28 Analysis of Tensors

Tensor *algebra* deals with lifeless vectors and tensors—objects that do not move, do not change, possess no dynamics. Whenever there is a need for tensors in physics, there is also a need to know the way these tensors change with position and time. Tensors that depend on position and time are called tensor fields and are the subject of this chapter.

In studying the algebra of tensors, we learned that they are generalizations of vectors. Once we have a vector space $\mathcal{V}$ and its dual space $\mathcal{V}^*$, we can take the tensor products of factors of $\mathcal{V}$ and $\mathcal{V}^*$ and create tensors of various kinds. Thus, once we know what a vector is, we can make up tensors from it.

In our discussion of tensor algebra, we did not concern ourselves with what a vector was; we simply assumed that it existed. Because all the vectors considered there were stationary, their mere existence was enough. However, in tensor analysis, where things keep changing from point to point (and over time), the existence of vectors at one point does not guarantee their existence at all points. Therefore, we now have to demand more from vectors than their mere existence. Tied to the concept of vectors is the notion of space, or space-time. Let us consider this first.

28.1 Differentiable Manifolds

Space is one of the undefinables in elementary physics. Length and time intervals are concepts that are "God given", and any definitions of these concepts will be circular. This is true as long as we are confined within a single space. In classical physics, this space is the three-dimensional Euclidean space in which every motion takes place. In special relativity, space is changed to Minkowski space-time. In nonrelativistic quantum mechanics, the underlying space is the (infinite-dimensional) Hilbert space, and time is the only dynamical parameter. In the general theory of relativity, gravitation and space-time are intertwined through the concept of curvature.

Mathematicians have invented a unifying theme that brings the common features of all spaces together. This unifying theme is the theory of differentiable manifolds. A rigorous understanding of differentiable manifolds is

S. Hassani, *Mathematical Physics*, DOI 10.1007/978-3-319-01195-0_28,

beyond the scope of this book. However, a working knowledge of manifold theory is surprisingly simple. Let us begin with a crude definition of a differentiable manifold.

differentiable manifold provisionally defined

Definition 28.1.1 A **differentiable manifold** is a collection of objects called **points** that are connected to each other in a smooth fashion such that the neighborhood of each point looks like the neighborhood of an m-dimensional (Cartesian) space; m is called the **dimension** of the manifold.

As is customary in the literature, we use "manifold" to mean "differentiable manifold".

Example 28.1.2 The following are examples of differentiable manifolds.

(a) The space $\mathbb{R}^n$ is an n-dimensional manifold.
(b) The surface of a sphere is a two-dimensional manifold.
(c) A torus is a two-dimensional manifold.
(d) The collection of all $n \times n$ real matrices whose elements are real functions having derivatives of all orders is an n^2-dimensional manifold. Here a point is an $n \times n$ matrix.
(e) The collection of all rotations in $\mathbb{R}^3$ is a three-dimensional manifold. (Here a point is a rotation.)
(f) Any smooth surface in $\mathbb{R}^3$ is a two-dimensional manifold.
(g) The unit n-sphere S^n, which is the collection of points in $\mathbb{R}^{n+1}$ satisfying

$$x_1^2 + \cdots + x_{n+1}^2 = 1,$$

is a manifold.

Any surface with sharp kinks, edges, or points cannot be a manifold. Thus, neither a cone nor a finite cylinder is a two-dimensional manifold. However, an infinitely long cylinder is a manifold.

Let U_P denote a neighborhood of P. When we say that this neighborhood looks like an m-dimensional Cartesian space, we mean that there exists a bijective map $\varphi : U_P \to \mathbb{R}^m$ from a neighborhood U_P of P to a neighborhood $\varphi(U_P)$ of $\varphi(P)$ in $\mathbb{R}^m$, such that as we move the point P continuously in U_P, its image moves *continuously* in $\varphi(U_P)$. Since $\varphi(P) \in \mathbb{R}^m$, we can define functions $x^i : U_P \to \mathbb{R}$ such that $\varphi(P) = (x^1(P), x^2(P), \ldots, x^m(P))$. These functions are called **coordinate functions** of φ. The numbers $x^i(P)$ are called **coordinates** of P. The neighborhood U_P together with its mapping φ form a **chart**, denoted by (U_P, φ).

coordinate functions and charts

Now let (V_P, μ) be another chart at P with coordinate functions $\mu(P) = (y^1(P), y^2(P), \ldots, y^m(P))$ (see Fig. 28.1). It is assumed that the map $\mu \circ \varphi^{-1} : \varphi(U_P \cap V_p) \to \mu(U_P \cap V_P)$, which maps a subset of $\mathbb{R}^m$ to another subset of $\mathbb{R}^m$, possesses derivatives of all orders. Then, we say that the two charts μ and φ are $\mathcal{C}^\infty$**-related**. Such a relation underlies the concept of smoothness in the definition of a manifold. A collection of charts that cover the manifold and of which each pair is $\mathcal{C}^\infty$-related is called a $\mathcal{C}^\infty$ **atlas**.

$\mathcal{C}^\infty$-related charts and atlases

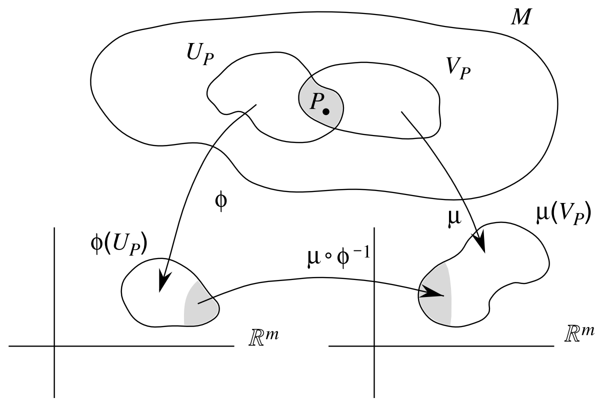

Fig. 28.1 Two charts (U_P, φ) and (V_P, μ), containing P are mapped into $\mathbb{R}^m$. The function $\mu \circ \varphi^{-1}$ is an ordinary function from $\mathbb{R}^m$ to $\mathbb{R}^m$

Example 28.1.3 For the two-dimensional unit sphere S^2 we can construct an atlas as follows. Let $P = (x, y, z)$ be a point in S^2. Then $x^2 + y^2 + z^2 = 1$, or

$$z = \pm\sqrt{1 - x^2 - y^2}.$$

The plus sign corresponds to the upper hemisphere, and the minus sign to the lower hemisphere. Let U_3^+ be the upper hemisphere *with the equator removed*. Then a chart (U_3^+, φ_3) with $\varphi_3 : U_3^+ \to \mathbb{R}^2$ can be constructed by projecting on the xy-plane: $\varphi_3(P) = (x, y)$. Similarly, (U_3^-, μ_3) with $\mu_3 : U_3^- \to \mathbb{R}^2$ given by $\mu_3(P) = (x, y)$ is a chart for the lower hemisphere.

Construction of an atlas for the sphere S^2.

In manifold theory the neighborhoods on which mappings of charts are defined have no boundaries (thus the word "open"). This is because it is more convenient to define limits on boundaryless (open) neighborhoods. Thus, in the above two charts the equator, which is the boundary for both hemispheres, must be excluded. With this exclusion U_3^+ and U_3^- cannot cover the entire S^2; hence, they do not form an atlas. More charts are needed to cover the unit two-sphere. Two such charts are the right and left hemispheres U_2^+ and U_2^-, for which $y > 0$ and $y < 0$, respectively. However, these two neighborhoods leave two points uncovered, the points $(1, 0, 0)$ and $(-1, 0, 0)$. Again this is because boundaries of the right and left hemispheres must be excluded. Adding the front and back hemispheres $U_1^\pm$ to the collection covers these two points. Then S^2 is completely covered and we have an atlas. There is, of course, a lot of overlap among charts. We now show that these overlaps are $\mathcal{C}^\infty$-related.

As an illustration, we consider the overlap between U_3^+ and U_2^+. This is the upper-right quarter of the sphere. Let (U_3^+, φ_3) and (U_2^+, φ_2) be charts with

$$\varphi_3(x, y, z) = (x, y), \qquad \varphi_2(x, y, z) = (x, z).$$

The inverses are therefore given by

$$\varphi_3^{-1}(x, y) = (x, y, z) = \bigl(x, y, \sqrt{1 - x^2 - y^2}\bigr),$$

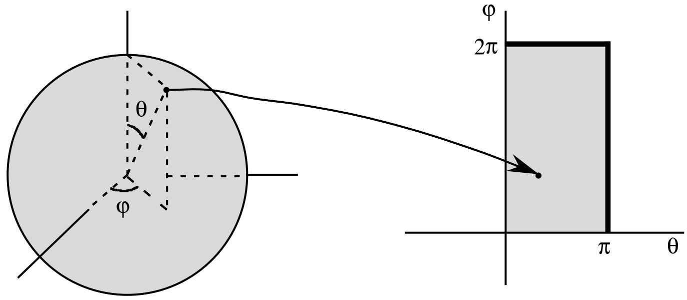

Fig. 28.2 A chart mapping points of S^2 into $\mathbb{R}^2$. Note that the map is not defined for $\theta = 0, \pi$, and therefore at least one more chart is required to cover the whole sphere

$$\varphi_2^{-1}(x, z) = (x, y, z) = \left(x, \sqrt{1 - x^2 - z^2}, z\right),$$

and

$$\varphi_2 \circ \varphi_3^{-1}(x, y) = \varphi_2\left(x, y, \sqrt{1 - x^2 - y^2}\right) = \left(x, \sqrt{1 - x^2 - y^2}\right).$$

Let us denote $\varphi_2 \circ \varphi_3^{-1}$ by F, so that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is described by two functions, the components of F:

$$F_1(x, y) = x \quad \text{and} \quad F_2(x, y) = \sqrt{1 - x^2 - y^2}.$$

The first component has derivatives of all orders at all points. The second component has derivatives of all orders at all points except at $x^2 + y^2 = 1$, which is excluded from the region of overlap of U_3^+ and U_2^+, for which z can never be zero. Thus, F has derivatives of all orders at all points of its domain of definition.

One can similarly show that all regions of overlap for all charts have this property, i.e., all charts are $\mathcal{C}^\infty$-related.

Example 28.1.4 For S^2 of the preceding example, we can find a new atlas in terms of new coordinate functions. Since $x_1^2 + x_2^2 + x_3^2 = 1$, we can use spherical coordinates $\theta = \cos^{-1} x_3$, $\varphi = \tan^{-1}(x_2/x_1)$. A chart is then given by $(S^2 - \{1\} - \{-1\}, \mu)$, where $\mu(P) = (\theta, \varphi)$ maps a point of S^2 onto a region in $\mathbb{R}^2$. This is schematically shown in Fig. 28.2. The singletons $\{1\}$ and $\{-1\}$ are the north and the south poles, respectively.

This chart cannot cover all of S^2, however, because when $\theta = 0$ (or π), the value of the azimuthal angle φ is not determined. In other words, $\theta = 0$ (or π) determines *one* point of the sphere (the north pole or the south pole), but its image in $\mathbb{R}^2$ is the whole range of φ values. Therefore, we must exclude $\theta = 0$ (or π) from the chart (S^2, μ). To cover these two points, we need more charts.

illustration of stereographic projection

Example 28.1.5 A third atlas for S^2 is the so-called **stereographic projection** shown in Fig. 28.3. In such a mapping the image of a point is obtained by drawing a line from the north pole to that point and extending it, if necessary, until it intersects the x_1x_2-plane. It can be verified that the mapping

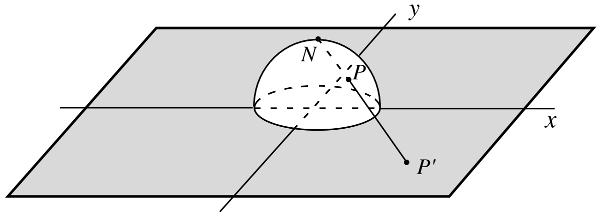

Fig. 28.3 Stereographic projection of S^2 into R^2. Note that the north pole has no image under this map; another chart is needed to cover the whole sphere

$\varphi : S^2 - \{1\} \to \mathbb{R}^2$ is given by

$$\varphi(x_1, x_2, x_3) = \left(\frac{x_1}{1 - x_3}, \frac{x_2}{1 - x_3}\right).$$

We see that this mapping fails for $x_3 = 1$, that is, the north pole. Therefore, the north pole must be excluded (thus, the domain $S^2 - \{1\}$). To cover the north pole we need another stereographic projection—this time from the south pole. Then the two mappings will cover all of S^2, and it can be shown that the two charts are $\mathcal{C}^\infty$-related (see Example 28.1.12).

The three foregoing examples illustrate the following fact, which can be shown to hold rigorously:

Box 28.1.6 *It is impossible to cover the whole S^2 with just one chart.*

Example 28.1.7 Let $\mathcal{V}$ be an m-dimensional real vector space. Fix any basis $\{\mathbf{e}_i\}$ in $\mathcal{V}$ with dual basis $\{\boldsymbol{\epsilon}^i\}$. Define $\phi : \mathcal{V} \to \mathbb{R}^m$ by $\phi(\mathbf{v}) = (\boldsymbol{\epsilon}^1(\mathbf{v}), \dots, \boldsymbol{\epsilon}^m(\mathbf{v}))$. Then the reader may verify that $(\mathcal{V}, \phi)$ is an atlas. Linearity of ϕ ensures that it has derivatives of all orders. This construction shows that $\mathcal{V}$ is a manifold of dimension m.

vector spaces are manifolds

If M and N are manifolds of dimensions m and n, respectively, we can construct their **product manifold** $M \times N$, a manifold of dimension $m + n$. A typical chart on $M \times N$ is obtained from charts on M and N as follows. Let (U, φ) be a chart on M and (V, μ) one on N. Then a chart on $M \times N$ is $(U \times V, \varphi \times \mu)$ where

product manifold defined

$$\varphi \times \mu(P, Q) = \big(\varphi(P), \mu(Q)\big) \in \mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n} \quad \text{for } P \in U,\ Q \in V.$$

Definition 28.1.8 Let M be a manifold. A subset N of M is called a **submanifold** of M if N is a manifold in its own right.

submanifold

A trivial, but important, example of submanifolds is the so-called **open submanifold**. If M is a manifold and U is an open subset[1] of M, then U

open submanifolds

[1] Recall that an open subset U is one each of whose points is the center of an open ball lying entirely in U.

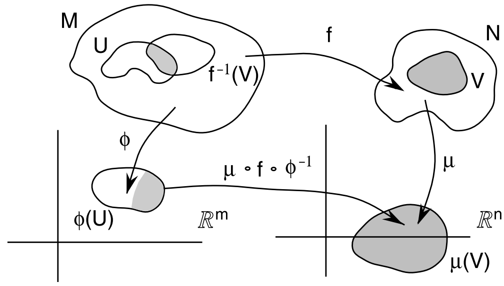

Fig. 28.4 Corresponding to every map $f : M \to N$ there exists a coordinate map $\mu \circ f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$

inherits a manifold structure from M by taking any chart $(U_\alpha, \varphi_\alpha)$ and restricting φ_α to $U \cap U_\alpha$. It is clear that $\dim U = \dim M$.

Having gained familiarity with manifolds, it is now appropriate to consider maps between them that are compatible with their structure.

differentiable maps and their coordinate expressions

Definition 28.1.9 Let M and N be manifolds of dimensions m and n, respectively. Let $f : M \to N$ be a map. We say that f is $\mathcal{C}^\infty$, or **differentiable**, if for every chart (U, φ) in M and every chart (V, μ) in N, the composite map $\mu \circ f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$, called the **coordinate expression for** f, is $\mathcal{C}^\infty$ wherever it is defined.[2]

function as a special kind of map

The content of this definition is illustrated in Fig. 28.4. A particularly important special case occurs when $N = \mathbb{R}$; then we call f a (real-valued) **function**. The collection of all $\mathcal{C}^\infty$ functions at a point $P \in M$ is denoted by $F^\infty(P)$: If $f \in F^\infty(P)$, then $f : U_P \to \mathbb{R}$ is $\mathcal{C}^\infty$ for some neighborhood U_P of P.

Let $f : M \to N$ be a differentiable map. Then f is automatically continuous. Now let V be an open subset of N. The set $f^{-1}(V)$ is an open subset of M by Proposition 17.4.6.[3]

Proposition 28.1.10 *Let M be an m-dimensional manifold, $f : M \to N$ a differentiable map, and V an open subset of N. Then $f^{-1}(V)$, the set of points of M mapped onto V, is an open m-dimensional submanifold of M.*

Just as the concept of isomorphism identified all vector spaces, algebras, and groups that were equivalent to one another, it is desirable to introduce a notion that brings together those manifolds that "look alike".

---

[2]The domain of $\mu \circ f \circ \varphi^{-1}$ is not all of $\mathbb{R}^m$, but only its open subset $\varphi(U)$. However, we shall continue to abuse the notation and write $\mathbb{R}^m$ instead of $\varphi(U)$. This way, we do not have to constantly change the domain as U changes. The domain is always clear from the context.

