lie algebra and g is an element of the group G. The operator gYg^{-1} may be better written as the push-forward notation, $g_* Y g^{-1}$.
[19] Most textbooks in mathematics adopt this definition. Arnold [Arn66; Arn78] uses the definition of its opposite sign which is convenient for physical systems related to rotation group (see (1.64)) in Chapters 4 and 9, characterized with the left-invariant metric. In fact, the difference between the left-invariant and right-invariant field, (1.58) and (1.59), results in different signs of the Lie bracket of right- and left-invariant fields (see [AzIz95]).

by the representation, $Y \to ad_X Y = [X, Y]$. The operator ad_X stands for the image of an element X under the linear ad-action.

1.8.2. *An example of the rotation group* $\boldsymbol{SO(3)}$

Consider the rotation group $G = \mathbf{SO(3)}$. Any element $A \in SO(3)$ is represented by a 3×3 orthogonal matrix ($AA^T = I$) of $\det A = 1$ (Appendix C), where A^T denotes transpose of A, i.e. $(A^T)^i_k = A^k_i$. Let $K = (\partial_x, \partial_y, \partial_z)$ be a cartesian right-handed frame. By the element A, the coordinate frame K is transformed to another frame $K' = (\partial_{x'}, \partial_{y'}, \partial_{z'}) = AK$, and a point $X = (x, y, z)$ in K is transformed to $X' = (x', y', z') = WX$ by the rules in §1.6.3, where $W = (A^{-1})^T$. Then, we have

$$(X')^T K' = (WX)^T AK = X^T W^T AK = X^T A^{-1} AK = XK.$$

Consider successive transformations $A' = A_2 A_1$, i.e. A_1 followed by A_2. Then we have $(X')^T = X^T (A_2 A_1)^{-1} = X^T A_1^{-1} A_2^{-1}$, that is, the components X evolve by the right-translation, resulting in the left-invariant vector field (§1.6).[20]

Let $\xi(t)$ be a curve (one-parameter subgroup) issuing from $e = \xi(0)$ with a tangent vector $\mathbf{a} = \dot{\xi}(0)$ on $SO(3)$. Then one has $\xi(t) = \exp[t\mathbf{a}] = e + t\mathbf{a} + O(t^2)$ for an infinitesimal parameter (time) t, where $\mathbf{a}$ is an element of the algebra $\mathbf{g}$ (usually written as $\mathbf{so}(3)$) and a skew-symmetric matrix due to the orthogonality of $\xi(t)$ (Fig. 1.19).

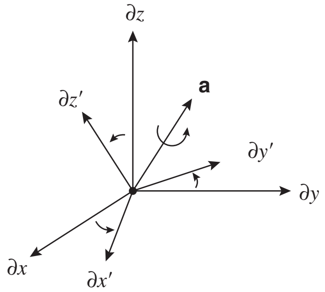

Fig. 1.19. $K = (\partial_x, \partial_y, \partial_z), K' = \xi(t)K$ with $e = \xi(0), \mathbf{a} = \dot{\xi}(0)$.

[20]The length of vector is also invariant by this transformation, i.e. *isometry*, since $\langle X', X' \rangle = \langle X, (W^T W)X \rangle = \langle X, X \rangle$, where $W^T W = (A^{-1})^T A^{-1} = (AA^T)^{-1} = I$.

Then, for $\mathbf{a}, \forall \mathbf{b} \in \mathbf{so}(3)$, the operation $ad_{\mathbf{a}} : \mathbf{g} \to \mathbf{g}$ is represented by

$$ad_{\mathbf{a}}\mathbf{b} = [\mathbf{a}, \mathbf{b}] = -(\mathbf{ab} - \mathbf{ba}), \tag{1.64}$$

where the minus sign in front of $(\mathbf{ab} - \mathbf{ba})$ is due to the definition (1.63). This is verified as follows. Since $\xi(t)^{-1} = \exp[-t\mathbf{a}] = (e - t\mathbf{a} + \cdots)$, we have

$$\begin{aligned}\xi(t)^{-1}\mathbf{b}\xi(t) &= (e - t\mathbf{a} + \cdots)\mathbf{b}(e + t\mathbf{a} + \cdots) \\ &= \mathbf{b} - t(\mathbf{ab} - \mathbf{ba}) + O(t^2).\end{aligned} \tag{1.65}$$

Its differentiation with respect to t results in Eq. (1.64).

In Chapter 4, we consider time trajectories over the rotation group $SO(3)$ such as $\xi(t)$ with time t. In such a case, it is convenient to define the bracket $[\mathbf{a}, \mathbf{b}]^{(\mathrm{L})}$ for the left-invariant field defined as

$$[\mathbf{a}, \mathbf{b}]^{(\mathrm{L})} := -[\mathbf{a}, \mathbf{b}] = \mathbf{ab} - \mathbf{ba} = \mathbf{c}. \tag{1.66}$$

In Appendix C.3, it is shown that, for $\mathbf{a}, \mathbf{b} \in \mathbf{so}(3)$, $\mathbf{c}$ is also skew-symmetric, and that the matrix equation $\mathbf{ab} - \mathbf{ba} = \mathbf{c}$ is equivalent to the cross-product equation (C.14),

$$\hat{\mathbf{c}} = \hat{\mathbf{a}} \times \hat{\mathbf{b}}, \tag{1.67}$$

where $\hat{\mathbf{a}}, \hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ are three-component (axial) vectors associated with the skew-symmetric matrices $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, respectively.

1.8.3. *Lie derivative and Lagrange derivative*

(a) **Derivative of a scalar function $f(x)$**

Suppose that a vector field $X = X^i \partial_i$ is given on a manifold M^n. As described in §1.2, with every such vector field, one can associate a *flow*, or one-parameter group of *diffeomorphisms* $\xi_t : M^n \to M^n$, for which $\xi_0 = id$[21] and $(\mathrm{d}/\mathrm{d}t)\xi_t x|_{t=0} = X(x)$. A first order differential operator $\mathcal{L}_X$ on a scalar function $f(x)$ on M (a function of coordinates x only) is defined as

$$\mathcal{L}_X f(x) := \frac{\mathrm{d}}{\mathrm{d}t}\,(\xi_t)^* f(x)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} f(\xi_t x)\bigg|_{t=0} = X^i \frac{\partial}{\partial x^i} f(x), \tag{1.68}$$

(see (1.44) and below, and (1.5)). This defines the derivative $\mathcal{L}_X f$ of a function f (a zero-form) by the time derivative of its pull-back $\xi_t^* f$ at $t = 0$, where the point $\xi_t x$ moves forward in accordance with the flow of

[21] The *id* is used here in order to emphasize that this is an *identity map*.

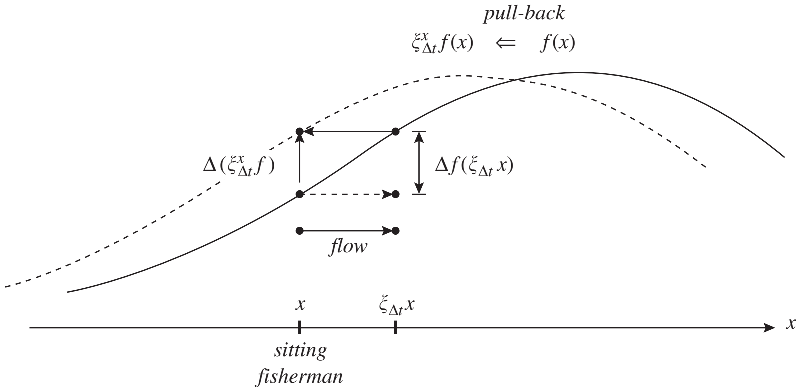

Fig. 1.20. Fisherman's derivative.

velocity X. Relatively observing, the pull-back $\xi_t^* f$ is evaluated at x, and its time derivative is defined by the Lie derivative. This is sometimes called as a *derivative of a fisherman* [AK98] sitting at a fixed place x (Fig. 1.20).[22]

In *fluid dynamics* however, the same derivative is called the *Lagrange* derivative, which refers to the third and fourth expressions,

$$\frac{\mathrm{D}f}{\mathrm{D}t} := \frac{\mathrm{d}}{\mathrm{d}t} f(\xi_t x) = X^i \frac{\partial}{\partial x^i} f(x).$$

Therefore we obtain that $(\mathrm{D}f/\mathrm{D}t)f = \mathcal{L}_X f$, which is valid for scalar functions. But this does not hold for vectors, as shown in the next subsection.

In the unsteady problem, the right-hand side is written as $(\partial_t + X^i \partial_i) f(x,t)$. The Lagrange derivative is understood as denoting the time derivative, with respect to the fluid particle $\xi_t x$ moving with the flow, of the function $f(x,t)$.

(b) Derivatives of a vector field $\boldsymbol{Y(x)}$

Now, suppose that we are given a second vector field $Y(x) = Y^i \partial_i$, and consider its time derivative along the X-flow generated by $X(x)$. To that end, let us denote the second Y-flow generated by $Y(x)$ as η_s with $\eta_0 = e = id$. The first flow ξ_t transports the vector $Y(x)$ in front of a fisherman sitting at a point x. After an infinitesimal time t, the fluid particle at x

[22]The Lie derivative $\mathcal{L}_X$ also acts on any form field α in the same way, $(\mathrm{d}/\mathrm{d}t)(\xi_t)^*\alpha$ as (1.78). On the contrary, to a vector field Y, the Lie derivative is defined by (1.69) in terms of the push-forward $(\xi_t)_*$. This definition is different from that of Arnold [1978, 1966] by the sign, but consistent with the present definition of (1.63).

will arrive at $\xi_t x$. We take the vector Y at this point $\xi_t x$ and translate it backwards to the original point x by the inverse map of the push-forward, that is $(\xi_t)^{-1}Y(\xi_t x)$, in precise $(\xi_t)_*^{-1}Y(\xi_t x)$. Its time derivative is the *Lie derivative of a vector* Y, given by

$$\mathcal{L}_X Y := \lim_{t\to 0} \frac{\xi_{t*}^{-1}Y(\xi_t x) - Y(x)}{t} = \lim_{t\to 0} \xi_{t*}^{-1} \left. \frac{Y\xi_t - \xi_{t*}Y}{t} \right|_x$$
$$= \lim_{t\to 0} \frac{1}{t}(Y\xi_t - \xi_{t*}Y). \tag{1.69}$$

The first expression is nothing but that of $ad_X Y(x)$ according to (1.63). Thus we have

$$\mathcal{L}_X Y = \left. \frac{\mathrm{d}}{\mathrm{d}t}\xi_{t*}^{-1}Y\xi_t \right|_{t=0} = ad_X Y = [X,Y], \tag{1.70}$$

where $[X,Y]$ is the *Lie bracket* (see (1.63)).