[3]Although Proposition 17.4.6 was shown for normed linear spaces, it really holds for all "spaces" for which the concept of open set is defined.

Definition 28.1.11 A bijective differentiable map whose inverse is also differentiable is called a **diffeomorphism**. Two manifolds between which a diffeomorphism exists are called **diffeomorphic**. Let M and N be manifolds. M is said to be **diffeomorphic to** N **at** $P \in M$ if there is a neighborhood U of P and a diffeomorphism $f : U \to f(U)$. Then f is called a **local diffeomorphism at** P.

diffeomorphism and local diffeomorphism defined

In our discussion of groups, we saw that the set of linear isomorphisms of a vector space $\mathcal{V}$ onto itself forms a group $GL(\mathcal{V})$. The set of diffeomorphisms of a manifold M onto itself also forms a group, which is denoted by $\text{Diff}(M)$.

diffeomorphisms of a manifold form a group

Example 28.1.12 The generalization of a sphere is the unit n-**sphere**, which is a subset of $\mathbb{R}^{n+1}$ defined by

n-sphere and its stereographic projection

$$S^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\}.$$

The stereographic projection defines an atlas for S^n as follows. For all points of S^n except $(0, 0, \dots, 1)$, the north pole, define the chart $\varphi_+ : S^n - \{1\} \equiv U^+ \to \mathbb{R}^n$ by

$$\varphi_+(x_1, \dots, x_{n+1}) = \left(\frac{x_1}{1 - x_{n+1}}, \dots, \frac{x_n}{1 - x_{n+1}}\right) \quad \text{for } (x_1, \dots, x_{n+1}) \in U^+.$$

To include the north pole, consider a second chart $\varphi_- : S^n - \{-1\} \equiv U^- \to \mathbb{R}^n$ defined by

$$\varphi_-(x_1, \dots, x_{n+1}) = \left(\frac{x_1}{1 + x_{n+1}}, \dots, \frac{x_n}{1 + x_{n+1}}\right) \quad \text{for } (x_1, \dots, x_{n+1}) \in U^-.$$

Next, let us find the inverses of these maps. We find the inverse of φ_+; that of φ_- can be found similarly. Let $\xi_k \equiv x_k/(1 - x_{n+1})$. Then one can readily show that

$$\sum_{k=1}^{n} \xi_k^2 = \frac{1 + x_{n+1}}{1 - x_{n+1}} \quad \Rightarrow \quad x_{n+1} = \frac{\sum_{k=1}^n \xi_k^2 - 1}{\sum_{k=1}^n \xi_k^2 + 1}$$

and

$$x_i = \frac{2\xi_i}{1 + \sum_{k=1}^n \xi_k^2} \quad \text{for } i = 1, 2, \dots, n.$$

From the definition of φ_+, we have

$$\varphi_+^{-1}(\xi_1, \dots, \xi_n) = (x_1, \dots, x_n, x_{n+1}) = \left(\frac{2\xi_1}{1 + \sum_{k=1}^n \xi_k^2}, \dots, \frac{2\xi_n}{1 + \sum_{k=1}^n \xi_k^2}, \frac{\sum_{k=1}^n \xi_k^2 - 1}{\sum_{k=1}^n \xi_k^2 + 1}\right). \tag{28.1}$$

On the overlap of U^+ and U^-, i.e., on all points of S^n except the north and the south poles, $\varphi_- \circ \varphi_+^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ can be calculated by noting that φ_- has the following effect on a typical entry of Eq. (28.1):

$$x_j \mapsto \frac{x_j}{1 + x_{n+1}} = \frac{\frac{2\xi_j}{1+\sum_{k=1}^n \xi_k^2}}{1 + \frac{\sum_{k=1}^n \xi_k^2 - 1}{\sum_{k=1}^n \xi_k^2 + 1}} = \frac{\xi_j}{\sum_{k=1}^n \xi_k^2}.$$

Therefore,

$$\varphi_- \circ \varphi_+^{-1}(\xi_1, \dots, \xi_n) = \left(\frac{\xi_1}{\sum_{k=1}^n \xi_k^2}, \dots, \frac{\xi_n}{\sum_{k=1}^n \xi_k^2}\right).$$

It is clear that $\varphi_- \circ \varphi_+^{-1}$ has derivatives of all orders except possibly at a point for which $\xi_i = 0$ for all i. But this would correspond to $x_{n+1} = 1$, which is excluded from the region of overlap.

28.2 Curves and Tangent Vectors

We noted above that *functions* are special cases of Definition 28.1.9. Another special case occurs when $M = \mathbb{R}$. This is important enough to warrant a separate definition.

differentiable curve

Definition 28.2.1 A **differentiable curve** in the manifold M is a $\mathcal{C}^\infty$ map of an interval of $\mathbb{R}$ to M.

initial and final points of a curve

This definition should be familiar from calculus, where $M = \mathbb{R}^3$ and a curve is given by its *parametric equation* $(f_1(t), f_2(t), f_3(t))$, or simply by $\mathbf{r}(t)$. The point $\gamma(a) \in M$ is called the **initial point**, and $\gamma(b) \in M$ is called the **final point** of the curve γ. A curve is closed if $\gamma(a) = \gamma(b)$.

We are now ready to consider what a vector at a point is. All the familiar vectors in classical physics, such as displacement, velocity, momentum, and so forth, are based on the displacement vector. Let us see how we can generalize such a vector so that it is compatible with the concept of a manifold.

In $\mathbb{R}^2$, we define the displacement vector from P to Q as a directed straight line that starts at P and ends at Q. Furthermore, the direction of the vector remains the same if we connect P to any other final point on the line PQ located beyond Q. This is because $\mathbb{R}^2$ is a flat space, a straight line is well-defined, and there is no ambiguity in the direction of the vector from P to Q.

Things change, however, if we move to a two-dimensional spherical surface such as the globe. How do we define the straight line from New York to Beijing? There is no satisfactory definition of the word "straight" on a curved surface. Let us say that "straight" means shortest distance. Then our shortest path would lie on a great circle passing through New York and Beijing. Define the "direction" of the trip as the "straight" arrow, say 1 km in length, connecting our present position to the next point 1 km away. As we move from New York to Beijing, going westward, the tip of the arrow keeps changing direction. Its direction in New York is slightly different from its direction in Chicago. In San Francisco the direction is changed even more, and by the time we reach Beijing, the tip of the arrow will be almost opposite to its original direction.

The reason for such a changing arrow is, of course, the curvature of the manifold. We can minimize this curvature effect if we do not go too far from New York. If we stay close to New York, the surface of the earth appears flat, and we can draw arrows between points. The closer the two points, the better the approximation to flatness. Clearly, the concept of a vector is a *local* concept, and the process of constructing a vector is a *limiting* process.

The limiting process in the globe example entailed the notions of "closeness". Such a notion requires the concept of distance, which is natural for a globe but not necessary for a general manifold. For most manifolds it is possible to define a metric that gives the "distance" between two points of the manifold. However, the concept of a vector is too general to require such an elaborate structure as a metric. The abstract usefulness of a metric is a result of its real-valuedness: given two points P_1 and P_2, the distance between them, $d(P_1, P_2)$, is a nonnegative real number. Thus, distances between different points can be compared.

We have already defined two concepts for manifolds (more basic than the concept of a metric) that together can replace the concept of a metric in defining a vector as a limit. These are the concepts of (real-valued) functions and curves. Let us see how functions and curves can replace metrics.

Let $\gamma : [a, b] \to M$ be a curve in the manifold M. Let $P \in M$ be a point of M that lies on γ such that $\gamma(c) = P$ for some $c \in [a, b]$. Let $f \in F^{\infty}(P)$. Restrict f to the neighboring points of P that lie on γ. Then the composite function $f \circ \gamma : \mathbb{R} \to \mathbb{R}$ is a real-valued function on $\mathbb{R}$.

We can compare values of $f \circ \gamma$ for various real numbers close to c—as in calculus. If $u \in [a, b]$ denotes[4] the variable, then $f \circ \gamma(u) = f(\gamma(u))$ gives the value of $f \circ \gamma$ at various u's. In particular, the difference $\Delta(f \circ \gamma) \equiv f(\gamma(u)) - f(\gamma(c))$ is a measure of how close the point $\gamma(u) \in M$ is to P. Going one step further, we define

$$\left. \frac{d(f \circ \gamma)}{du} \right|_{u=c} = \lim_{u \to c} \frac{f(\gamma(u)) - f(\gamma(c))}{u - c}, \tag{28.2}$$

the usual derivative of an ordinary function of one variable. However, this derivative depends on γ and on the point P. The function f is merely a *test function*. We could choose any other function to test how things change with movement along γ. What is important is not which function we choose, but how the curve γ causes it to change with movement along γ away from P. This change is determined by the directional derivative along γ at P, as given by (28.2). A directional derivative determines a tangent which, in turn, suggests a tangent *vector*. That is why the tangent vector at P along γ is defined to be the directional derivative itself!

[4] We usually use u or t to denote the (real) argument of the map $\gamma : [a, b] \to M$.

The use of derivative as tangent vector may appear strange to the novice, especially physicists encountering it for the first time, but it has been familiar to mathematicians for a long time. It is hard for the beginner to imagine vectors being charged with the responsibility of measuring the rate of change of functions. It takes some mental adjustment to get used to this idea. The following simple illustration may help with establishing the vector-derivative connection.

illustration of the equality of vectors and directional derivatives

Example 28.2.2 Let us take the familiar case of a plane and consider the vector $\mathbf{a} = a_x\hat{\mathbf{e}}_x + a_y\hat{\mathbf{e}}_y$. What kind of a directional derivative can correspond to $\mathbf{a}$? First we need a curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ that is somehow associated with $\mathbf{a}$. It is not hard to convince oneself that the most natural association is that of vectors to tangents. Thus, we seek a curve whose tangent is (parallel to) $\mathbf{a}$. The easiest (but not the only) way is simply to take the straight line along $\mathbf{a}$; that is, let $\gamma(u) = (a_x u, a_y u)$. The directional derivative at $u = 0$ for an arbitrary function $f : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\left.\frac{d(f \circ \gamma)}{du}\right|_{u=0} = \lim_{u\to 0} \frac{f(\gamma(u)) - f(\gamma(0))}{u} = \lim_{u\to 0} \frac{f(a_x u, a_y u) - f(0,0)}{u}. \tag{28.3}$$

Taylor expansion in two dimensions yields

$$f(a_x u, a_y u) = f(0,0) + a_x u \left.\frac{\partial f}{\partial x}\right|_{u=0} + a_y u \left.\frac{\partial f}{\partial y}\right|_{u=0} + \cdots .$$

Substituting in (28.3), we obtain

$$\begin{aligned}\left.\frac{d(f \circ \gamma)}{du}\right|_{u=0} &= \lim_{u\to 0} \frac{a_x u(\partial f/\partial x)_{u=0} + a_y u(\partial f/\partial y)_{u=0} + \cdots}{u} \\ &= a_x \frac{\partial f}{\partial x} + a_y \frac{\partial f}{\partial y} = \left(a_x \frac{\partial}{\partial x} + a_y \frac{\partial}{\partial y}\right) f.\end{aligned}$$

This clearly shows the connection between directional derivatives and vectors. In fact, the correspondences $\partial/\partial x \leftrightarrow \hat{\mathbf{e}}_x$ and $\partial/\partial y \leftrightarrow \hat{\mathbf{e}}_y$ establish this connection very naturally.

Note that the curve γ chosen above is by no means unique. In fact, there are infinitely many curves that have the same tangent at $u = 0$ and give the same directional derivative.

Since vectors are the same as derivatives, we expect them to have the properties shared by derivatives:

tangent vector defined

Definition 28.2.3 Let M be a differentiable manifold. A **tangent vector** at $P \in M$ is an operator $\mathbf{t} : F^\infty(P) \to \mathbb{R}$ such that for every $f, g \in F^\infty(P)$ and $\alpha, \beta \in \mathbb{R}$

1. $\mathbf{t}$ is linear: $\mathbf{t}(\alpha f + \beta g) = \alpha\mathbf{t}(f) + \beta\mathbf{t}(g)$;
2. $\mathbf{t}$ satisfies the **derivation property**:

derivation property of tangent vectors

$$\mathbf{t}(fg) = g(P)\mathbf{t}(f) + f(P)\mathbf{t}(g).$$

The operator $\mathbf{t}$ is an abstraction of the derivative operator. Note that $\mathbf{t}(f)$, $g(P)$, $f(P)$, and $\mathbf{t}(g)$ are all real *numbers*.

The reader may easily check that if addition and scalar multiplication of tangent vectors are defined in an obvious way, the set of all tangent vectors at $P \in M$ becomes a vector space, called the **tangent space** at P and denoted by $\mathcal{T}_P(M)$. If U is an open subset of M (therefore, an open submanifold of M), then it is clear that

tangent space defined

$$\mathcal{T}_P(U) = \mathcal{T}_P(M) \quad \text{for all } P \in U. \tag{28.4}$$

Definition 28.2.3 was motivated by Eqs. (28.2) and (28.3). Let us go backwards and see if (28.2) is indeed a tangent, that is, if it satisfies the two conditions of Definition 28.2.3.

Proposition 28.2.4 *Let γ be a $\mathcal{C}^\infty$ curve in M such that $\gamma(c) = P$. Define $\vec{\boldsymbol{\gamma}}(c) : F^\infty(P) \to \mathbb{R}$ by*

vectors tangent to a curve

$$\big(\vec{\boldsymbol{\gamma}}(c)\big)(f) \equiv \frac{d}{du} f \circ \gamma \bigg|_{u=c}, \qquad f \in F^\infty(P).$$

*Then $\vec{\boldsymbol{\gamma}}(c)$ is a tangent vector at P called the vector **tangent to** γ **at** c.*

Proof We have to show that the two conditions of Definition 28.2.3 are satisfied for $f, g \in F^\infty(P)$ and $\alpha, \beta \in \mathbb{R}$. The first condition is trivial. For the second condition, we use the product rule for ordinary differentiation as follows:

$$\begin{aligned}
\big(\vec{\boldsymbol{\gamma}}(c)\big)(fg) &= \frac{d}{du}(fg) \circ \gamma \bigg|_{u=c} \equiv \frac{d}{du}\big[(f \circ \gamma)(g \circ \gamma)\big]\bigg|_{u=c} \\
&= \left[\frac{d}{du}(f \circ \gamma)\bigg|_{u=c}\right](g \circ \gamma)_{u=c} + (f \circ \gamma)_{u=c}\left[\frac{d}{du}(g \circ \gamma)\bigg|_{u=c}\right] \\
&= \big[\big(\vec{\boldsymbol{\gamma}}(c)\big)(f)\big] g\big(\gamma(c)\big) + f\big(\gamma(c)\big)\big[\big(\vec{\boldsymbol{\gamma}}(c)\big)(g)\big] \\
&= \big[\big(\vec{\boldsymbol{\gamma}}(c)\big)(f)\big] g(P) + f(P)\big[\big(\vec{\boldsymbol{\gamma}}(c)\big)(g)\big].
\end{aligned}$$

Note that in going from the first equality to the second, we used the fact that by definition, the product of two functions evaluated at a point is the product of the values of the two functions at that point. □

Let us now consider a special curve and corresponding tangent vector that is of extreme importance in applications. Let $\varphi = (x^1, x^2, \dots, x^m)$ be a coordinate system at P, where $x^i : M \to \mathbb{R}$ is the ith coordinate function. Then φ is a bijective $\mathcal{C}^\infty$ mapping from the manifold M into $\mathbb{R}^m$. Its inverse, $\varphi^{-1} : \mathbb{R}^m \to M$, is also a $\mathcal{C}^\infty$ mapping. Now, the ith coordinate of P is the real number $u \equiv x^i(P)$. Suppose that all coordinates of P are held fixed except the ith one, which is allowed to vary with u describing this variation.

Definition 28.2.5 Let (U_P, φ) be a chart at $P \in M$. Then the curve $\gamma^i : \mathbb{R} \to M$, defined by

$$\gamma^i(u) = \varphi^{-1}\big(x^1(P), \dots, x^{i-1}(P), u, x^{i+1}(P), \dots, x^m(P)\big)$$

coordinate curve, coordinate vector field, and coordinate frames

is called the ith **coordinate curve** through P. The tangent vector to this curve at P is denoted by $\partial_i|_P$ and is called the ith **coordinate vector field** at P. The collection of all vector fields at P is called a **coordinate frame** at P. The variable u is arbitrary in the sense that it can be replaced by any (good) function of u.