The last expression of (1.69) suggests another useful expression of $[X,Y]$, which is given by

$$\mathcal{L}_X Y = [X,Y] := \lim_{t\to 0, s\to 0} \frac{1}{st}(\eta_s\xi_t - \xi_t\eta_s)$$
$$= \left. \frac{\partial}{\partial t}\frac{\partial}{\partial s}\,(\eta_s\xi_t - \xi_t\eta_s)\right|_{t=0,s=0}. \tag{1.71}$$

According to Appendix C, the two flows ξ_t and η_s generated by X and Y can be written in the form [AK98, §2]:

$$\xi_t : x \mapsto x + tX(x) + O(t^2), \quad t \to 0, \tag{1.72}$$
$$\eta_s : x \mapsto x + sY(x) + O(s^2), \quad s \to 0. \tag{1.73}$$

Recalling that $\eta_s\xi_t(x)$ for diffeomorphisms is given by $\eta_s(\xi_t(x))$, i.e. the composition rule, we have

$$\eta_s\xi_t = e + tX + O(t^2) + sY(\xi_t) + O(s^2)$$
$$= e + tX + sY + stX^j\partial_j Y + O(t^2, s^2), \tag{1.74}$$

where $\xi_t = e + tX + O(t^2)$. The expression of $\xi_t\eta_s$ is obtained by exchanging the pairs (t,X) and (s,Y). Thus finally, we have

$$\eta_s\xi_t - \xi_t\eta_s = st\left(X^j\frac{\partial Y}{\partial x^j} - Y^j\frac{\partial X}{\partial x^j}\right) + O(st^2, s^2t). \tag{1.75}$$

The first term may be written as $st[X, Y]$ according to (1.71). Thus the non-commutativity of two diffeomorphisms ξ_t and η_s is proportional to $[X, Y]$, where

$$[X, Y] = \{X, Y\} := \{X, Y\}^k \partial_k, \tag{1.76}$$

$$\{X, Y\}^k := X^j \frac{\partial Y^k}{\partial x^j} - Y^j \frac{\partial X^k}{\partial x^j} \tag{1.77}$$

and $\{X, Y\}$ is the *Poisson bracket*. The degree of non-commutativity of ξ_t and η_s is interpreted graphically in Fig. 1.21. According to the definition (1.70), by using the expression $L_X := X^i \partial_i$, the Lie derivative of the vector field Y with respect to X is given by

$$\mathcal{L}_X Y = [X, Y] = L_X L_Y - L_Y L_X = [L_X, L_Y] = L_{\{X,Y\}}. \tag{1.78}$$

If they commute, i.e. $\xi_t \eta_s = \eta_s \xi_t$, then obviously we have $[X, Y] = 0$. This suggests that the coordinate bases commute since the coordinate curves are defined to intersect. In fact, for $X = \partial_\alpha, Y = \partial_\beta$, we obtain from (1.78)

$$[\partial_\alpha, \partial_\beta] = \partial_\alpha \partial_\beta - \partial_\beta \partial_\alpha = 0. \tag{1.79}$$

In general, we have

$$\xi_t \eta_s - \eta_s \xi_t = [X, Y]st + O(st^2, s^2 t).$$

If $[X, Y] = 0$, then we obtain $\xi_t \eta_s - \eta_s \xi_t = O(st^2, s^2 t)$.

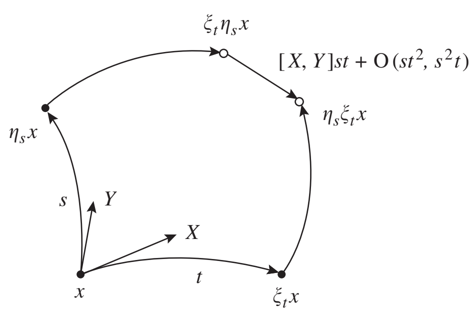

Fig. 1.21. Graphic interpretation of $[X, Y]$ for infinitesimal s and t.

In unsteady problem of *fluid dynamics*, the Lagrange derivative of the vector $Y = Y^k(x,t)\partial_k$ is defined by

$$\frac{\mathrm{D}}{\mathrm{D}t}Y = \frac{\mathrm{D}}{\mathrm{D}t}Y^k(\xi_t x)\partial_k := \frac{\partial Y^k}{\partial t}\partial_k + X^j \frac{\partial Y^k}{\partial x^j}\partial_k. \tag{1.80}$$

This derivative makes sense in the gauge-theoretical formulation described in §7.5 and denotes the derivative following a fluid particle moving with the velocity $X^j\partial_j$, whereas the Lie derivative characterizes a frozen field (see the remark just below).