Let $c = x^i(P)$. Then for $f \in F^\infty(P)$, we have

$$\begin{aligned}(\partial_i|_P)f &= \big(\vec{\boldsymbol{\gamma}}_i(c)\big)(f) = \frac{d}{du} f \circ \gamma^i\bigg|_{u=c}\\ &= \frac{d}{du} f\big(\varphi^{-1}\big(x^1(P), \ldots, x^{i-1}(P), u, x^{i+1}(P), \ldots, x^m(P)\big)\big)\bigg|_{u=c}\\ &\equiv \frac{\partial f}{\partial x^i}\bigg|_P \quad \Rightarrow \quad \partial_i|_P = \frac{\partial}{\partial x^i}\bigg|_P,\end{aligned} \tag{28.5}$$

where the last equality is a (natural) *definition* of the partial derivative of f with respect to the ith coordinate evaluated at the point P. This partial derivative is again a $\mathcal{C}^\infty$ function at P. We therefore have the following:

Proposition 28.2.6 *The coordinate frame* $\{\partial_i|_P\}_{i=1}^m$ *at* P *is a set of operators* $\partial_i(P) : F^\infty(P) \to \mathbb{R}$ *given by*

$$\begin{aligned}(\partial_i|_P)f &= \frac{\partial f}{\partial x^i}\bigg|_P\\ &\equiv \frac{d}{du} f\big(\varphi^{-1}\big(x^1(P), \ldots, x^{i-1}(P), u, x^{i+1}(P), \ldots, x^m(P)\big)\big)\bigg|_{u=c}.\end{aligned} \tag{28.6}$$

Another common notation for $\partial f/\partial x^i$ is $f_{,i}$

Example 28.2.7 Pick a point $P = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ on the sphere S^2 in a chart (U_P, μ) given by $\mu(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) = (\theta, \varphi)$. If θ is kept constant and φ is allowed to vary over values given by u, then the coordinate curve associated with φ is given by

$$\gamma_\varphi(u) = \mu^{-1}(\theta, u) = (\sin\theta\cos u, \sin\theta\sin u, \cos\theta).$$

As u varies, $\gamma_\varphi(u)$ describes a curve on S^2. This curve is simply a circle of radius $\sin\theta$. The tangent to this curve at any point is $\partial/\partial\varphi$, or simply ∂_φ, the derivative with respect to the coordinate φ.

Similarly, the curve $\gamma_\theta(u)$ describes a great circle on S^2 with tangent $\partial_\theta \equiv \partial/\partial\theta$.

The vector space $\mathcal{T}_P(M)$ of all tangents at P was mentioned earlier. In the case of S^2 this tangent space is simply a plane tangent to the sphere at a point. Also, the two vectors, ∂_θ and ∂_φ encountered in Example 28.2.7 are clearly linearly independent. Thus, they form a basis for the tangent plane. This argument can be generalized to any manifold. The following theorem is such a generalization (for a proof, see [Bish 80, pp. 51–53]):

Theorem 28.2.8 *Let M be an m-dimensional manifold and $P \in M$. Then the set $\{\partial_i|_P\}_{i=1}^m$ forms a basis of $\mathcal{T}_P(M)$. In particular, $\mathcal{T}_P(M)$ is m-dimensional. An arbitrary vector, $\mathbf{t} \in \mathcal{T}_P(M)$, can be written as*

$$\mathbf{t} = \alpha^i \partial_i|_P, \quad \textit{where } \alpha^i = \mathbf{t}(x^i).$$

Remember Einstein's summation convention!

The last statement can be derived by letting both sides operate on x^j and using Eq. (28.6). Let $M = \mathcal{V}$, a vector space. Choose a basis $\{\mathbf{e}_i\}$ in $\mathcal{V}$ with its dual considered as coordinate functions. Then, at every $\mathbf{v} \in \mathcal{V}$, there is a *natural* isomorphism $\phi : \mathcal{V} \to \mathcal{T}_\mathbf{v}(\mathcal{V})$ mapping a vector $\mathbf{u} = \alpha^i \mathbf{e}_i \in \mathcal{V}$ onto $\alpha^i \partial_i|_\mathbf{v} \in \mathcal{T}_\mathbf{v}(\mathcal{V})$. The reader may verify that this isomorphism is coordinate independent; i.e., if one chooses any other basis of $\mathcal{V}$ with its corresponding dual, then $\phi(\mathbf{v})$ will be the same vector as before, expressed in the new coordinate basis. Thus,

Box 28.2.9 *If $\mathcal{V}$ is a vector space, then for all $\mathbf{v} \in \mathcal{V}$, one can identify $\mathcal{T}_\mathbf{v}(\mathcal{V})$ with $\mathcal{V}$ itself.*

Suppose we have two coordinate systems at P, $\{x^i\}$ with tangents $\partial_i|_P$ and $\{y^j\}$ with tangents $\nabla_j|_P$. Any $\mathbf{t} \in \mathcal{T}_P(M)$ can be expressed either in terms of $\partial_i|_P$ or in terms of $\nabla_j|_P$: $\mathbf{t} = \alpha^i \partial_i|_P = \beta^j \nabla_j|_P$. We can use this relation to obtain α^i in terms of β^j: From Theorem 28.2.8, we have

$$\alpha^i = \mathbf{t}(x^i) = (\beta^j \nabla_j|_P)(x^i) \equiv \left[\beta^j \frac{\partial}{\partial y^j}\bigg|_P\right](x^i) = \beta^j \frac{\partial x^i}{\partial y^j}\bigg|_P. \tag{28.7}$$

In particular, if $\mathbf{t} = \nabla_k|_P$, then $\beta^j = \mathbf{t}(y^j) = [\nabla_k|_P](y^j) = \delta_k^j$, and (28.7) gives $\alpha^i = \partial x^i / \partial y^k$. Thus, using Eq. (28.5),

$$\frac{\partial}{\partial y^j}\bigg|_P = \frac{\partial x^i}{\partial y^j} \frac{\partial}{\partial x^i}\bigg|_P. \tag{28.8}$$

For any function $f \in F^\infty(P)$, Eq. (28.8) yields

$$\left[\frac{\partial}{\partial y^j}\bigg|_P\right] f = \frac{\partial f}{\partial y^j}\bigg|_P = \frac{\partial x^i}{\partial y^j}\bigg|_P \left[\frac{\partial}{\partial x^i}\bigg|_P\right] f = \frac{\partial x^i}{\partial y^j}\bigg|_P \frac{\partial f}{\partial x^i}\bigg|_P.$$

This is the chain rule for differentiation.

Example 28.2.10 Let us find the coordinate curves and the coordinate frame at $P = (x, y, z)$ on S^2. We use the coordinates of Example 28.1.3. In particular, consider φ_3, whose inverse is given by

$$\varphi_3^{-1}(x, y) = \left(x, y, \sqrt{1 - x^2 - y^2}\right).$$

The coordinate curve $\gamma_2(u)$ along y is obtained by letting y be a function[5] of u:

$$\gamma_2(u) = \varphi_3^{-1}\big(x, h(u)\big) = \Big(x, h(u), \sqrt{1-x^2-h^2(u)}\Big),$$

where $h(0) = y$ and $h'(0) = \alpha$, a constant. To find the coordinate vector field at P, let $f \in F^\infty(P)$, and note that

$$\begin{aligned}\partial_2 f &= \frac{d}{du} f\big(\gamma_2(u)\big)\bigg|_{u=0} = \frac{d}{du} f\Big(x, h(u), \sqrt{1-x^2-h^2(u)}\Big)\bigg|_{u=0} \\ &= \frac{\partial f}{\partial y}\frac{dh}{du}\bigg|_{u=0} + \frac{\partial f}{\partial z}\bigg[\frac{1}{2}\big(-2h(u)\big)\frac{dh}{du}\frac{1}{\sqrt{1-x^2-h^2(u)}}\bigg]_{u=0} \\ &= \alpha\bigg(\frac{\partial f}{\partial y} - \frac{y}{z}\frac{\partial f}{\partial z}\bigg) = \alpha\bigg(\frac{\partial}{\partial y} - \frac{y}{z}\frac{\partial}{\partial z}\bigg) f.\end{aligned}$$

So, choosing the function h in such a way that $\alpha = 1$,

$$\partial_2 = \partial_y - \frac{y}{z}\partial_z,$$

where ∂_y and ∂_z are the coordinate vector fields of $\mathbb{R}^3$. The coordinate vector field ∂_1 can be obtained similarly.

28.3 Differential of a Map

Now that we have constructed tangent spaces and defined bases for them, we are ready to consider the notion of the differential (derivative) of a map between manifolds.

differential of a map at a point

Definition 28.3.1 Let M and N be manifolds of dimensions m and n, respectively, and let $\psi : M \to N$ be a $\mathcal{C}^\infty$ map. Let $P \in M$, and let $Q = \psi(P) \in N$ be the image of P. Then there is induced a map $\psi_{*P} : \mathcal{T}_P(M) \to \mathcal{T}_Q(N)$, called the **differential of** ψ **at** P and given as follows. Let $\mathbf{t} \in \mathcal{T}_P(M)$ and $f \in F^\infty(Q)$. The action of $\psi_{*P}(\mathbf{t}) \in \mathcal{T}_Q(N)$ on f is defined as

$$\big(\psi_{*P}(\mathbf{t})\big)(f) \equiv \mathbf{t}(f \circ \psi). \tag{28.9}$$

The reader may check that the differential of a composite map is the composite of the corresponding differentials, i.e.,

$$(\psi \circ \phi)_{*P} = \psi_{*\phi(P)} \circ \phi_{*P}. \tag{28.10}$$

Furthermore, if ψ is a local diffeomorphism at P, then ψ_{*P} is a vector space isomorphism. The inverse of this statement—which is called the **inverse mapping theorem**, and is much harder to prove (see [Abra 88, pp. 116 and 196])—is also true:

[5] See the last statement of Definition 28.2.5.

Theorem 28.3.2 (Inverse mapping theorem) *If $\psi : M \to N$ is a map and $\psi_{*P} : \mathcal{T}_P(M) \to \mathcal{T}_{\psi(P)}(N)$ is a vector space isomorphism, then ψ is a local diffeomorphism at P.*

inverse mapping theorem

Let us see how Eq. (28.9) looks in terms of coordinate functions. Suppose that $\{x^i\}_{i=1}^m$ are coordinates at P and $\{y^a\}_{a=1}^n$ are coordinates at $Q = \psi(P)$. We note that $y^a \circ \psi$ is a real-valued $\mathcal{C}^\infty$ function on M. Thus, we may write (with the function expressed in terms of coordinates)

$$y^a \circ \psi \equiv f^a\bigl(x^1, \dots, x^m\bigr).$$

We also have $\mathbf{t} = \alpha^i \partial_i|_P$. Similarly, $\psi_{*P}(\mathbf{t}) = \beta^a (\partial/\partial y^a)|_Q$ because $\{(\partial/\partial y^a)|_Q\}$ form a basis. Theorem 28.2.8 and Definition 28.3.1 now give

$$\begin{aligned}\beta^a &= \psi_{*P}(\mathbf{t})\bigl(y^a\bigr) = \mathbf{t}\bigl(y^a \circ \psi\bigr) = \mathbf{t}\bigl(f^a\bigr) \\ &= \bigl[\alpha^i \partial_i|_P\bigr]\bigl(f^a\bigr) = \alpha^i \frac{\partial f^a}{\partial x^i}\bigg|_P \equiv \sum_{i=1}^m \alpha^i \frac{\partial f^a}{\partial x^i}\bigg|_P .\end{aligned}$$

This can be written in matrix form as

$$\begin{pmatrix}\beta^1\\ \beta^2\\ \vdots\\ \beta^n\end{pmatrix} = \begin{pmatrix}\partial f^1/\partial x^1 & \partial f^1/\partial x^2 & \dots & \partial f^1/\partial x^m\\ \partial f^2/\partial x^1 & \partial f^2/\partial x^2 & \dots & \partial f^2/\partial x^m\\ \vdots & \vdots & & \vdots\\ \partial f^n/\partial x^1 & \partial f^n/\partial x^2 & \dots & \partial f^n/\partial x^m\end{pmatrix}\begin{pmatrix}\alpha^1\\ \alpha^2\\ \vdots\\ \alpha^m\end{pmatrix}. \tag{28.11}$$

The $n \times m$ matrix is denoted by J and is called the **Jacobian matrix** of ψ with respect to the coordinates x^i and y^a. On numerous occasions the two manifolds are simply Cartesian spaces, so that $\psi : \mathbb{R}^m \to \mathbb{R}^n$. In such a case, f^α is naturally written as ψ^α, and the Jacobian matrix will have elements of the form $\partial\psi^\alpha/\partial x^i$.

Jacobian matrix of a differentiable map

An important special case of the differential of a map is that of a constant map. Let $\psi : M \to \{Q\} \in N$ be such a map; it maps all points of M onto a single point Q of N. For any $f \in F^\infty(Q)$, the function $f \circ \psi \in F^\infty(P)$ is constant for all $P \in M$. Let $\mathbf{t} \in \mathcal{T}_P(M)$ be an arbitrary vector. Then

$$\bigl(\psi_{*P}(\mathbf{t})\bigr)(f) \equiv \mathbf{t}(f \circ \psi) = 0 \quad \forall f \Rightarrow \psi_{*P}(\mathbf{t}) = 0 \quad \forall \mathbf{t} \tag{28.12}$$

because $\mathbf{t}(c) = 0$ for any constant c. So,

Differential of a constant map is the zero map.

Box 28.3.3 *If $\psi : M \to \{Q\} \in N$ is a constant map, so that it maps the entire manifold M onto a point Q of N, then $\psi_{*P} : \mathcal{T}_P(M) \to \mathcal{T}_Q(N)$ is the zero map.*

Two other special cases merit closer attention: $M = \mathbb{R}$ for arbitrary N, and $N = \mathbb{R}$ for arbitrary M. In either case $\mathcal{T}_c(\mathbb{R})$ is one-dimensional with the basis vector $(d/du)|_c$. When $M = \mathbb{R}$, the mapping becomes a curve, $\gamma : \mathbb{R} \to N$. The only vector whose image we are interested in is $\mathbf{t} = (d/du)|_c$, with $\gamma(c) = P$. From (28.9) using Proposition 28.2.4 in the last step, we have

$$\left[\gamma_{*c} \frac{d}{du}\bigg|_c\right] f = \frac{d}{du} f \circ \gamma\bigg|_{u=c} = \big(\vec{\boldsymbol{\gamma}}(c)\big)(f).$$

This tells us that the differential of a curve at c is simply its tangent vector at $\gamma(c)$. It is common to leave out the constant vector $(d/du)|_c$, and write γ_{*c} for the LHS.

components of tangents to curves

Example 28.3.4 It is useful to have an expression for the components of the tangent to a curve γ at an arbitrary point on it. Since γ maps the real line to M, with a coordinate patch established on M, we can write γ as $\gamma = (\gamma^1, \dots, \gamma^m)$ where $\gamma^i = x^i \circ \gamma$ are ordinary functions of one variable. Proposition 28.2.4 then yields

$$\begin{aligned}\gamma_{*t} f &= \frac{d}{du} f \circ \gamma\bigg|_{u=t} = \frac{d}{du} f\big(\gamma(u)\big)\bigg|_{u=t} = \frac{d}{du} f\big(\gamma^1(u), \dots, \gamma^m(u)\big)\bigg|_{u=t} \\ &= \frac{\partial f}{\partial x^i}\frac{d\gamma^i}{du}\bigg|_{u=t} \equiv \frac{\partial f}{\partial x^i}\frac{d\gamma^i}{dt} = \dot{\gamma}^i \partial_i f,\end{aligned}$$

or

$$\boxed{\gamma_{*t} \equiv \dot{\gamma}^i \partial_i, \quad \text{where } \dot{\gamma}^i = \frac{d\gamma^i}{dt}.} \tag{28.13}$$

For this reason, γ_{*t} is sometimes denoted by $\dot{\gamma}$.