Remark. A vector field Y defined along the integral curve ξ_t generated by the tangent field X is said to be *invariant* if $Y(\xi_t x) = (\xi_t)_* Y(x)$. Substituting this in the previous expression of (1.69), it is readily seen that $\mathcal{L}_X Y = 0$, or rewriting it,

$$\mathcal{L}_X Y = \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \right) \partial_i = 0. \tag{1.81}$$

In unsteady problem, $X^j\partial_j Y^i$ is also written as $\mathrm{D}Y^i/\mathrm{D}t$ given by the right-hand side of (1.80). Then, using the operator $\mathrm{D}/\mathrm{D}t = \partial_t + X^j\partial_j$, the above equation (1.81) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}Y = (Y^j\partial_j)X. \tag{1.82}$$

In *fluid dynamics*, the equation $\mathcal{L}_X Y = 0$ is called the equation of *frozen field* (Fig. 1.22).[23] If we set $\phi = \xi_t$ in (1.37) together with $X = Y(x)$ and $Y = Y(\xi_t x)$, then the equation $Y(\xi_t x) = (\xi_t)_* Y(x)$ represents the

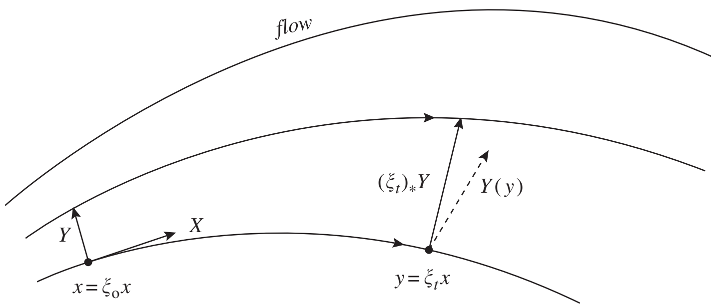

Fig. 1.22. Frozen field $(\xi_t)_*Y$ (push-forward of Y) coincides with $Y(y)$ at $y = \xi_t x$.

[23]The Jacobi field Y ($= J$ below) satisfies this equation (see §8.4).

push-forward transformation. Therefore, writing $\xi_t x = y_t(x)$, the solution of (1.82) is given by Eq. (1.38),

$$Y^\alpha(t) = Y^j(0)\frac{\partial y_t^\alpha}{\partial x^j}, \tag{1.83}$$

which is called the *Cauchy's solution* [Cau1816] in the fluid dynamics.

1.9. Diffeomorphisms of a Circle S^1

> *A smooth sequence of diffeomorphisms is a mathematical concept of a flow and the unit circle S^1 is one of the simplest base manifolds for physical fields.*

Diffeomorphism of the manifold S^1 (a unit circle in $\mathbb{R}^2$, see Fig. 1.3) is represented by a map $g : x \in \mathbb{R}^1 \to g(x) \in \mathbb{R}^1$ (where $g \in C^\infty$) with every point of x or $g(x)$ is identified with $x+1$ or $g(x)+1$ respectively.[24] Collection of all such maps constitutes a group $\mathcal{D}(S^1)$ of diffeomorphisms with the composition law:

$$h = g \circ f, \quad \text{i.e.} \quad h(x) = g(f(x)) \in \mathcal{D}(S^1),$$

for $f, g \in \mathcal{D}(S^1)$. The diffeomorphism is a map of infinite degrees of freedom (i.e. having *pointwise* degrees of freedom). In Chapter 5, the diffeomorphism is assumed to be orientation-preserving in the sense that $g'(x) > 0$, where the prime denotes $\partial/\partial x$.

Consider a flow $\xi_t(x)$ which is a smooth sequence of diffeomorphisms with the time parameter as t (see (1.72)). Its tangent field at ξ_t is defined by

$$\dot{\xi}_t(x) := \frac{\mathrm{d}}{\mathrm{d}t}\xi_t(x)\bigg|_t = \lim_{\tau\to 0}\frac{\xi_\tau(x) - id}{\tau} \circ \xi_t(x) = u(x) \circ \xi_t(x),$$

in a right-invariant form. The tangent field $X(x)$ at the identity (id) is given by $u(x) = \mathrm{d}\xi_t(x)/\mathrm{d}t|_{t=0}$.

Alternatively, with the language of differentiable manifolds, the tangent field $X(x)$ is represented as

$$X(x) = u(x)\partial_x \in TS^1, \tag{1.84}$$

where TS^1 is a tangent bundle (§1.3.1) over the manifold S^1. The tangent bundle TS^1 allows a global product structure $S^1 \times \mathbb{R}^1$ as shown in Fig. 1.23(a). The figure (b) is obtained from (a) by cutting it along one

[24] By the map $\phi(x) = e^{i2\pi x}$, there is a periodititY $\phi(x+1) = \phi(x)$ for $x \in \mathbb{R}^1$.

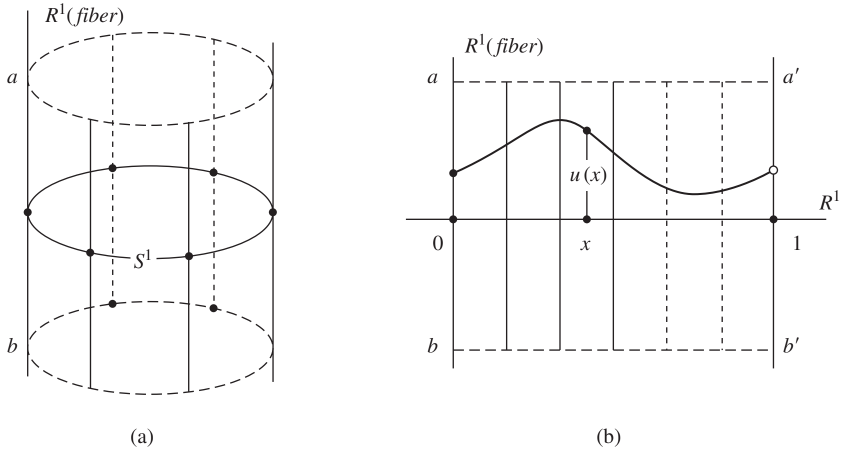

Fig. 1.23. Tangent bundle $S^1 \times \mathbb{R}^1$ with the circle S^1 and the fiber $\mathbb{R}^1$.

fiber ab and developing it flat, where $a'b'$ is identified with ab. The solid curve in the figure describes a particular vector field $u(x)$ on S^1, which is called a *cross-section* of the tangent bundle TS^1.