When $N = \mathbb{R}$, we are dealing with a real-valued function $f : M \to \mathbb{R}$. The differential of f at P is $f_{*P} : \mathcal{T}_P(M) \to \mathcal{T}_c(\mathbb{R})$, where $c = f(P)$. Since $\mathcal{T}_c(\mathbb{R})$ is one-dimensional, for a tangent $\mathbf{t} \in \mathcal{T}_P(M)$, we have $f_{*P}(\mathbf{t}) = a(d/du)|_c$. Let $g : \mathbb{R} \to \mathbb{R}$ be an arbitrary function on $\mathbb{R}$. Then $[f_{*P}(\mathbf{t})](g) = a(dg/du)_c$, or, by definition of the LHS, $\mathbf{t}(g \circ f) = a(dg/du)_c$. To find a we choose the function $g(u) = u$, i.e., the identity function; then $dg/du = 1$ and $\mathbf{t}(g \circ f) = \mathbf{t}(f) = a$. We thus obtain $f_{*P}(\mathbf{t}) = \mathbf{t}(f)(d/du)|_c$. Since $\mathcal{T}_c(\mathbb{R})$ is a flat one-dimensional vector space, all vectors are the same and there is no need to write $(d/du)|_c$. Thus, we define the **differential of** f, denoted by $df \equiv f_*$, as a map $df : \mathcal{T}_P(M) \to \mathbb{R}$ given by

differential of a real-valued function

$$df(\mathbf{t}) = \mathbf{t}(f). \tag{28.14}$$

In particular, if f is the coordinate function x^i and $\mathbf{t}$ is the tangent to the jth coordinate curve $\partial_j|_P$, we obtain

$$dx^i\big|_P(\partial_j|_P) = [\partial_j|_P]\big(x^i\big) = \frac{\partial x^i}{\partial x^j}\bigg|_P = \delta^i_j. \tag{28.15}$$

This shows that

Box 28.3.5 *$\{dx^i|_P\}_{i=1}^m$ is dual to the basis $\{\partial_j|_P\}_{j=1}^m$ of $\mathcal{T}_P(M)$.*

Example 28.3.6 Let $f : M \to \mathbb{R}$ be a real-valued function on M. Let x^i be coordinates at P. We want to express df in terms of coordinate functions. For $\mathbf{t} \in \mathcal{T}_P(M)$ we can write $\mathbf{t} = \alpha^i \partial_i|_P$ and

Remember Einstein's summation convention!

$$df(\mathbf{t}) = \mathbf{t}(f) = \alpha^i [\partial_i|_P](f) = \alpha^i \partial_i(f),$$

where in the last step, we suppressed the P. Theorem 28.2.8 and Eq. (28.14) yield $\alpha^i = \mathbf{t}(x^i) = (dx^i)(\mathbf{t})$. We thus have

$$df(\mathbf{t}) = \partial_i(f)\big[\big(dx^i\big)(\mathbf{t})\big] = \big[\partial_i(f)\big(dx^i\big)\big](\mathbf{t}).$$

Since this is true for all $\mathbf{t}$, we get

$$df = \partial_i(f)\big(dx^i\big) \equiv \sum_{i=1}^{m} \partial_i(f)\big(dx^i\big) = \sum_{i=1}^{m} \frac{\partial f}{\partial x^i} dx^i. \tag{28.16}$$

This is the classical formula for the differential of a function f. If we choose y^j, the jth member of a new coordinate system, for f, we obtain

$$dy^j = \sum_{i=1}^{m} \frac{\partial y^j}{\partial x^i} dx^i \equiv \frac{\partial y^j}{\partial x^i} dx^i, \tag{28.17}$$

which is the transformation dual to Eq. (28.8).

Consider a map ϕ from the product manifold $M \times N$ to another manifold L. Then

$$\phi_* : \mathcal{T}_P(M) \times \mathcal{T}_Q(N) \to \mathcal{T}_{\phi(P,Q)}(L).$$

We want to find $\phi_*(\mathbf{t}, \mathbf{s})$ for $\mathbf{t} \in \mathcal{T}_P(M)$ and $\mathbf{s} \in \mathcal{T}_Q(N)$. First define the maps $\phi_Q : M \to L$ and $\phi_P : N \to L$ by $\phi_Q(P) = \phi(P, Q)$ and $\phi_P(Q) = \phi(P, Q)$. Then

$$\phi_{Q*} : \mathcal{T}_P(M) \to \mathcal{T}_{\phi(P,Q)}(L) \quad \text{and} \quad \phi_{P*} : \mathcal{T}_Q(N) \to \mathcal{T}_{\phi(P,Q)}(L).$$

Now let $\alpha(t)$ and $\beta(t)$ be the tangent curves associated with $\mathbf{t}$ and $\mathbf{s}$ passing through P and Q, respectively. Let $f \in F^\infty(P, Q)$. Then,

$$\begin{aligned}
\phi_*(\mathbf{t}, \mathbf{s})(f) &= \frac{d}{dt}\big[(f \circ \phi)\big(\alpha(t), \beta(t)\big)\big]_{t=0} \\
&= \frac{d}{dt}\big[(f \circ \phi)\big(\alpha(t), \beta(0)\big)\big]_{t=0} + \frac{d}{dt}\big[(f \circ \phi)\big(\alpha(0), \beta(t)\big)\big]_{t=0} \\
&= \frac{d}{dt}\big[(f \circ \phi)\big(\alpha(t), Q\big)\big]_{t=0} + \frac{d}{dt}\big[(f \circ \phi)\big(P, \beta(t)\big)\big]_{t=0}
\end{aligned}$$

where the second line follows from the chain rule (or partial derivatives) and the third line from the fact that α passes through P and β through Q. From the definitions of ϕ_P and ϕ_Q, we can rewrite the last line as

$$\begin{aligned}
\phi_*(\mathbf{t}, \mathbf{s})(f) &= \frac{d}{dt}\big[(f \circ \phi_Q)\big(\alpha(t)\big)\big]_{t=0} + \frac{d}{dt}\big[(f \circ \phi_P)\big(\beta(t)\big)\big]_{t=0} \\
&\equiv \phi_{Q*}(\mathbf{t}) f + \phi_{P*}(\mathbf{s}) f.
\end{aligned}$$

We thus have the following:

Proposition 28.3.7 *The differential of* $\phi : M \times N \to L$ *at* (P, Q) *is a map* $\phi_* : \mathcal{T}_P(M) \times \mathcal{T}_Q(N) \to \mathcal{T}_{\phi(P,Q)}(L)$ *given by* $\phi_*(\mathbf{t}, \mathbf{s}) = \phi_{Q*}(\mathbf{t}) + \phi_{P*}(\mathbf{s})$, *where* $\phi_Q : M \to L$ *and* $\phi_P : N \to L$ *are defined by* $\phi_Q(P) = \phi(P, Q) = \phi_P(Q)$.

The following is a powerful theorem that constructs a submanifold out of a differentiable map (for a proof, see [Warn 83, p. 31]):

Theorem 28.3.8 *Assume that* $\psi : M \to N$ *is a* $\mathcal{C}^\infty$ *map, that* Q *is a point in the range of* ψ, *and that* $\psi_* : \mathcal{T}_P(M) \to \mathcal{T}_Q(N)$ *is surjective for all* $P \in \psi^{-1}(Q)$. *Then* $\psi^{-1}(Q)$ *is a submanifold of* M *and* $\dim \psi^{-1}(Q) = \dim M - \dim N$.

Compare this theorem with Proposition 28.1.10. There, V was an open subset of N, and since $f^{-1}(V)$ is open, it is automatically an *open* submanifold. The difficulty in proving Theorem 28.3.8 lies in the fact that $\psi^{-1}(Q)$ is *closed* because $\{Q\}$, a single point of N, is closed.

We can justify the last statement of the theorem as follows. From Eq. (28.12), we readily conclude that $\mathcal{T}_P(\psi^{-1}(Q)) = \ker \psi_{*P}$. The dimension theorem, applied to $\psi_{*P} : \mathcal{T}_P(M) \to \mathcal{T}_Q(N)$, now gives

$$\dim \mathcal{T}_P(M) = \dim \ker \psi_{*P} + \operatorname{rank} \psi_{*P}$$

$$\Rightarrow \quad \dim M = \dim \psi^{-1}(Q) + \dim N,$$

where the last equality follows from the surjectivity of ψ_{*P}.

Example 28.3.9 Consider a $\mathcal{C}^\infty$ map $f : \mathbb{R}^n \to \mathbb{R}$. Let $c \in \mathbb{R}$ such that the partial derivatives of f are defined and not all zero for all points of $f^{-1}(c)$. Then, according to Eq. (28.11), a vector $\alpha^i \partial_i \in \mathcal{T}_P(\mathbb{R}^n)$ is mapped by f_* to the vector $\alpha^i (\partial f/\partial x^i)_{f=c} d/dt$. Since $\partial f/\partial x^i$ are not all zero, by properly choosing α^i, we can make $\alpha^i (\partial f/\partial x^i)_{f=c} d/dt$ sweep over all real numbers. Therefore, f_* is surjective, and by Theorem 28.3.8, $f^{-1}(c)$ is an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$. A noteworthy special case is the function defined by

$$f\left(x^1, x^2, \ldots, x^n\right) = \left(x^1\right)^2 + \left(x^2\right)^2 + \cdots + \left(x^n\right)^2$$

and $c = r^2 > 0$. Then, $f^{-1}(c)$, an $(n-1)$-sphere of radius r, is a submanifold of $\mathbb{R}^n$.

28.4 Tensor Fields on Manifolds

So far we have studied vector spaces, learned how to construct tensors out of vectors, touched on manifolds (the abstraction of spaces), seen how to construct vectors at a single point in a manifold by the use of the tangent-at-a-curve idea, and even found the dual vectors $dx^i|_P$ to the coordinate vectors $\partial_i|_P$ at a point P of a manifold. We have everything we need to study the analysis of tensors.

28.4.1 Vector Fields

We are familiar with the concept of a vector field in 3D: Electric field, magnetic field, gravitational field, velocity field, and so forth are all familiar notions. We now want to generalize the concept so that it is applicable to a general manifold. To begin with, let us consider the following definition.

Definition 28.4.1 The union of all tangent spaces at different points of a manifold M is denoted by $T(M)$ and called the **tangent bundle** of M:

tangent bundle defined

$$T(M) = \bigcup_{P\in M} \mathcal{T}_P(M)$$

It can be shown ([Bish 80, pp. 158–164]) that $T(M)$ is a *manifold* of dimension $2\dim M$.

Definition 28.4.2 A **vector field X** on a subset U of a manifold M is a mapping $\mathbf{X} : U \to T(M)$ such that $\mathbf{X}(P) \equiv \mathbf{X}|_P \equiv \mathbf{X}_P \in \mathcal{T}_P(M)$. The set of vector fields on M is denoted by $\mathcal{X}(M)$. Let M and N be manifolds and $F : M \to N$ a differentiable map. We say that the two vector fields $\mathbf{X} \in \mathcal{X}(M)$ and $\mathbf{Y} \in \mathcal{X}(N)$ are F**-related** if $F_*(\mathbf{X}_P) = \mathbf{Y}_{F(P)}$ for all $P \in M$. This is sometimes written simply as $F_*\mathbf{X} = \mathbf{Y}$.

vector field defined

vector fields related by a map

It is worthwhile to point out that $F_*\mathbf{X}$ is not, in general, a vector field on N. To be a vector field, $F_*\mathbf{X}$ must be defined at all points of N. The natural way to define $F_*\mathbf{X}$ at $Q \in N$ is $[F_*\mathbf{X}(Q)](f) = \mathbf{X}(f \circ F)(P)$ where P is the preimage of Q, i.e., $F(P) = Q$. But there may not exist any such P (F may not be onto), or there may be more than one P (F may not be one-to-one) with such property. Therefore, this natural construction does not lead to a vector field on N. If $F_*\mathbf{X}$ happens to be a vector field on N, then it is clearly F-related to $\mathbf{X}$. In terms of the coordinates x^i, at each point $P \in M$,

$$\mathbf{X}_P \equiv \mathbf{X}|_P = X^i_P \partial_i|_P,$$

where the real numbers X^i_P are components of $\mathbf{X}_P$ in the basis $\{\partial_i|_P\}$. As P moves around in U, the real numbers X^i_P keep changing. Thus, we can think of X^i_P as a function of P and define the real-valued function $X^i : M \to \mathbb{R}$ by $X^i(P) \equiv X^i_P$. Therefore, the components of a vector field are real-valued functions on M.

Example 28.4.3 Let $M = \mathbb{R}^3$. At each point $P = (x, y, z) \in \mathbb{R}^3$, let $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z)$ be a basis for $\mathbb{R}^3$. Let $\mathcal{V}_P$ be the vector space at P. Then $T(\mathbb{R}^3)$ is the collection of all vector spaces $\mathcal{V}_P$ for all P.

We can determine the value of an electric field at a point in $\mathbb{R}^3$ by first specifying the point, as $P_0 = (x_0, y_0, z_0)$, for example. This uniquely determines the tangent space $\mathcal{T}_{P_0}(\mathbb{R}^3)$. Once we have the vector space, we can ask what the components of the electric field are in that space. These components are given by three numbers: $E_x(x_0, y_0, z_0)$, $E_y(x_0, y_0, z_0)$, and $E_z(x_0, y_0, z_0)$. The argument is the same for any other vector field.

To specify a "point" in $T(\mathbb{R}^3)$, we need three numbers to determine the location in $\mathbb{R}^3$ and another three numbers to determine the components of a vector field at that point. Thus, a "point" in $T(\mathbb{R}^3)$ is given by six "coordinates" (x, y, z, E_x, E_y, E_z), and $T(\mathbb{R}^3)$ is a six-dimensional manifold.

We know how a tangent vector $\mathbf{t}$ at a point $P \in M$ acts on a function $f \in F^\infty(P)$ to give a real number $\mathbf{t}(f)$. We can extend this, point by point, for a vector field $\mathbf{X}$ and define a function $\mathbf{X}(f)$ by

$$\big[\mathbf{X}(f)\big](P) \equiv \mathbf{X}_P(f), \quad P \in U, \tag{28.18}$$

where U is a subset of M on which both $\mathbf{X}$ and f are defined. The RHS is well-defined because we know how $\mathbf{X}_P$, the vector at P, acts on functions at P to give the real number $[\mathbf{X}_P](f)$. On the LHS, we have $\mathbf{X}(f)$, which maps the point P onto a real number. Thus, $\mathbf{X}(f)$ is indeed a real-valued function on M. We can therefore define vector fields directly as operators on $\mathcal{C}^\infty$ functions satisfying

$$\mathbf{X}(\alpha f + \beta g) = \alpha \mathbf{X}(f) + \beta \mathbf{X}(g),$$

$$\mathbf{X}(fg) = \big[\mathbf{X}(f)\big]g + \big[\mathbf{X}(g)\big]f.$$

$\mathcal{C}^\infty$ vector fields

A prototypical vector field is the coordinate vector field ∂_i. In general, $\mathbf{X}(f)$ is not a $\mathcal{C}^\infty$ function even if f is. A vector field that produces a $\mathcal{C}^\infty$ function $\mathbf{X}(f)$ for every $\mathcal{C}^\infty$ function f is called a $\mathcal{C}^\infty$ **vector field**. Such a vector field has components that are $\mathcal{C}^\infty$ functions on M.

The set of tangent vectors $\mathcal{T}_P(M)$ at a point $P \in M$ form an m-dimensional vector space. The set of vector fields $\mathcal{X}(M)$—which yield a vector at every point of the manifold—also constitutes a vector space. However, this vector space is (uncountably) infinite-dimensional.

A property of $\mathcal{X}(M)$ that is absent in $\mathcal{T}_P(M)$ is composition.[6] This suggests the possibility of defining a "product" on $\mathcal{X}(M)$ to turn it into an algebra. Let $\mathbf{X}$ and $\mathbf{Y}$ be vector fields. For $\mathbf{X} \circ \mathbf{Y}$ to be a vector field, it has to satisfy the derivation property. But

$$\begin{aligned}
\mathbf{X} \circ \mathbf{Y}(fg) &= \mathbf{X}\big(\mathbf{Y}(fg)\big) = \mathbf{X}\big(\mathbf{Y}(f)g + f\mathbf{Y}(g)\big) \\
&= \big(\mathbf{X}\big(\mathbf{Y}(f)\big)\big)g + \mathbf{Y}(f)\mathbf{X}(g) + \mathbf{X}(f)\mathbf{Y}(g) + f\big(\mathbf{X}\big(\mathbf{Y}(g)\big)\big) \\
&\neq \big(\mathbf{X} \circ \mathbf{Y}(f)\big)g + f\big(\mathbf{X} \circ \mathbf{Y}(g)\big).
\end{aligned}$$

However, the reader may verify that $\mathbf{X} \circ \mathbf{Y} - \mathbf{Y} \circ \mathbf{X}$ does indeed satisfy the derivation property. Therefore, by defining the binary operation $\mathcal{X}(M) \times$

[6]Recall that a typical element of $\mathcal{T}_P(M)$ is a map $\mathbf{t} : F^\infty(P) \to \mathbb{R}$ for which composition is meaningless.

$\mathfrak{X}(M) \to \mathfrak{X}(M)$ as

$$[\mathbf{X}, \mathbf{Y}] \equiv \mathbf{X} \circ \mathbf{Y} - \mathbf{Y} \circ \mathbf{X},$$

The set of vector fields forms a Lie algebra under Lie bracket multiplication.

$\mathfrak{X}(M)$ becomes an algebra, called the **Lie algebra** of vector fields of M. The binary operation is called the **Lie bracket**. Although it was not mentioned at the time, we have encountered another example of a Lie algebra in Chap. 4, namely $\mathcal{L}(\mathcal{V})$ under the binary operation of the commutation relation. Lie brackets have the following two properties:

$$[\mathbf{X}, \mathbf{Y}] = -[\mathbf{Y}, \mathbf{X}],$$

$$\big[[\mathbf{X}, \mathbf{Y}], \mathbf{Z}\big] + \big[[\mathbf{Z}, \mathbf{X}], \mathbf{Y}\big] + \big[[\mathbf{Y}, \mathbf{Z}], \mathbf{X}\big] = 0.$$

These two relations are the defining properties of all Lie algebras. The last relation is called the **Jacobi identity**. $\mathfrak{X}(M)$ with Lie brackets is an example of an infinite-dimensional Lie algebra; $\mathcal{L}(\mathcal{V})$ with commutators is an example of a finite-dimensional Lie algebra.

Jacobi identity

We shall have occasion to use the following theorem in our treatment of Lie groups and algebras in Chap. 29:

Theorem 28.4.4 *Let M and N be manifolds and $F : M \to N$ a differentiable map. Assume that $\mathbf{X}_i \in \mathfrak{X}(M)$ is F-related to $\mathbf{Y}_i \in \mathfrak{X}(N)$ for $i = 1, 2$. Then $[\mathbf{X}_1, \mathbf{X}_2]$ is F-related to $[\mathbf{Y}_1, \mathbf{Y}_2]$, i.e.,*

$$F_*[\mathbf{X}_1, \mathbf{X}_2] = [F_*\mathbf{X}_1, F_*\mathbf{X}_2].$$

Proof Let f be an arbitrary function on N. Then

$$\begin{aligned}\big(F_*[\mathbf{X}_1, \mathbf{X}_2]\big) f &\equiv [\mathbf{X}_1, \mathbf{X}_2](f \circ F) = \mathbf{X}_1\big(\mathbf{X}_2(f \circ F)\big) - \mathbf{X}_2\big(\mathbf{X}_1(f \circ F)\big) \\ &= \mathbf{X}_1\big(\big[F_*\mathbf{X}_2(f)\big] \circ F\big) - \mathbf{X}_2\big(\big[F_*\mathbf{X}_1(f)\big] \circ F\big) \\ &= F_*\mathbf{X}_1\big(F_*\mathbf{X}_2(f)\big) - F_*\mathbf{X}_2\big(F_*\mathbf{X}_1(f)\big) \\ &= [F_*\mathbf{X}_1, F_*\mathbf{X}_2] f,\end{aligned}$$

where we used Eq. (28.9) in the first, second, and third lines, and the result of Problem 28.8 in the second line. □

It is convenient to visualize vector fields as streamlines. In fact, most of the terminology used in three-dimensional vector analysis, such as flux, divergence, and curl, have their origins in the flow of fluids and the associated velocity vector fields. The streamlines are obtained—in nonturbulent flow—by starting at one point and drawing a curve whose tangent at all points is the velocity vector field. For a smooth flow this curve is unique. There is an exact analogy in manifold theory.

Definition 28.4.5 Let $\mathbf{X} \in \mathfrak{X}(M)$ be defined on an open subset U of M. An **integral curve of X** in U is a curve γ whose range lies in U and for every t in the domain of γ, the vector tangent to γ satisfies $\gamma_{*t} = \mathbf{X}(\gamma(t))$. If $\gamma(0) = P$, we say that γ **starts at** P.

integral curve of a vector field

Let us choose a coordinate system on M. Then $\mathbf{X} \equiv X^i \partial_i$, where X^i are $\mathcal{C}^\infty$ functions on M, and, by (28.13), $\gamma_* = \dot{\gamma}^i \partial_i$. The equation for the integral curve of $\mathbf{X}$ will therefore become

$$\dot{\gamma}^i \partial_i = X^i\big(\gamma(t)\big)\partial_i, \quad \text{or} \quad \frac{d\gamma^i}{dt} = X^i\big(\gamma^1(t), \dots, \gamma^m(t)\big), \quad i = 1, 2, \dots, m.$$

Since γ^i are simply coordinates of points on M, we rewrite the equation above as

$$\frac{dx^i}{dt} = X^i\big(x^1(t), \dots, x^m(t)\big), \quad i = 1, 2, \dots, m. \tag{28.19}$$

This is a system of first-order differential equations that has a unique (local) solution once the initial value $\gamma(0)$ of the curve, i.e., the coordinates of the starting point P, is given. The precise statement for existence and uniqueness of integral curves is contained in the following theorem.

Theorem 28.4.6 *Let $\mathbf{X}$ be a $\mathcal{C}^\infty$ vector field defined on an open subset U of M. Suppose $P \in U$, and $c \in \mathbb{R}$. Then there is a positive number ϵ and a unique integral curve γ of $\mathbf{X}$ defined on $|t - c| \le \epsilon$ such that $\gamma(c) = P$.*

Example 28.4.7 (Examples of integral curves)

(a) Let $M = \mathbb{R}$ with coordinate function x. The vector field $\mathbf{X} = x\partial_x$ has an integral curve with initial point x_0 given by the DE $dx/dt = x(t)$, which has the solution $x(t) = e^t x_0$.

(b) Let $M = \mathbb{R}^n$ with coordinate functions x^i. The vector field $\mathbf{X} = \sum a^i \partial_i$ has an integral curve, with initial point $\mathbf{r}_0$, given by the system of DEs $dx^i/dt = a^i$, which has the solution $x^i(t) = a^i t + x_0^i$, or $\mathbf{r} = \mathbf{a}t + \mathbf{r}_0$. The curve is therefore a straight line parallel to $\mathbf{a}$ going through $\mathbf{r}_0$.

(c) Let $M = \mathbb{R}^n$ with coordinate functions x^i. Consider the vector field

$$\mathbf{X} = \sum_{i,j=1}^{n} a^i_j x^j \partial_i.$$

The integral curve of this vector field, with initial point $\mathbf{r}_0$, is given by the system of DEs $dx^i/dt = \sum_{j=1}^n a^i_j x^j$, which can be written in vector form as $d\mathbf{r}/dt = \mathsf{A}\mathbf{r}$ where A is a constant matrix. By differentiating this equation several times, one can convince oneself that $d^k\mathbf{r}/dt^k = \mathsf{A}^k\mathbf{r}$. The Taylor expansion of $\mathbf{r}(t)$ then yields

$$\mathbf{r}(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k\mathbf{r}}{dt^k}\bigg|_{t=0} t^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathsf{A}^k \mathbf{r}_0 = e^{t\mathsf{A}}\mathbf{r}_0.$$

(d) Let $M = \mathbb{R}^2$ with coordinate x, y. The reader may verify that the vector field $\mathbf{X} = -y\partial_x + x\partial_y$ has an integral curve through (x_0, y_0) given by

$$x = x_0 \cos t - y_0 \sin t,$$

$$y = x_0 \sin t + y_0 \cos t,$$

i.e., a circle centered at the origin passing through (x_0, y_0).

Going back to the velocity vector field analogy, we can think of integral curves as the path of particles flowing with the fluid. If we think of the entire fluid as a manifold M, the flow of particles can be thought of as a transformation of M. To be precise, let M be an arbitrary manifold, and $\mathbf{X} \in \mathcal{X}(M)$. At each point P of M, there is a unique local integral curve γ_P of $\mathbf{X}$ starting at P defined on an open subset U of M. The map $F_t : U \to M$ defined by $F_t(P) = \gamma_P(t)$ is a (local) transformation of M. The collection of such maps with different t's is called the **flow** of the vector field $\mathbf{X}$. The uniqueness of the integral curve γ_P implies that F_t is a local diffeomorphism. In fact, the collection of maps $\{F_t\}_{t\in\mathbb{R}}$ forms a (local) **one-parameter group of transformations** in the sense that

flow of a vector field

$$F_t \circ F_s = F_{t+s}, \qquad F_0 = \text{id}, \qquad (F_t)^{-1} = F_{-t}. \tag{28.20}$$

One has to keep in mind that F_t at a point $P \in M$ is, in general, defined only *locally* in t, i.e., only for t in some open interval that depends on P. For some special, but important, cases this interval can be taken to be the entire $\mathbb{R}$ for all P, in which case we speak of a **global one-parameter group of transformations**, and $\mathbf{X}$ is called a **complete vector field** on M.

Global 1-parameter group of transformations; complete vector fields

The symbol F_t used for the flow of the vector field $\mathbf{X}$ does not contain its connection to $\mathbf{X}$. In order to make this connection, it is common to define

$$F_t \equiv \exp(t\mathbf{X}). \tag{28.21}$$

This definition, with no significance attached to "exp" at this point, converts Eq. (28.20) into

$$\begin{aligned} \exp(t\mathbf{X}) \circ \exp(s\mathbf{X}) &= \exp\big[(t+s)\mathbf{X}\big], \\ \exp(0\mathbf{X}) &= \text{id}, \\ \big[\exp(t\mathbf{X})\big]^{-1} &= \exp(-t\mathbf{X}), \end{aligned} \tag{28.22}$$

which notationally justifies the use of "exp". We shall see in our discussion of Lie groups that this choice of notation is not accidental.

Using this notation, we can write

$$\mathbf{X}_P(f) \equiv \frac{d}{dt} f \circ F_t(P)\bigg|_{t=0} = \frac{d}{dt} f \circ \exp(t\mathbf{X})\bigg|_{t=0}.$$

One usually leaves out the function f and writes

$$\mathbf{X}_P = \frac{d}{dt}\exp(t\mathbf{X})\bigg|_{t=0}, \tag{28.23}$$

where it is understood that the LHS acts on some f that must compose on the RHS to the left of the exponential. Similarly, we have

$$(F_*\mathbf{X})_{F(P)} = \frac{d}{dt}F(\exp t\mathbf{X})\bigg|_{t=0},$$
$$G_{*F(P)}\underbrace{\left(\frac{d}{dt}F(\exp t\mathbf{X})\bigg|_{t=0}\right)}_{=F_*(\mathbf{X})} = \frac{d}{dt}G\circ F(\exp t\mathbf{X})\bigg|_{t=0}, \tag{28.24}$$

where $F : M \to N$ and $G : N \to K$ are maps between manifolds.

Example 28.4.8 In this example, we derive a useful formula that gives the value of a function at a neighboring point of $P \in M$ located on the integral curve of $\mathbf{X} \in \mathcal{X}(M)$ going through P. We first note that since $\mathbf{X}_P$ is tangent to γ_P at $P = \gamma(0)$, by Proposition 28.2.4 we have

$$\mathbf{X}_P(f) = \frac{d}{dt}f\big(\gamma_P(t)\big)\bigg|_{t=0} = \frac{d}{dt}f\big(F_t(P)\big)\bigg|_{t=0}.$$

Next we use the definition of derivative and the fact that $F_0(P) = P$ to write

$$\lim_{t\to 0}\frac{1}{t}\big[f\big(F_t(P)\big) - f(P)\big] = \mathbf{X}_P(f).$$

Now, if we assume that t is very small, we have

$$f\big(F_t(P)\big) = f(P) + t\mathbf{X}_P(f) + \cdots, \tag{28.25}$$

which is a Taylor series with only the first two terms kept.

28.4.2 Tensor Fields

We have defined vector spaces $\mathcal{T}_P(M)$ at each point of M. We have also constructed coordinate bases, $\{\partial_i|_P\}_{i=1}^m$, for these vector spaces. At the end of Sect. 28.2, we showed that the differentials $\{dx^i|_P\}_{i=1}^m$ form a basis that is dual to $\{\partial_i|_P\}_{i=1}^m$. Let us concentrate on this dual space, which we will denote by $\mathcal{T}_P^*(M)$.

cotangent bundle of a manifold

Taking the union of all $\mathcal{T}_P^*(M)$ at all points of M, we obtain the **cotangent bundle** of M:

$$T^*(M) = \bigcup_{P\in M}\mathcal{T}_P^*(M). \tag{28.26}$$

This is the dual space of $T(M)$ at each point of M. We can now define the analogue of the vector field for the cotangent bundle.

differential one-form

Definition 28.4.9 A **differential one-form $\boldsymbol{\theta}$** on a subset U of a manifold M is a mapping $\boldsymbol{\theta} : U \to T^*(M)$ such that $\boldsymbol{\theta}(P) \equiv \boldsymbol{\theta}_P \in \mathcal{T}_P^*(M)$. The collection of all one-forms on M is denoted by $\mathcal{X}^*(M)$.

If $\boldsymbol{\theta}$ is a one-form and $\mathbf{X}$ is a vector field on M, then $\boldsymbol{\theta}(\mathbf{X})$ is a real-valued function on M defined naturally by $[\boldsymbol{\theta}(\mathbf{X})](P) \equiv (\boldsymbol{\theta}_P)(\mathbf{X}_P)$. The first factor on the RHS is a linear functional at P, and the second factor is a vector at P. So, the pairing of the two factors produces a real number. A prototypical one-form is the coordinate differential, dx^i.

Associated with a differentiable map $\psi : M \to N$, we defined a differential ψ_* that mapped a tangent space of M to a tangent space of N. The dual of ψ_* (Definition 2.5.4) is denoted by ψ^* and is called the **pullback** of ψ. It takes a one-form on N to a one-form on M. In complete analogy to the case of vector fields, $\boldsymbol{\theta}$ can be written in terms of the basis $\{dx^i\}$: $\boldsymbol{\theta} = \theta_i dx^i$. Here θ_i, the components of $\boldsymbol{\theta}$, are real-valued functions on M.

pullback of a differentiable map

With the vector spaces $\mathcal{T}_P(M)$ and $\mathcal{T}_P^*(M)$ at our disposal, we can construct various kinds of tensors at each point P. The union of all these tensors is a manifold, and a tensor field can be defined as usual. Thus, we have the following definition.

Definition 28.4.10 Let $\mathcal{T}_P(M)$ and $\mathcal{T}_P^*(M)$ be the tangent and cotangent spaces at $P \in M$. Then the set of tensors of type (r,s) on $\mathcal{T}_P(M)$ is denoted by $\mathcal{T}_{s,P}^r(M)$. The **bundle of tensors** of type (r,s) over M, denoted by $T_s^r(M)$, is

bundle of tensors and tensor fields

$$T_s^r(M) = \bigcup_{P \in M} \mathcal{T}_{s,P}^r(M).$$

A **tensor field** $\mathbf{T}$ of type (r,s) over a subset U of M is a mapping $\mathbf{T} : U \to T_s^r(M)$ such that $\mathbf{T}(P) \equiv \mathbf{T}_P \equiv \mathbf{T}|_P \in \mathcal{T}_{s,P}^r(M)$.

In particular, $T_0^0(M)$ is the set of real-valued functions on M, $T_0^1(M) = T(M)$, and $T_1^0(M) = T^*(M)$. Furthermore, since $\mathbf{T}$ is a multilinear map, the parentheses are normally reserved for vectors and their duals, and as indicated in Definition 28.4.10, the value of $\mathbf{T}$ at $P \in M$ is written as $\mathbf{T}_P$ or $\mathbf{T}|_P$. The reader may check that the map

$$\mathbf{T} : \underbrace{\mathcal{X}^*(M) \times \cdots \times \mathcal{X}^*(M)}_{r \text{ times}} \times \underbrace{\mathcal{X}(M) \times \cdots \times \mathcal{X}(M)}_{s \text{ times}} \to T_0^0(M)$$

defined by

$$\big[\mathbf{T}\big(\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^r, \mathbf{v}_1, \ldots, \mathbf{v}_s\big)\big](P) = \mathbf{T}_P\big(\boldsymbol{\omega}^1|_P, \ldots, \boldsymbol{\omega}^r|_P, \mathbf{v}_1|_P, \ldots, \mathbf{v}_s|_P\big)$$

has the property that

A crucial property of tensors

$$\begin{aligned} \mathbf{T}\big(\ldots, f\boldsymbol{\omega}^j + g\boldsymbol{\theta}^j, \ldots\big) &= f\mathbf{T}\big(\ldots, \boldsymbol{\omega}^j, \ldots\big) + g\mathbf{T}\big(\ldots, \boldsymbol{\theta}^j, \ldots\big), \\ \mathbf{T}(\ldots, f\mathbf{v}_k + g\mathbf{u}_k, \ldots) &= f\mathbf{T}(\ldots, \mathbf{v}_k, \ldots) + g\mathbf{T}(\ldots, \mathbf{u}_k, \ldots) \end{aligned} \tag{28.27}$$

for any two *functions* f and g on M. Thus,[7]

[7] In mathematical jargon, $\mathcal{X}(M)$ and $\mathcal{X}^*(M)$ are called **modules** over the (ring of) real-valued functions on M. Rings are a generalization of the real numbers (field of real numbers) whose elements have all the properties of a field except that they may have no inverse. A module over a field is a vector space.

Box 28.4.11 *A tensor is linear in vector fields and* 1*-forms, even when the coefficients of linear expansion are functions.*

The components of **T** with respect to coordinates x^i are the m^{r+s} real-valued functions

$$T^{i_1 i_2 \dots i_r}_{j_1 j_2 \dots j_s} = \mathbf{T}\big(dx^{i_1}, dx^{i_2}, \dots, dx^{i_r}, \partial_{j_1}, \partial_{j_2}, \dots, \partial_{j_s}\big).$$

If tensor fields are to be of any use, we must be able to differentiate them. We shall consider three types of derivatives with different applications. We study one of them here, another in the next section, and the third in Chap. 36.