If $a'b'$ is identified with ba by twisting the strip, then a Möbius band is formed. The resulting fiber bundle is not trivial, i.e. not a product space (see, e.g. [Sch80]). The Möbius band is a *double-fold cover*, i.e. two-sheeted cover of the circle S^1 (see Fig. 1.4).

For two diffeomorphisms ξ_t and η_t corresponding to the vector fields X and Y respectively, the *Lie bracket* (commutator) is given by (1.76) and (1.77):

$$[X, Y] = (uv' - vu')\partial_x, \tag{1.85}$$

where $X = u(x)\partial_x, Y = v(x)\partial_x \in TS^1$. This is sometimes called *Witt algebra* [AzIz95].

1.10. Transformation of Tensors and Invariance

1.10.1. *Transformations of vectors and metric tensors*

We considered the transformations of vectors and covectors in §1.4 (see also (1.10)) together with the invariance of the value of covectors based on (1.44).

Here, we consider a transformation and an invariant property of tensors.[25] Let M be an n-manifold with a Riemannian metric and covered with a family of local (curvilinear) coordinate systems, $\{U : x^1, \dots, x^n\}$, $\{V : y^1, \dots, y^n\}, \dots$, where $U, V, \dots$ are open sets (called *patches*) with coordinates $x, y, \dots$. A point $p \in U \cap V$, lying in two overlapping patches U and V, has two sets of coordinates $x_{(p)}$ and $y_{(p)}$ which are related differentiably by the functions $y^k(x)$:

$$y^k_{(p)} = y^k(x^1_{(p)}, \dots, x^n_{(p)}), \quad k = 1, \dots, n.$$

In the corresponding tangent spaces, the vectors are represented as $X = X^i \partial/\partial x^i \in T_p U$ and $Y = Y^k \partial/\partial y^k \in T_p V$. The coordinate bases are transformed according to

$$\frac{\partial}{\partial x^i} = \frac{\partial y^k}{\partial x^i}\frac{\partial}{\partial y^k} = W^k_i \frac{\partial}{\partial y^k}, \quad \text{where} \quad W^k_i = \frac{\partial y^k}{\partial x^i}, \tag{1.86}$$

by the chain rule (in an analogous way to (1.39)), W^k_i being the transformation matrix. Suppose that the components of the vectors are related by

$$Y^k = W^k_i X^i, \quad (\text{written as } Y = WX), \tag{1.87}$$

as is the case of the push-forward transformation (1.38), then the vectors are invariant in the sense:

$$X = X^i \frac{\partial}{\partial x^i} = X^i W^k_i \frac{\partial}{\partial y^k} = Y^k \frac{\partial}{\partial y^k} = Y.$$

Equation (1.87) is the rule of transformation of *vector components*. In physical problems, the logic is reversed. The vector, like the velocity vector of a particle, should be the same (may be written as $X = Y$) in both coordinate frames. Then the components must be transformed according to the rule (1.87).

The metric tensor is defined by (1.29). According to the basis transformation (1.86), we obtain

$$\begin{aligned} g_{ij}(x) &= \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle = \left\langle W^k_i \frac{\partial}{\partial y^k}, W^l_j \frac{\partial}{\partial y^l} \right\rangle \\ &= W^k_i W^l_j \left\langle \frac{\partial}{\partial y^k}, \frac{\partial}{\partial y^l} \right\rangle = W^k_i W^l_j g_{kl}(y). \end{aligned} \tag{1.88}$$

[25]See §3.11 for differentiation of tensors.

This is the transformation rule of the tensor g_{ij}. Using (1.87) for the transformation of two pairs (X, Y) and (ξ, η) of tangent vectors, where $X, \xi \in T_pU$ and $Y, \eta \in T_pV$, we have the invariance of the inner product with the transformation (1.88):

$$\begin{aligned} G(\xi, X) = \langle \xi, X \rangle(x) = g_{ij}(x)\xi^i X^j = W_i^k W_j^l g_{kl}(y)\xi^i X^j \\ = g_{kl}(y)\eta^k Y^l = \langle \eta, Y \rangle(y), \end{aligned} \tag{1.89}$$

where $Y^k = W_i^k X^i$ and $\eta^l = W_j^l \xi^j$.

1.10.2. *Covariant tensors*

The inner product $G(\xi, X)$ in the previous section is an example of covariant tensor of rank 2. In general, a *covariant tensor* of rank n is defined by

$$Q^{(n)} : E_1 \times E_2 \times \cdots \times E_n \to \mathbb{R},$$

a multilinear real-valued function of n-tuple vectors, written as $Q^{(n)}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ which is linear in each entry $\boldsymbol{v}_i$ $(i = 1, \ldots, n)$, where E_k is the tangent vector space for the kth entry.