Derivatives can be defined only for objects that can be added. For functions of a single (real or complex) variable, this is done almost subconsciously: We take the difference between the values of the function at two nearby points and divide by the length of the interval between the two points. We extended this definition to operators in Chap. 4 with practically no change. For functions of more than one variable, one chooses a direction (a vector) and considers change in the function along that direction. This leads to the concept of **directional derivative**, or partial derivative when the vector happens to be along one of the axes.

directional derivative

In all the above cases, the objects being differentiated reside in the *same* space: $f(t)$ and $f(t + \Delta t)$ are both real (complex) numbers; $\mathbf{H}(t)$ and $\mathbf{H}(t + \Delta t)$ both belong to $\mathcal{L}(\mathcal{V})$. When we try to define derivatives of tensor fields, however, we run immediately into trouble: $\mathbf{T}_P$ and $\mathbf{T}_{P'}$ cannot be compared because they belong to two different spaces, one to $\mathcal{T}^r_{s,P}(M)$ and the other to $\mathcal{T}^r_{s,P'}(M)$. To make comparisons, we need first to establish a "connection" between the two spaces. This connection has to be a vector space isomorphism so that there is one and only one vector in the second space that is to be compared with a given vector in the first space. The problem is that there are infinitely many isomorphisms between any given two vector spaces. No "natural" isomorphism exists between $\mathcal{T}^r_{s,P}(M)$ and $\mathcal{T}^r_{s,P'}(M)$; thus the diversity of tensor "derivatives!" We narrow down this diversity by choosing a specific vector at $\mathcal{T}^r_{s,P}(M)$ and seeking a natural way of defining the derivative along that vector by associating a "natural" isomorphism corresponding to the vector. There are a few methods of doing this. We describe one of them here.

difficulty associated with differentiating tensors

First, let us see what happens to tensor fields under a diffeomorphism of M onto itself. Let $F : M \to M$ be such a diffeomorphism. The differential F_{*P} of this diffeomorphism is an isomorphism of $\mathcal{T}_P(M)$ and $\mathcal{T}_{F(P)}(M)$. This isomorphism induces an isomorphism of the vector spaces $\mathcal{T}^r_{s,P}(M)$ and $\mathcal{T}^r_{s,F(P)}(M)$—also denoted by F_{*P}—by Eq. (26.10). Let us denote by F_* a map of $T(M)$ onto $T(M)$ whose restriction to $\mathcal{T}_P(M)$ is F_{*P}. If **T** is a tensor field on M, then $F_*(\mathbf{T})$ is also a tensor field, whose value at $F(Q)$ is obtained by letting F_{*Q} act on $\mathbf{T}(Q)$:

$$\big[F_*(\mathbf{T})\big]\big(F(Q)\big) = F_{*Q}\big(\mathbf{T}(Q)\big),$$

or, letting $P = F(Q)$ or $Q = F^{-1}(P)$,

$$\big[F_*(\mathbf{T})\big](P) = F_{*F^{-1}(P)}\big(\mathbf{T}\big(F^{-1}(P)\big)\big). \tag{28.28}$$

Now, let $\mathbf{X}$ be a vector field and $P \in M$. The flow of $\mathbf{X}$ at P defines a local diffeomorphism $F_t : U \to F_t(U)$ with $P \in U$. The differential F_{t*} of this diffeomorphism is an isomorphism of $\mathcal{T}_P(M)$ and $\mathcal{T}_{F_t(P)}(M)$. As discussed above, this isomorphism induces an isomorphism of the vector space $\mathcal{T}^r_{s,P}(M)$ onto itself. The derivative we are after is defined by comparing a tensor field evaluated at P with the image of the same tensor field under the isomorphism F_{t*}^{-1}. The following definition makes this procedure more precise.

Definition 28.4.12 Let $P \in M$, $\mathbf{X} \in \mathcal{X}(M)$, and F_t the flow of $\mathbf{X}$ defined in a neighborhood of P. The **Lie derivative** of a tensor field $\mathbf{T}$ at P **with respect to X** is denoted by $(L_{\mathbf{X}}\mathbf{T})_P$ and defined by

Lie derivative of tensor fields with respect of a vector field

$$(L_{\mathbf{X}}\mathbf{T})_P = \lim_{t\to 0} \frac{1}{t}\big[F_{t*}^{-1}\mathbf{T}_{F_t(P)} - \mathbf{T}_P\big] \equiv \frac{d}{dt}F_{t*}^{-1}\mathbf{T}_{F_t(P)}\bigg|_{t=0}. \tag{28.29}$$

Let us calculate the derivative in Eq. (28.29) at an arbitrary value of t. For this purpose, let $Q \equiv F_t(P)$. Then

$$\begin{aligned}
\frac{d}{dt}F_{t*}^{-1}\mathbf{T}_{F_t(P)} &\equiv \lim_{\Delta t\to 0}\frac{1}{\Delta t}\big[F_{(t+\Delta t)*}^{-1}\mathbf{T}_{F_{t+\Delta t}(P)} - F_{t*}^{-1}\mathbf{T}_{F_t(P)}\big] \\
&= F_{t*}^{-1}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\big[F_{\Delta t*}^{-1}\mathbf{T}_{F_{t+\Delta t}(P)} - \mathbf{T}_{F_t(P)}\big] \\
&= F_{t*}^{-1}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\big[F_{\Delta t*}^{-1}\mathbf{T}_{F_{\Delta t}(Q)} - \mathbf{T}_Q\big] \equiv F_{t*}^{-1}(L_{\mathbf{X}}\mathbf{T})_Q.
\end{aligned}$$

Since Q is arbitrary, we can remove it from the equation and write, as the generalization of Eq. (28.29),

$$L_{\mathbf{X}}\mathbf{T} = F_{t*}\frac{d}{dt}F_{t*}^{-1}\mathbf{T}. \tag{28.30}$$

An important special case of the definition above is the Lie derivative of a vector field with respect to another. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{X}(M)$. To evaluate the RHS of (28.29), we apply the first term in the brackets to an arbitrary function f,

$$\begin{aligned}
\big[F_{t*}^{-1}\mathbf{Y}_{F_t(P)}\big](f) &= \mathbf{Y}_{F_t(P)}\big(f\circ F_t^{-1}\big) = \mathbf{Y}\big(f\circ F_t^{-1}\big)\big|_{F_t(P)} \\
&= \mathbf{Y}(f\circ F_{-t})\big|_{F_t(P)} = \mathbf{Y}\big(f - t\mathbf{X}(f)\big)\big|_{F_t(P)} \\
&= (\mathbf{Y}f)_{F_t(P)} - t\mathbf{Y}\big(\mathbf{X}(f)\big)\big|_{F_t(P)} \\
&= (\mathbf{Y}f)_P + t\big[\mathbf{X}(\mathbf{Y}f)\big]_P - t\big\{\big[\mathbf{Y}(\mathbf{X}f)\big]_P \\
&\quad + t\big[\mathbf{X}\big(\mathbf{Y}(\mathbf{X}f)\big)\big]_P\big\} \\
&= \mathbf{Y}_P(f) + t\mathbf{X}_P\circ\mathbf{Y}_P(f) - t\mathbf{Y}_P\circ\mathbf{X}_P(f) \\
&= \mathbf{Y}_P(f) + t[\mathbf{X}_P, \mathbf{Y}_P](f) = \mathbf{Y}_P(f) + t[\mathbf{X}, \mathbf{Y}]_P(f).
\end{aligned}$$

The first equality on the first line follows from (28.9), the second equality from the meaning of $\mathbf{Y}_{F_t(P)}$; the second equality on the second line and the fourth line follow from (28.25). Finally, the fifth line follows if we ignore the t^2 term. Therefore,

$$(L_{\mathbf{X}}\mathbf{Y})_P(f) = \lim_{t\to 0}\frac{1}{t}\big[F_{t*}^{-1}\mathbf{Y}_{F_t(P)} - \mathbf{Y}_P\big](f)$$
$$= \lim_{t\to 0}\frac{1}{t}\big\{t[\mathbf{X},\mathbf{Y}]_P\big\}(f) = [\mathbf{X},\mathbf{Y}]_P(f).$$

Lie derivative is commutator

Since this is true for all P and f, we get

$$L_{\mathbf{X}}\mathbf{Y} = [\mathbf{X},\mathbf{Y}]. \tag{28.31}$$

This and other properties of the Lie derivative are summarized in the following proposition.

Proposition 28.4.13 *Let* $\mathsf{T} \in T_s^r(M)$ *and* T' *be arbitrary tensor fields and* $\mathbf{X}$ *a given vector field. Then*

Properties of Lie derivative

1. $L_{\mathbf{X}}$ *satisfies a derivation property in the algebra of tensor fields, i.e.,*

$$L_{\mathbf{X}}\big(\mathsf{T}\otimes\mathsf{T}'\big) = (L_{\mathbf{X}}\mathsf{T})\otimes\mathsf{T}' + \mathsf{T}\otimes\big(L_{\mathbf{X}}\mathsf{T}'\big).$$

2. $L_{\mathbf{X}}$ *is type-preserving, i.e.,* $L_{\mathbf{X}}\mathsf{T}$ *is a tensor field of type* (r,s).
3. $L_{\mathbf{X}}$ *commutes with the operation of contraction of tensor fields; in particular, in combination with property* 1, *we have*

$$L_{\mathbf{X}}\langle\boldsymbol{\theta},\mathbf{Y}\rangle = \langle L_{\mathbf{X}}\boldsymbol{\theta},\mathbf{Y}\rangle + \langle\boldsymbol{\theta}, L_{\mathbf{X}}\mathbf{Y}\rangle.$$

4. $L_{\mathbf{X}} f = \mathbf{X} f$ *for every function* f.
5. $L_{\mathbf{X}}\mathbf{Y} = [\mathbf{X},\mathbf{Y}]$ *for every vector field* $\mathbf{Y}$.

Proof Except for the last property, which we demonstrated above, the rest follow directly from definitions and simple manipulations. The details are left as exercises. □

Although the Lie derivative of a vector field is nicely given in terms of commutators, no such simple relation exists for the Lie derivative of a 1-form. However, if we work in a given coordinate frame, then a useful expression for the Lie derivative of a 1-form can be obtained. Applying $L_{\mathbf{X}}$ to $\langle\boldsymbol{\theta},\mathbf{X}\rangle$, we obtain

$$\underbrace{L_{\mathbf{X}}\langle\boldsymbol{\theta},\mathbf{Y}\rangle}_{=\mathbf{X}(\langle\boldsymbol{\theta},\mathbf{Y}\rangle)} = \langle L_{\mathbf{X}}\boldsymbol{\theta},\mathbf{Y}\rangle + \langle\boldsymbol{\theta}, L_{\mathbf{X}}\mathbf{Y}\rangle = \langle L_{\mathbf{X}}\boldsymbol{\theta},\mathbf{Y}\rangle + \big\langle\boldsymbol{\theta},[\mathbf{X},\mathbf{Y}]\big\rangle. \tag{28.32}$$

In particular, if $\mathbf{Y} = \partial_i$ and we write $\mathbf{X} = X^j\partial_j$, $\boldsymbol{\theta} = \theta_j dx^j$, then the LHS becomes $\mathbf{X}(\theta_i) = X^j\partial_j\theta_i$, and the RHS can be written as

$$(L_{\mathbf{X}}\boldsymbol{\theta})_i + \big\langle\boldsymbol{\theta}, \underbrace{\big[X^j\partial_j, \partial_i\big]}_{-(\partial_i X^j)\partial_j}\big\rangle.$$

It follows that

$$L_{\mathbf{X}}\boldsymbol{\theta} \equiv (L_{\mathbf{X}}\boldsymbol{\theta})_i dx^i = \big(X^j \partial_j \theta_i + \theta_j \partial_i X^j\big) dx^i. \tag{28.33}$$

We give two other useful properties of the Lie derivative applicable to all tensors. From the Jacobi identity one can readily deduce that

$$L_{[\mathbf{X},\mathbf{Y}]}\mathbf{Z} = L_{\mathbf{X}} L_{\mathbf{Y}}\mathbf{Z} - L_{\mathbf{Y}} L_{\mathbf{X}}\mathbf{Z}.$$

Similarly, Eq. (28.32) yields

$$L_{[\mathbf{X},\mathbf{Y}]}\boldsymbol{\theta} = L_{\mathbf{X}} L_{\mathbf{Y}}\boldsymbol{\theta} - L_{\mathbf{Y}} L_{\mathbf{X}}\boldsymbol{\theta}.$$

Putting these two equations together, recalling that a general tensor is a linear combination of tensor products of vectors and 1-forms, and that the Lie derivative obeys the product rule of differentiation, we obtain

$$L_{[\mathbf{X},\mathbf{Y}]}\mathbf{T} = L_{\mathbf{X}} L_{\mathbf{Y}}\mathbf{T} - L_{\mathbf{Y}} L_{\mathbf{X}}\mathbf{T} \tag{28.34}$$

for any tensor field $\mathbf{T}$. Furthermore, Eq. (28.33) and the linearity of the Lie bracket imply that $L_{\alpha\mathbf{X}+\beta\mathbf{Y}} = \alpha L_{\mathbf{X}} + \beta L_{\mathbf{Y}}$ when acting on vectors and 1-forms. It follows by the same argument as above that

$$L_{\alpha\mathbf{X}+\beta\mathbf{Y}}\mathbf{T} = \alpha L_{\mathbf{X}}\mathbf{T} + \beta L_{\mathbf{Y}}\mathbf{T} \quad \forall \mathbf{T} \in \mathcal{T}^r_s(M). \tag{28.35}$$

Equation (28.32) gives a rule for calculating the Lie derivative of a 1-form, i.e., it tells us how to evaluate $L_{\mathbf{X}}\boldsymbol{\theta}$ on a vector $\mathbf{Y}$. We can generalize this for a p-form $\boldsymbol{\omega}$. Write the evaluation of $\boldsymbol{\omega}$ on p vectors as p contractions as in Eq. (26.8):

$$\boldsymbol{\omega}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p) = \mathbf{C}^p_p \cdots \mathbf{C}^2_2 \mathbf{C}^1_1 (\boldsymbol{\omega} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2 \otimes \cdots \otimes \mathbf{X}_p).$$

Now apply the Lie derivative on both sides and use its derivation property and the fact the it commutes with contractions to get

$$L_{\mathbf{X}}\big(\boldsymbol{\omega}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p)\big) = \mathbf{C}^p_p \cdots \mathbf{C}^2_2 \mathbf{C}^1_1 L_{\mathbf{X}} (\boldsymbol{\omega} \otimes \mathbf{X}_1 \otimes \mathbf{X}_2 \otimes \cdots \otimes \mathbf{X}_p).$$

The left-hand side is just $\mathbf{X}(\boldsymbol{\omega}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p))$. For the right-hand side, we use

$$\begin{aligned} &L_{\mathbf{X}}(\boldsymbol{\omega} \otimes \mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_p) \\ &\quad = (L_{\mathbf{X}}\boldsymbol{\omega}) \otimes \mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_p + \sum_{i=1}^p \boldsymbol{\omega} \otimes \mathbf{X}_1 \otimes \cdots \otimes L_{\mathbf{X}}\mathbf{X}_i \otimes \cdots \otimes \mathbf{X}_p. \end{aligned}$$

Applying the contractions, using $L_{\mathbf{X}}\mathbf{X}_i = [\mathbf{X}, \mathbf{X}_i]$, and putting everything together, we obtain

$$\begin{aligned} &\mathbf{X}\big(\boldsymbol{\omega}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p)\big) \\ &\quad = (L_{\mathbf{X}}\boldsymbol{\omega})(\mathbf{X}_1, \ldots, \mathbf{X}_p) + \sum_{i=1}^p \boldsymbol{\omega}\big(\mathbf{X}_1, \ldots, [\mathbf{X}, \mathbf{X}_i], \ldots, \mathbf{X}_p\big) \end{aligned}$$

which finally gives the rule by which $L_{\mathbf{X}}\boldsymbol{\omega}$ acts on p vectors:

$$(L_{\mathbf{X}}\boldsymbol{\omega})(\mathbf{X}_1, \dots, \mathbf{X}_p)$$
$$= \mathbf{X}\bigl(\boldsymbol{\omega}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)\bigr) - \sum_{i=1}^{p} \boldsymbol{\omega}\bigl(\mathbf{X}_1, \dots, [\mathbf{X}, \mathbf{X}_i], \dots, \mathbf{X}_p\bigr). \quad (28.36)$$

28.5 Exterior Calculus

Skew-symmetric tensors are of special importance to applications. We studied these tensors in their algebraic format in Chap. 26. Let us now investigate them as they reside on manifolds.

differential forms, or simply forms, defined

Definition 28.5.1 Let M be a manifold and Q a point of M. Let $\Lambda_Q^p(M)$ denote the space of all antisymmetric tensors of rank p over the tangent space at Q. Let $\Lambda^p(M)$ be the union of all $\Lambda_Q^p(M)$ for all $Q \in M$. A **differential** p**-form** $\boldsymbol{\omega}$ is a mapping $\boldsymbol{\omega} : U \to \Lambda^p(M)$ such that $\boldsymbol{\omega}(Q) \in \Lambda_Q^p(M)$ where U is, as usual, a subset of M. To emphasize their domain of definition, we sometimes use the notation $\Lambda^p(U)$.

Since $\{dx^i\}_{i=1}^m$ is a basis for $T_Q^*(M)$ at every $Q \in M$, $\{dx^{i_1} \wedge \cdots \wedge dx^{i_p}\}$ is a basis for the p-forms. All the algebraic properties established in Chap. 26 apply to these p-forms at every point $Q \in M$.

The concept of a pullback has been mentioned a number of times in connection with linear maps. The most frequent use of pullbacks takes place in conjunction with the p-forms.

defining pullback for differential forms

Definition 28.5.2 Let M and N be manifolds and $\psi : M \to N$ a differentiable map. The **pullback map** on p-forms is the map $\psi^* : \Lambda^p(N) \to \Lambda^p(M)$ defined by

$$\psi^*\boldsymbol{\rho}(\mathbf{X}_1, \dots, \mathbf{X}_p) = \boldsymbol{\rho}(\psi_*\mathbf{X}_1, \dots, \psi_*\mathbf{X}_p) \quad \text{for } \boldsymbol{\rho} \in \Lambda^p(N).$$

For $p = 0$, i.e., for functions on M, $\psi^*\boldsymbol{\omega} \equiv \boldsymbol{\omega} \circ \psi$.

It can be shown that

$$\psi^*(\boldsymbol{\omega} \wedge \boldsymbol{\eta}) = \psi^*\boldsymbol{\omega} \wedge \psi^*\boldsymbol{\eta}, \qquad (\psi \circ \phi)^* = \phi^* \circ \psi^*. \quad (28.37)$$

Since $\boldsymbol{\omega}$ varies from point to point, we can define its derivatives. Recall that $T_0^0(M)$ is the collection of real-valued functions on M. Since the dual of $\mathbb{R}$ is $\mathbb{R}$, we conclude that $\Lambda^0(M)$, the collection of zero-forms, is the union of all real-valued functions on M. Also recall that if f is a zero-form, then df, the differential of f, is a one-form. Thus, the differential operator d creates a one-form from a zero-form. The fact that this can be generalized to p-forms is the subject of the next theorem (for a proof, see [Abra 88, pp. 111–112]).

Theorem 28.5.3 *For each point Q of M, there exists a neighborhood U and a unique operator $d : \Lambda^p(U) \to \Lambda^{p+1}(U)$, called the* ***exterior derivative*** *operator, such that for any $\boldsymbol{\omega} \in \Lambda^p(U)$ and $\boldsymbol{\eta} \in \Lambda^q(U)$,*

1. $d(\boldsymbol{\omega} + \boldsymbol{\eta}) = d\boldsymbol{\omega} + d\boldsymbol{\eta}$ *if $q = p$; otherwise the sum is not defined.*
2. $d(\boldsymbol{\omega} \wedge \boldsymbol{\eta}) = (d\boldsymbol{\omega}) \wedge \boldsymbol{\eta} + (-1)^p \boldsymbol{\omega} \wedge (d\boldsymbol{\eta})$; *this is called the* ***antiderivation*** *property of d with respect to the wedge product.*
3. $d(d\boldsymbol{\omega}) = 0$ *for any differential form $\boldsymbol{\omega}$; stated differently, $d \circ d = 0$.*
4. $df = (\partial_i f)dx^i$ *for any real-valued function f.*
5. d ***is natural with respect to*** *pullback; that is, $d_M \circ \psi^* = \psi^* \circ d_N$ for any differentiable map $\psi : M \to N$. Here d_M (d_N) is the exterior derivative operating on differential forms of M (N).*

exterior derivative and its antiderivation property

Example 28.5.4 Let $M = \mathbb{R}^3$ and $\boldsymbol{\omega} = a_i dx^i$ a 1-form on M. The exterior derivative of $\boldsymbol{\omega}$ is

$$d\boldsymbol{\omega} = (da_i) \wedge dx^i = \left(\partial_j a_i dx^j\right) \wedge dx^i = \sum_{j<i} (\partial_j a_i - \partial_i a_j) dx^j \wedge dx^i.$$

We see that the components of $d\boldsymbol{\omega}$ are the components of $\boldsymbol{\nabla} \times \mathbf{A}$ where $\mathbf{A} = (a_1, a_2, a_3)$. It follows that the curl of a vector in $\mathbb{R}^3$ is the exterior derivative of the 1-form constructed out of the components of the vector.

Example 28.5.5 In relativistic electromagnetic theory the electric and magnetic fields are combined to form the electromagnetic field tensor. This is a skew-symmetric tensor field of rank 2, which can be written as[8]

The homogeneous Maxwell's equations are written in terms of differential forms.

$$\begin{aligned} \mathsf{F} = &-E_x dt \wedge dx - E_y dt \wedge dy - E_z dt \wedge dz \\ &+ B_z dx \wedge dy - B_y dx \wedge dz + B_x dy \wedge dz, \end{aligned} \tag{28.38}$$

where t is the time coordinate and the units are such that c, the velocity of light, is equal to 1.

Let us take the exterior derivative of F. In the process, we use $df = (\partial_i f)dx^i$, $d(dx^i \wedge dx^j) = 0$, and in dE_i or dB_j we include only the terms that give a nonzero contribution:

$$\begin{aligned} d\mathsf{F} = &-\left(\frac{\partial E_x}{\partial y} dy + \frac{\partial E_x}{\partial z} dz\right) \wedge dt \wedge dx - \left(\frac{\partial E_y}{\partial x} dx + \frac{\partial E_y}{\partial z} dz\right) \wedge dt \wedge dy \\ &- \left(\frac{\partial E_z}{\partial x} dx + \frac{\partial E_z}{\partial y} dy\right) \wedge dt \wedge dz + \left(\frac{\partial B_z}{\partial t} dt + \frac{\partial B_z}{\partial z} dz\right) \wedge dx \wedge dy \\ &- \left(\frac{\partial B_y}{\partial t} dt + \frac{\partial B_y}{\partial y} dy\right) \wedge dx \wedge dz + \left(\frac{\partial B_x}{\partial t} dt + \frac{\partial B_x}{\partial x} dx\right) \wedge dy \wedge dz. \end{aligned}$$

[8] Note how in the wedge product, the first factor has a lower index (is an "earlier" coordinate) than the second factor. If this restriction is to be removed, we need to introduce a factor of $\frac{1}{2}$ for each component (see Example 28.5.12).

Collecting all similar terms and taking into account changes of sign due to the antisymmetry of the exterior products gives

$$
\begin{aligned}
d\mathbf{F} &= \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} + \frac{\partial B_z}{\partial t}\right) dt \wedge dx \wedge dy \\
&\quad + \left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} - \frac{\partial B_y}{\partial t}\right) dt \wedge dx \wedge dz \\
&\quad + \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} + \frac{\partial B_x}{\partial t}\right) dt \wedge dy \wedge dz \\
&\quad + \left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}\right) dx \wedge dy \wedge dz \\
&= \left[\left(\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t}\right)_z\right] dt \wedge dx \wedge dy \\
&\quad + \left[\left(\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t}\right)_y\right] dt \wedge dz \wedge dx \\
&\quad + \left[\left(\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t}\right)_x\right] dt \wedge dy \wedge dz + (\boldsymbol{\nabla} \cdot \mathbf{B}) dx \wedge dy \wedge dz.
\end{aligned}
$$

Each component of $d\mathbf{F}$ vanishes because of Maxwell's equations.

The example above shows that

> **Box 28.5.6** *The two homogeneous Maxwell's equations can be written as $d\mathbf{F} = 0$, where $\mathbf{F}$ is defined by Eq.* (28.38).

The exterior derivative is a very useful concept in the theory of differential forms, as illustrated in the preceding example. However, that is not the only differentiation available to the differential forms. We have already defined the Lie derivative for arbitrary tensors. Since differential forms are (antisymmetrized) linear combinations of covariant tensors, Lie differentiation is defined for them as well. In fact, since differential forms have no contravariant parts, one uses the pullback map F_t^* in the definition of the Lie derivative instead of F_{t*}^{-1}:

$$
L_{\mathbf{X}} \boldsymbol{\omega} = \left(F_t^*\right)^{-1} \frac{d}{dt} F_t^* \boldsymbol{\omega}. \tag{28.39}
$$

The two derivatives defined so far have the following convenient property, whose proof is left as an exercise for the reader:

Theorem 28.5.7 *The exterior derivative d is natural with respect to $L_{\mathbf{X}}$ (or commutes with $L_{\mathbf{X}}$) for $\mathbf{X} \in \mathcal{X}(M)$; that is, $d \circ L_{\mathbf{X}} = L_{\mathbf{X}} \circ d$.*

In the last chapter, we defined the interior product i_θ for p-vectors, where θ is a 1-form. With our shift of emphasis from p-vectors to p-forms in this chapter, we need to shift the role of vectors and forms.

Definition 28.5.8 Let $\mathbf{X}$ be a vector field and $\boldsymbol{\omega}$ a p-form on a manifold M. Then the interior product $i_{\mathbf{X}} : \Lambda^p(M) \to \Lambda^{p-1}(M)$ is defined as follows:

interior product of a vector field and a differential form

$$i_{\mathbf{X}}\boldsymbol{\omega}(\mathbf{X}_1, \dots, \mathbf{X}_{p-1}) = \boldsymbol{\omega}(\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_{p-1}).$$

If $\boldsymbol{\omega} \in \Lambda^0(M)$, i.e., if $\boldsymbol{\omega}$ is just a function, we set $i_{\mathbf{X}}\boldsymbol{\omega} = 0$. Another notation commonly used for $i_{\mathbf{X}}\boldsymbol{\omega}$ is $\mathbf{X} \rfloor \boldsymbol{\omega}$.

The interior product $i_{\mathbf{X}}$ has the antiderivation property of Theorem 27.0.2:

Theorem 28.5.9 *Let $\boldsymbol{\omega}$ be a p-form and $\boldsymbol{\eta}$ a q-form on a manifold M. Then*

$$i_{\mathbf{X}}(\boldsymbol{\omega} \wedge \boldsymbol{\eta}) = (i_{\mathbf{X}}\boldsymbol{\omega}) \wedge \boldsymbol{\eta} + (-1)^p \boldsymbol{\omega} \wedge (i_{\mathbf{X}}\boldsymbol{\eta}).$$

We have introduced three types of derivation on the algebra of differential forms: the exterior derivative, the Lie derivative, and the interior product. The following theorem connects all three derivations in a most useful way (see A[Abra 88, pp. 115–116]):

Theorem 28.5.10 *Let $\boldsymbol{\omega} \in \Lambda^p(M)$, $f \in \Lambda^0(M)$, and $\mathbf{X} \in \mathfrak{X}(M)$. Let $i_{\mathbf{X}} : \Lambda^p(M) \to \Lambda^{p-1}(M)$, $d : \Lambda^p(M) \to \Lambda^{p+1}(M)$, and $L_{\mathbf{X}} : \Lambda^p(M) \to \Lambda^p(M)$ be the interior product, the exterior derivative, and the Lie derivative, respectively. Then*

Relation between d, $L_{\mathbf{X}}$, and $i_{\mathbf{X}}$

1. $i_{\mathbf{X}} df = L_{\mathbf{X}} f$.
2. $L_{\mathbf{X}} = i_{\mathbf{X}} \circ d + d \circ i_{\mathbf{X}}$.
3. $L_{f\mathbf{X}}\boldsymbol{\omega} = f L_{\mathbf{X}}\boldsymbol{\omega} + df \wedge i_{\mathbf{X}}\boldsymbol{\omega}$.

If $\mathbf{X} = X^j \partial_j$ and $\boldsymbol{\omega} = \omega_{i_1 i_2 \dots i_{p+1}} dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_{p+1}}$, then the reader may verify that $i_{\mathbf{X}}\boldsymbol{\omega} = X^i \omega_{i i_1 \dots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p}$. In particular, we have the useful formula

$$\begin{aligned} & i_{\mathbf{X}}\big(dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_{p+1}}\big) \\ &\quad = X^j \delta^{i_1 i_2 \dots i_{p+1}}_{j\, j_1 \dots j_p} dx^{j_1} \wedge dx^{j_2} \wedge \cdots \wedge dx^{j_p} \\ &\quad = X^j \left(\sum_{\pi} \epsilon_\pi \delta^{i_1}_{\pi(j)} \delta^{i_2}_{\pi(j_1)} \dots \delta^{i_{p+1}}_{\pi(j_p)} \right) dx^{j_1} \wedge dx^{j_2} \wedge \cdots \wedge dx^{j_p}. \end{aligned} \tag{28.40}$$

Theorem 28.5.11 *For a p-form $\boldsymbol{\omega}$, we have*

$$\begin{aligned} & d\boldsymbol{\omega}(\mathbf{X}_1, \dots, \mathbf{X}_{p+1}) \\ &\quad = \sum_{i=1}^{p+1} (-1)^{i+1} \mathbf{X}_i \big(\boldsymbol{\omega}(\mathbf{X}_1, \dots, \hat{\mathbf{X}}_i, \dots, \mathbf{X}_{p+1})\big) \\ &\qquad + \sum_{1 \le i < j \le p+1} (-1)^{i+j} \boldsymbol{\omega}\big([\mathbf{X}_i, \mathbf{X}_j], \mathbf{X}_1, \dots, \hat{\mathbf{X}}_i, \dots, \hat{\mathbf{X}}_j, \dots, \mathbf{X}_{p+1}\big) \end{aligned}$$

where the circumflex on a symbol means that symbol is to be omitted.

Proof We use mathematical induction on p. From item 2 of Theorem 28.5.10, we have

$$L_{\mathbf{X}}\boldsymbol{\omega} = i_{\mathbf{X}}(d\boldsymbol{\omega}) + d(i_{\mathbf{X}}\boldsymbol{\omega})$$

or

$$(L_{\mathbf{X}}\boldsymbol{\omega})(\mathbf{X}_1,\dots,\mathbf{X}_p) = \underbrace{\big(i_{\mathbf{X}}(d\boldsymbol{\omega})\big)(\mathbf{X}_1,\dots,\mathbf{X}_p)}_{=d\boldsymbol{\omega}(\mathbf{X},\mathbf{X}_1,\dots,\mathbf{X}_p)} + \big(d(i_{\mathbf{X}}\boldsymbol{\omega})\big)(\mathbf{X}_1,\dots,\mathbf{X}_p)$$

or

$$d\boldsymbol{\omega}(\mathbf{X},\mathbf{X}_1,\dots,\mathbf{X}_p) = (L_{\mathbf{X}}\boldsymbol{\omega})(\mathbf{X}_1,\dots,\mathbf{X}_p) - \big(d(i_{\mathbf{X}}\boldsymbol{\omega})\big)(\mathbf{X}_1,\dots,\mathbf{X}_p).$$

For the first term on the right-hand side, we use Eq. (28.36). For the second term, we use the induction hypothesis because $i_{\mathbf{X}}\boldsymbol{\omega}$ is a $(p-1)$-form. A straightforward manipulation then leads to the desired result. □

Example 28.5.12 Let $\mathbf{p} = p_\alpha dx^\alpha$ be the momentum one-form and write the electromagnetic field tensor as[9] $\mathsf{F} = \frac{1}{2}F_{\alpha\beta}dx^\alpha \wedge dx^\beta$, where α and β run over the values 0, 1, 2, and 3 with 0 being the time index. Let

$$\frac{d\mathbf{p}}{d\tau} \equiv \left(\frac{dp_\alpha}{d\tau}\right)dx^\alpha$$

analysis of the Lorentz force law in the language of forms

be the derivative of momentum with respect to the proper time, τ. Also, let $\mathbf{u} = u^\beta\partial_\beta$ be the velocity four-vector of a charged particle. Then the **Lorentz force law** can be written simply as $d\mathbf{p}/d\tau = q\mathsf{F}(\mathbf{u}) \equiv -qi_{\mathbf{u}}\mathsf{F}$, where q is the electric charge of the particle whose 4-velocity is $\mathbf{u}$. Note that F, a two-form, contracts with $\mathbf{u}$, a vector, to give a one-form on the RHS. Thus, both sides are of the same type. Let us write this equation in component form:

$$\begin{aligned}\frac{dp_\alpha}{d\tau}dx^\alpha &= -q\frac{1}{2}F_{\alpha\beta}i_{\mathbf{u}}\big(dx^\alpha\wedge dx^\beta\big) = -\frac{1}{2}qF_{\alpha\beta}\big(u^\gamma\delta^{\alpha\beta}_{\gamma\mu}dx^\mu\big)\\ &= \frac{1}{2}qF_{\alpha\beta}u^\gamma\big(\delta^\alpha_\gamma\delta^\beta_\mu - \delta^\alpha_\mu\delta^\beta_\gamma\big)dx^\mu\\ &= \frac{1}{2}qF_{\alpha\beta}\big(u^\beta dx^\alpha - u^\alpha dx^\beta\big)\\ &= \frac{1}{2}q(F_{\alpha\beta}-F_{\beta\alpha})u^\beta dx^\alpha = \big(qF_{\alpha\beta}u^\beta\big)dx^\alpha.\end{aligned} \tag{28.41}$$

Equating the components on both sides, we get $dp_\alpha/d\tau = qF_{\alpha\beta}u^\beta$, which may be familiar to the reader. To make the equation even more familiar, consider the component $\alpha = 1$,

$$\frac{dp_1}{d\tau} = qF_{1\beta}u^\beta = q\big[F_{10}u^0 + F_{12}u^2 + F_{13}u^3\big], \tag{28.42}$$

[9] The factor $\frac{1}{2}$ is introduced here to avoid restricting the sum over α and β.

and recall that $u^\alpha = dx^\alpha/d\tau$, where

$$(d\tau)^2 = (dt)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 = (dt)^2(1-v^2)$$

and $\mathbf{v} = (dx^1/dt, dx^2/dt, dx^3/dt)$ is the 3-velocity of the particle. Since $x^0 = t$, we get

$$u^0 = \frac{dt}{d\tau} = \frac{1}{\sqrt{1-v^2}}, \qquad u^i = \frac{dx^i}{d\tau} = \frac{v_i}{\sqrt{1-v^2}} \quad \text{for } i = 1, 2, 3.$$

Substituting this in (28.42) and remembering that $F_{10} = -F_{01} = E_1$, $F_{12} = B_3$, and $F_{13} = -F_{31} = -B_2$, we obtain

$$\frac{dp_1}{dt\sqrt{1-v^2}} = q\left[E_1\frac{1}{\sqrt{1-v^2}} + B_3\frac{v_2}{\sqrt{1-v^2}} - B_2\frac{v_3}{\sqrt{1-v^2}}\right],$$

or

$$\frac{dp_1}{dt} = q\left[E_1 + (v_2B_3 - v_3B_2)\right] = \left[q(\mathbf{E} + \mathbf{v}\times\mathbf{B})\right]_1.$$

The other components are obtained similarly. Thus, in vector form we have

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}),$$

where $\mathbf{p}$ now represents the 3-momentum of the particle. This is the Lorentz force law for electromagnetism in its familiar form. Again, note the simplification offered by the language of forms.