A covector $\alpha = a_i \mathrm{d}x^i$ on a vector $\boldsymbol{v} = v^j \partial_j = v^j \partial/\partial x^j$ is an example of $Q^{(1)}$, a covariant tensor of rank 1. In fact, we have $\alpha(\boldsymbol{v}) = a_i v^j \mathrm{d}x^i(\partial_j) = a_i v^i$. An example of $Q^{(2)}$ is $G(A, X) = g_{ij} A^i X^j$. Both are shown to be invariant with the coordinate transformation (see §1.6.3 for $\alpha(\boldsymbol{v})$).

In general, the values of $Q^{(n)}$ must be independent of the basis with respect to which components of the vectors are expressed. In components, we have

$$\begin{aligned} \bar{Q}^{(n)}(x) := Q^{(n)}(\boldsymbol{v}_1(x), \ldots, \boldsymbol{v}_n(x)) = Q^{(n)}(v_1^{k_1}\partial_{k_1}, \ldots, v_n^{k_n}\partial_{k_n}) \\ = Q^{(n)}_{k_1,\ldots,k_n}(x) v_1^{k_1} \cdots v_n^{k_n} \end{aligned}$$

at $x \in U$, where

$$Q^{(n)}_{k_1,\ldots,k_n}(x) = Q^{(n)}(\partial/\partial x^{k_1}, \ldots, \partial/\partial x^{k_n}).$$

Considering that the bases are transformed according to (1.86) and $Q^{(n)}$ is multilinear, we have the transformation rule,

$$Q^{(n)}_{k_1,\ldots,k_n}(x) = W^{l_1}_{k_1} \cdots W^{l_n}_{k_n} Q^{(n)}_{l_1,\ldots,l_n}(y). \tag{1.90}$$

Owing to the transformation (1.87), it is obvious that we have the invariance, $\bar{Q}^{(n)}(x) = \bar{Q}^{(n)}(y)$.

From two covectors $\alpha = a_i \mathrm{d}x^i$ and $\beta = b_j \mathrm{d}x^j$, one can form a covariant tensor of rank 2 by the *tensor product* $\otimes$ as follows: $\alpha \otimes \beta : E \times E \to \mathbb{R}$, defined by

$$\begin{aligned} \alpha \otimes \beta(\boldsymbol{v}, \boldsymbol{w}) &:= \alpha(\boldsymbol{v})\beta(\boldsymbol{w}) = a_i \mathrm{d}x^i \otimes b_j \mathrm{d}x^j(\boldsymbol{v}, \boldsymbol{w}) \\ &= Q_{kl} v^k w^l, \\ Q_{kl} &= Q^{(2)}(\partial_k, \partial_l) = a_i b_j \mathrm{d}x^i \otimes \mathrm{d}x^j(\partial_k, \partial_l) \end{aligned} \tag{1.91}$$

where $\boldsymbol{v} = v^k \partial_k, \boldsymbol{w} = w^l \partial_l \in E$.

1.10.3. *Mixed tensors*

A mixed tensor of rank 2 is defined by

$$M^i_j(x) = M^{(2)}\left(\mathrm{d}x^i, \frac{\partial}{\partial x^j}\right),$$

which is a first order covariant and first order contravariant tensor. According to (1.50), a 1-form base $\mathrm{d}x^i$ is transformed as

$$\mathrm{d}x^i = \frac{\partial x^i}{\partial y^j}\,\mathrm{d}y^j = \hat{W}^i_j\,\mathrm{d}y^j, \quad \text{where} \quad \hat{W}^i_j := \partial x^i / \partial y^j. \tag{1.92}$$

Thus, using (1.86) and (1.92), we obtain the transformation rule of the mixed tensor M:

$$M^i_j(x) = W^\beta_j \hat{W}^i_\alpha M^\alpha_\beta(y). \tag{1.93}$$

The transformation matrix $\hat{W} = \partial x / \partial y$ is the inverse of $W = \partial y / \partial x$, i.e. $\hat{W} = W^{-1}$, since one can verify $W\hat{W} = I$, i.e.

$$(W\hat{W})^k_j = \frac{\partial y^k}{\partial x^\beta}\frac{\partial x^\beta}{\partial y^j} = \frac{\partial y^k}{\partial y^j} = \delta^k_j.$$

Let us consider such mixed tensors through several examples.

(i) *Transformation*: A mixed tensor $M^{(2)}$ of rank 2 arises from the matrix of a linear transformation $W = (W^k_i)$. In the coordinate patch V, taking a covariant vector $\alpha = A_i \mathrm{d}y^i \in E^*$ (cotangent space, §1.5.1) and a contravariant vector $Y = Y^k \partial_k \in E$, the mixed tensor $M^{(2)} : E^* \times E \to R$ is defined by $M^{(2)}(\alpha, Y) \equiv \alpha[Y] = A_i Y^i$. Next, consider the transformation ϕ and its matrix $W(= \phi_*) : E(U) \to E(V)$, i.e. $Y = WX$ defined by (1.86) and

(1.87). The corresponding pull-back is given by $\phi^*\alpha = a_i \mathrm{d}x^i$ (see (1.46)), and the component a_i is expressed by (1.47):

$$a_i = A_j \frac{\partial y^j}{\partial x^i} = A_j W_i^j. \tag{1.94}$$

Thus, we have the invariance of the value of the mixed tensor as follows (using $Y^j = W_i^j X^i$):

$$M_W^{(2)}(\alpha, Y) := \alpha[Y] = A_j W_i^j X^i = a_i X^i = \phi^*\alpha[X]. \tag{1.95}$$

According to (1.93), the transformation of the tensor $W = (W_\beta^\alpha)$ is $(W_j^\beta \hat{W}_\alpha^i)\ W_\beta^\alpha = W_j^i$. Namely, the transformation of W is an identity: $W \to W$.

(ii) *Vector-valued one-form*: Next example of the mixed tensor is the tensor product, $M^{(2)} = \boldsymbol{v} \otimes \alpha : E^* \times E \to \mathbb{R}$, of a vector and a covector, defined by

$$\begin{aligned} M_V^{(2)}(\beta, \boldsymbol{w}) &:= \boldsymbol{v} \otimes \alpha(\beta, \boldsymbol{w}) = v^j \partial_j \otimes a_i \mathrm{d}x^i(\beta, \boldsymbol{w}) \\ &= \partial_j(\beta) v^j a_i \mathrm{d}x^i(\boldsymbol{w}) = b_j M_i^j w^i, \end{aligned} \tag{1.96}$$

where $\boldsymbol{w} = w^i \partial_i, \beta = b_i \mathrm{d}x^i$ and $M_i^j = v^j a_i$. The value of the tensor $\boldsymbol{v} \otimes \alpha$ on a vector $X = X^i \partial_i$ takes the value of a vector (rather than a scalar) as follows:

$$M_V^{(2)}(X) = \boldsymbol{v} \otimes \alpha(X) = \boldsymbol{v} \otimes a_i \mathrm{d}x^i(X) = \boldsymbol{v} a_i X^i. \tag{1.97}$$

In this sense, $M_V^{(2)} = \boldsymbol{v} \otimes \alpha$ is interpreted also as a *vector-valued 1-form.*

In particular, the following I is the identity mixed-tensor:

$$I := \partial_i \otimes \mathrm{d}x^i. \tag{1.98}$$

In fact, we have

$$I(X) = I(X^\alpha \partial_\alpha) = \partial_i \otimes \mathrm{d}x^i(X^\alpha \partial_\alpha) = X^\alpha \delta_\alpha^i \partial_i = X.$$

(iii) *Covariant derivative*: The third example is the covariant differentiation ∇, which is an essential building block in the differential geometry and also in *Physics*, and investigated in the subsequent chapters as well.

Consider a vector $X = X^k \partial_k$ and the transformation $X = \hat{W} Y$ with $\hat{W}_i^k = \partial x^k / \partial y^i$. It may appear that, just like the tensor $W_i^k = \partial y^k / \partial x^i$, the derivative $\partial X^k/\partial x^j$ is also a mixed tensor. But this is not the case. In fact, we have

$$\begin{aligned}\frac{\partial X^k}{\partial x^j} &= \frac{\partial}{\partial x^j} Y^\alpha \hat{W}^k_\alpha = \frac{\partial}{\partial x^j}\left(Y^\alpha \frac{\partial x^k}{\partial y^\alpha}\right)\\ &= \frac{\partial Y^\alpha}{\partial y^\beta}\frac{\partial y^\beta}{\partial x^j}\frac{\partial x^k}{\partial y^\alpha} + Y^\alpha \frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta}\frac{\partial y^\beta}{\partial x^j} = W^\beta_j \hat{W}^k_\alpha \frac{\partial Y^\alpha}{\partial y^\beta} + W^\beta_j \frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta} Y^\alpha.\end{aligned} \tag{1.99}$$

Only the first term follows the transformation rule (1.93), while the second does not. In order to overcome this difficulty, the differential geometry introduces the following *linear* operator ∇_{∂_j} on the product of a scalar X^k and a vector ∂_k, defined by

$$\begin{aligned}\nabla_{\partial_j}(X^k\partial_k) &= (\nabla_{\partial_j}X^k)\partial_k + X^\alpha(\nabla_{\partial_j}\partial_\alpha)\\ &:= \frac{\partial X^k}{\partial x^j}\partial_k + X^\alpha \Gamma^k_{j\alpha}\partial_k,\end{aligned} \tag{1.100}$$

where Γ^k_{ij} is the *Christoffel symbol*, which can be represented in terms of the derivatives of the metric tensors $g_{\alpha\beta}$ (see §2.4 and 3.3.2). From (1.100), the following mixed tensor is defined:

$$X^k_{;j} := \frac{\partial X^k}{\partial x^j} + \Gamma^k_{j\alpha}X^\alpha. \tag{1.101}$$

In fact, it can be verified that this tensor is transformed like a mixed tensor according to $X^k_{;j}(x) = X^\alpha_{;\beta}(y) W^\beta_j \hat{W}^k_\alpha$ (see e.g. [Eis47]).

In order to write it in the form of a vector-valued 1-form just as (1.97), it is useful to define

$$\nabla X = \partial_k \otimes (\nabla X^k), \tag{1.102}$$

$$\nabla X^k = \mathrm{d}X^k + \Gamma^k_{i\alpha}X^\alpha \mathrm{d}x^i = \frac{\partial X^k}{\partial x^i}\mathrm{d}x^i + \Gamma^k_{i\alpha}X^\alpha\mathrm{d}x^i = X^k_{;i}\mathrm{d}x^i.$$

Then the value of ∇X on a vector $v = v^j\partial_j$ is found to be a vector, which is given by

$$\begin{aligned}\nabla X(v) &= \partial_k \otimes \nabla X^k(v^j\partial_j) = v^j \nabla X^k(\partial_j)\partial_k\\ &= \left(v^j\frac{\partial X^k}{\partial x^j} + \Gamma^k_{j\alpha}X^\alpha v^j\right)\partial_k = v^j X^k_{;j}\partial_k.\end{aligned} \tag{1.103}$$

The operator ∇X^k is called a *connection 1-form* (see §3.5), and $\nabla X(v)$ is called the covariant derivative of X with respect to the vector v.

(iv) *Riemann tensors*: In the differential geometry, a fourth order tensor called the Riemann's curvature tensor plays a central role. This is defined as $R^l_{ijk} = \partial_j \Gamma^l_{ik} - \partial_k \Gamma^l_{jl} + \Gamma^l_{jm}\Gamma^m_{ki} - \Gamma^l_{km}\Gamma^m_{ji}$ (see §2.4 and 3.9.2). It can be verified (e.g. [Eis47]) that this tensor is transformed according to

$$R^l_{ijk}(x) = R^\delta_{\alpha\beta\gamma}(y)\hat{W}^l_\delta W^\alpha_i W^\beta_j W^\gamma_k, \tag{1.104}$$

showing that R^l_{ijk} is a mixed tensor of rank 4, the third order covariant and the first order contravariant tensor.

(v) *General mixed tensor*: In general, a *mixed tensor* of rank n is defined by

$$M^{(n)} : E_1^* \times \cdots \times E_q^* \times E_1 \times \cdots \times E_p \to \mathbb{R}.$$

This is a p times covariant and q times contravariant tensor $(p + q = n)$ and a multilinear real-valued function of p-tuple vectors and q-tuple covectors, written as $M^{(n)}(\alpha_1, \ldots, \alpha_q, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$, which is linear in each entry α_i $(i = 1, \ldots, q)$ and $\boldsymbol{v}_i$ $(i = 1, \ldots, p)$. The values of $M^{(n)}$ is independent of the basis by which the components of the vectors are expressed. In components, we have

$$\hat{M}^{(n)}(x) = M^{(n)}(\alpha_1, \ldots, \alpha_q, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_p) = a_{1k_1} \cdots a_{qk_q} M^{k_1 \cdots k_q}_{l_1 \cdots l_p} v_1^{l_1} \cdots v_p^{l_p},$$

where

$$M^{k_1 \cdots k_q}_{l_1 \cdots l_p} = M^{(n)}(\mathrm{d}x^{k_1}, \ldots, \mathrm{d}x^{k_q}, \partial_{l_1}, \ldots, \partial_{l_p}).$$

1.10.4. *Contravariant tensors*

In the second example (ii) of the mixed tensor, we obtained the expression, $M^{(2)} = b_j M^j_i w^i$. According to the rule (1.31) of §1.3, the lower-index component b_j is related to the upper-index component B^k (the vector counterpart of b_j) by means of the metric tensor g_{jk} as $b_j = g_{jk}B^k$. Similarly, according to (1.33), the upper-index component w^i is related with its covector counter part $W = W_l \mathrm{d}x^l$ as $w^i = g^{il}W_l$ by means of the inverse of the metric tensor $g^{il} = (g^{-1})^{il}$. Hence, we have

$$M^{(2)} = b_j M^j_i w^i = (g_{jk}M^j_i)B^k w^i = (g^{il}M^j_i)b_j W_l. \tag{1.105}$$

Thus it is found that a covariant tensor M_{ki} of rank 2 is obtained by lowering the upper index of the mixed tensor of rank 2:

$$M_{ki} = g_{jk}M^j_i.$$

Similarly, a contravariant tensor M^{jl} of rank 2 is obtained by raising the lower index:

$$M^{jl} = g^{il} M_i^j.$$

In this way, we have found the equivalence:

$$M^{jl} b_j W_l = M_i^j b_j w^i = M_{ki} B^k w^i.$$

In tensor analysis, one can use the same letter M for the derived tensors by lowering or raising the indices by means of the metric tensor.

In general, a *contravariant tensor* of rank n is defined by

$$P^{(n)} : E_1^* \times E_2^* \times \cdots \times E_n^* \to \mathbb{R},$$

a multilinear real-valued function of n-tuple covectors, written as $P^{(n)}(\alpha_1, \ldots, \alpha_n)$ which is linear in each entry α_i $(i = 1, \ldots, n)$. The values of $P^{(n)}$ is independent of the basis. In components, we have

$$\bar{P}^{(n)} = P(\alpha_1, \ldots, \alpha_n) = a_{1\ k_1} \cdots a_{k\ k_n} P^{k_1, \ldots, k_n},$$

where

$$P^{k_1, \ldots, k_n} = P(\mathrm{d}x^{k_1}, \ldots, \mathrm{d}x^{k_n}).$$

From the two vectors $\boldsymbol{v} = v^i \partial_i$ and $\boldsymbol{w} = w^j \partial_j$, one can form a contravariant tensor of rank 2 by the *tensor product*: $\boldsymbol{v} \otimes \boldsymbol{w}$, defined by

$$\boldsymbol{v} \otimes \boldsymbol{w}\ (\mathrm{d}x^i, \mathrm{d}x^j) := \mathrm{d}x^i(\boldsymbol{v})\, \mathrm{d}x^j(\boldsymbol{w}) = v^i w^j. \tag{1.106}$$