A combination that is very useful is that of the exterior derivative and the Hodge star operator. Recall that the latter is defined by

$$*\left(dx^{i_1}\wedge\cdots\wedge dx^{i_p}\right) = \frac{1}{(m-p)!}\epsilon^{i_1\dots i_p}_{i_{p+1}\dots i_m}dx^{i_{p+1}}\wedge\cdots\wedge dx^{i_m}, \qquad (28.43)$$

where m is the dimension of the manifold.

Example 28.5.13 Let us calculate $*\mathsf{F}$ and $d(*\mathsf{F})$ where $\mathsf{F} = \frac{1}{2}F_{\alpha\beta}dx^\alpha \wedge dx^\beta$ is the electromagnetic field tensor. We have

$$*\mathsf{F} = *\left(\frac{1}{2}F_{\alpha\beta}dx^\alpha\wedge dx^\beta\right) = \frac{1}{2}F_{\alpha\beta}*\left(dx^\alpha\wedge dx^\beta\right) = \frac{1}{2}F_{\alpha\beta}\frac{1}{2!}\epsilon^{\alpha\beta}_{\mu\nu}dx^\mu\wedge dx^\nu$$

and

$$d(*\mathsf{F}) = d\left(\frac{1}{4}F_{\alpha\beta}\epsilon^{\alpha\beta}_{\mu\nu}dx^\mu\wedge dx^\nu\right) = \frac{1}{4}\epsilon^{\alpha\beta}_{\mu\nu}F_{\alpha\beta,\gamma}dx^\gamma\wedge dx^\mu\wedge dx^\nu,$$

where $F_{\alpha\beta,\gamma} \equiv \partial F_{\alpha\beta}/\partial x^\gamma$. We can now use the components $F_{j0} = E_j$, $F_{12} = B_3$, $F_{13} = -B_2$, and $F_{23} = B_1$ to write $d(*\mathsf{F})$ in terms of $\mathbf{E}$ and $\mathbf{B}$. After a long but straightforward calculation, we obtain

$$d(*\mathbf{F}) = \left[\left(\frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{B}\right)_z\right] dt \wedge dx \wedge dy + \left[\left(\frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{B}\right)_y\right] dt \wedge dz \wedge dx + \left[\left(\frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{B}\right)_x\right] dt \wedge dy \wedge dz + (\nabla \cdot \mathbf{E})\, dx \wedge dy \wedge dz. \tag{28.44}$$

The inhomogeneous pair of Maxwell's equations is

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} + 4\pi \mathbf{J}, \qquad \nabla \cdot \mathbf{E} = 4\pi\rho, \tag{28.45}$$

Maxwell's inhomogeneous equations in the language of forms

where ρ and $\mathbf{J}$ are charge and current densities, respectively. We can put these two densities together to form a four-current one-form with ρ as the zeroth component: $\mathbf{J} = J_\alpha dx^\alpha$. Thus,

$$\begin{aligned} *\mathbf{J} &= J_\alpha (*dx^\alpha) = J_\alpha \frac{1}{3!}\epsilon^\alpha_{\mu\nu\rho} dx^\mu \wedge dx^\nu \wedge dx^\rho \\ &= J_0 dx \wedge dy \wedge dz + J_x dt \wedge dy \wedge dz + J_y dt \wedge dz \wedge dx \\ &\quad + J_z dt \wedge dx \wedge dy \\ &= \rho dx \wedge dy \wedge dz - J^x dt \wedge dy \wedge dz - J^y dt \wedge dz \wedge dx \\ &\quad - J^z dt \wedge dx \wedge dy, \end{aligned} \tag{28.46}$$

where we have used the facts that $\rho = J^0 = J_0$ and $\mathbf{J} = (J^x, J^y, J^z) = -(J_x, J_y, J_z)$.

Comparing Eqs. (28.44), (28.45), and (28.46), we note that

Box 28.5.14 *In the language of forms, the inhomogeneous pair of Maxwell's equations has the simple appearance* $d(*\mathbf{F}) = 4\pi(*\mathbf{J})$.

Problem 28.15 shows that the relation $d^2\boldsymbol{\omega} = 0$ is equivalent—at least in $\mathbb{R}^3$—to the vanishing of the curl of the gradient and the divergence of the curl. It is customary in physics to try to go backwards as well, that is, given that $\nabla \times \mathbf{E} = 0$, to assume that $\mathbf{E} = \nabla f$ for some function f. Similarly, we want to believe that $\nabla \cdot \mathbf{B} = 0$ implies that $\mathbf{B} = \nabla \times \mathbf{A}$.

closed and exact forms

What is the analogue of the above statement for a general p-form? A form $\boldsymbol{\omega}$ that satisfies $d\boldsymbol{\omega} = 0$ is called a **closed form**. An **exact form** is one that can be written as the exterior derivative of another form. Thus, *every exact form is automatically closed*. This is the **Poincaré lemma**. The converse of this lemma is true only if the region of definition of the form is topologically simple, as explained in the following.

regions that are contractable to a point

Consider a p-form $\boldsymbol{\omega}$ defined on a region U of a manifold M. If all closed curves in U can be shrunk to a point in U without encountering any points at which $\boldsymbol{\omega}$ is ill-defined, we say that U is **contractable** to a point. If $\boldsymbol{\omega}$ is not defined for a point P on M, then any U that contains P is not contractable to a point. We can now state the converse of the Poincaré lemma (for a proof, see [Bish 80, p. 175]):

converse of the Poincaré lemma

Theorem 28.5.15 (Converse of the Poincaré lemma) *Let U be a region in a manifold M such that U is contractable to a point. Let $\boldsymbol{\omega}$ be a p-form on U such that $d\boldsymbol{\omega} = 0$. Then there exists a $(p-1)$-form $\boldsymbol{\eta}$ on U such that $\boldsymbol{\omega} = d\boldsymbol{\eta}$.*

Example 28.5.16 The electromagnetic field tensor $\mathbf{F} = \frac{1}{2} F_{\alpha\beta} dx^\alpha \wedge dx^\beta$ is a two-form that satisfies $d\mathbf{F} = 0$. The converse of the Poincaré lemma says that if $\mathbf{F}$ is well behaved in a region U of $\mathbb{R}^4$, then there must exist a one-form $\boldsymbol{\eta}$ such that $\mathbf{F} = d\boldsymbol{\eta}$.

Let us write this one-form in terms of coordinates as $\boldsymbol{\eta} = A_\alpha dx^\alpha$. Then $d\boldsymbol{\eta} = A_{\alpha,\beta} dx^\beta \wedge dx^\alpha$, and we have

$$\frac{1}{2} F_{\alpha\beta} dx^\alpha \wedge dx^\beta = A_{\beta,\alpha} dx^\alpha \wedge dx^\beta$$
$$\Rightarrow \quad \frac{1}{2}(F_{\alpha\beta} - A_{\beta,\alpha} + A_{\alpha,\beta}) dx^\alpha \wedge dx^\beta = 0.$$

Since $dx^\alpha \wedge dx^\beta$ are linearly independent *and their coefficients are antisymmetric*, each of the latter must vanish. Thus,

$$F_{\alpha\beta} = A_{\beta,\alpha} - A_{\alpha,\beta} = \frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta}.$$

The four-vector A^α is simply the four-potential of relativistic electromagnetic theory.

gauge invariance in the language of forms

Note that the $(p-1)$-form of Theorem 28.5.15 is not unique. In fact, if $\boldsymbol{\alpha}$ is any $(p-2)$-form, then $\boldsymbol{\omega}$ can be written as

$$\boldsymbol{\omega} = d(\boldsymbol{\eta} + d\boldsymbol{\alpha})$$

because $d(d\boldsymbol{\alpha})$ is identical to zero. This freedom of choice in selecting $\boldsymbol{\eta}$ is called **gauge invariance**, and its generalization plays an important role in the physics of fundamental interactions.[10]

Historical Notes

Jules Henri Poincaré (1854–1912): The development of mathematics in the nineteenth century began under the shadow of a giant, Carl Friedrich Gauss; it ended with the domination by a genius of similar magnitude, Henri Poincaré. Both were universal mathematicians in the supreme sense, and both made important contributions to astronomy and mathematical physics. If Poincaré's discoveries in number theory do not equal those of Gauss, his achievements in the theory of functions are at least on the same level—even when one takes into account the theory of elliptic and modular functions, which must be credited to Gauss and which represents in that field his most important discovery, although it was not published during his lifetime. If Gauss was the initiator in the theory

Jules Henri Poincaré
1854–1912

[10]Gauge invariance and gauge theories are discussed in detail in Chap. 35.

of differentiable manifolds, Poincaré played the same role in *algebraic topology*. Finally, Poincaré remains the most important figure in the theory of differential equations and the mathematician who after Newton did the most remarkable work in celestial mechanics. Both Gauss and Poincaré had very few students and liked to work alone; but the similarity ends there. Where Gauss was very reluctant to publish his discoveries, Poincaré's list of papers approaches five hundred, which does not include the many books and lecture notes he published as a result of his teaching at the Sorbonne.

Poincaré's parents both belonged to the upper middle class, and both their families had lived in Lorraine for several generations. His paternal grandfather had two sons: Léon, Henri's father, was a physician and a professor of medicine at the University of Nancy; Antoine had studied at the École Polytechnique and rose to high rank in the engineering corps. One of Antoine's sons, Raymond, was several times prime minister and was president of the French Republic during World War I; the other son, Lucien, occupied high administrative functions in the university. Poincaré's mathematical ability became apparent while he was still a student in the *lycée*. He won first prizes in the *concours génération* (a competition among students from all French *lycées*) and in 1873 entered the École Polytechnique at the top of his class; his professor at Nancy is said to have referred to him as a "monster of mathematics." After graduation, he followed courses in engineering at the École des Mines and worked briefly as an engineer while writing his thesis for the doctorate in mathematics which he obtained in 1879. Shortly afterward he started teaching at the University of Caen, and in 1881 he became a professor at the University of Paris, where he taught until his untimely death in 1912. At the early age of thirty-three he was elected to the Académie des Sciences and in 1908 to the Académie Française. He was also the recipient of innumerable prizes and honors both in France and abroad.

Before he was thirty years of age, Poincaré became world famous with his epoch-making discovery of the "automorphic functions" of one complex variable (or, as he called them, the "fuchsian" and "kleinean" functions). Much has been written on the "competition" between C.F. Klein and Poincaré in the discovery of automorphic functions. However, Poincaré's ignorance of the mathematical literature when he started his researches is almost unbelievable. He hardly knew anything on the subject beyond Hermite's work on the modular functions; he certainly had never read Riemann, and by his own account had not even heard of the "Dirichlet principle," which he was to use in such imaginative fashion a few years later. Nevertheless, Poincaré's idea of associating a fundamental domain to any fuchsian group does not seem to have occurred to Klein, nor did the idea of "using" non-Euclidean geometry, which is never mentioned in his papers on modular functions up to 1880.

Poincaré was one of the few mathematicians of his time who understood and admired the work of Lie and his continuators on "continuous groups," and in particular the only mathematician who in the early 1900s realized the depth and scope of E. Cartan's papers. In 1899 Poincaré proved what is now called the *Poincaré–Birkhoff–Witt theorem* which has become fundamental in the modern theory of Lie algebras. The theory of differential equations and its applications to dynamics was clearly at the center of Poincaré's mathematical thought; from his first (1878) to his last (1912) paper, he attacked the theory from all possible angles and very seldom let a year pass without publishing a paper on the subject. The most extraordinary production of Poincaré's, also dating from his prodigious period of creativity (1880–1883) (reminding us of Gauss's *Tagebuch* of 1797–1801), is the qualitative theory of differential equations. It is one of the few examples of a mathematical theory that sprang apparently out of nowhere and that almost immediately reached perfection in the hands of its creator. Everything was new in the first two of the four big papers that Poincaré published on the subject between 1880 and 1886.

For more than twenty years Poincaré lectured at the Sorbonne on mathematical physics; he gave himself to that task with his characteristic thoroughness and energy, with the result that he became an expert in practically all parts of theoretical physics, and published more than seventy papers and books on the most varied subjects, with a predilection for the theories of light and of electromagnetic waves. On two occasions he played an important part in the development of the new ideas and discoveries that revolutionized physics at the end of the nineteenth century. His remark on the possible connection between X-rays and the phenomenon of phosphorescence was the starting point of H. Becquerel's experiments that led him to the discovery of radioactivity. On the other hand, Poincaré was active from 1899 on in the discussions concerning Lorentz's theory of the electron; Poincaré was the first to observe that the Lorentz transformations form a group; and many physicists consider that Poincaré shares with Lorentz and Einstein the credit for the invention of the special theory of relativity. The main leitmotiv of Poincaré's mathematical work is clearly the idea of "continuity": Whenever he attacks a problem in analysis, we almost immediately see him investigating what happens when the conditions of the problem are allowed to vary continuously. He was therefore bound to encounter at every turn what we now call *topological* problems. He himself said in 1901, "Every problem I had attacked led me to *Analysis situs*," particularly the researches on differential equations and on the periods of multiple integrals. Starting in 1894 he inaugurated in a remarkable series of six papers—written during a period of ten years—the modern methods of algebraic topology.

Whereas Poincaré has been accused of being too conservative in physics, he certainly was very open-minded regarding new mathematical ideas. The quotations in his papers show that he read extensively, if not systematically, and was aware of all the latest developments in practically every branch of mathematics. He was probably the first mathematician to use Cantor's theory of sets in analysis. Up to a certain point, he also looked with favor on the axiomatic trend in mathematics, as it was developing toward the end of the nineteenth century, and he praised Hilbert's *Grundlagen der Geometrie*. However, he obviously had a blind spot regarding the formalization of mathematics, and poked fun repeatedly at the efforts of the disciples of Peano and Russell in that direction; but, somewhat paradoxically, his criticism of the early attempts of Hilbert was probably the starting point of some of the most fruitful of the later developments of metamathematics. Poincaré stressed that Hilbert's point of view of defining objects by a system of axioms was admissible only if one could prove a priori that such a system did not imply contradiction, and it is well known that the proof of noncontradiction was the main goal of the theory that Hilbert founded after 1920. Poincaré seems to have been convinced that such attempts were hopeless, and K. Gödel's theorem proved him right.

28.6 Integration on Manifolds

integration of differential forms in $\mathbb{R}^n$

We mentioned in Chap. 26 that certain exterior products are interpreted as volume elements. We now exploit this notion and define integration on manifolds. Starting with $\mathbb{R}^n$, considered as a manifold, we define the integral of an n-form $\boldsymbol{\omega}$ as follows. Choose a coordinate system $\{x^i\}_{i=1}^n$ in $\mathbb{R}^n$, write $\boldsymbol{\omega} = f\,dx^1 \wedge \cdots \wedge dx^n$, and define the integral of the n-form as

$$\int_{\mathbb{R}^n, x} \boldsymbol{\omega} \equiv \int_{\mathbb{R}^n} f\big(x^1, \ldots, x^n\big) dx^1 \ldots dx^n,$$

where to avoid dealing with infinities, one assumes that f vanishes outside a bounded region. The second symbol in the lower part of the integral sign indicates the variables of integration. Let us now change the coordinates, say to $\{y^j\}_{j=1}^n$. Using Eq. (28.17), which gives the transformation rule for 1-forms when changing coordinates, and Eq. (2.32), which defines the determinant in terms of n-forms, we obtain

$$\boldsymbol{\omega} = f\,dx^1 \wedge \cdots \wedge dx^n = f \det\left(\frac{\partial x^i}{\partial y^j}\right) dy^1 \wedge \cdots \wedge dy^n,$$

where f is now understood to be a function of the y's through the x's. So, in terms of the new coordinates, the integral becomes

$$\int_{\mathbb{R}^n, y} \boldsymbol{\omega} = \int_{\mathbb{R}^n} f\big(x^1(\mathbf{y}), \dots, x^n(\mathbf{y})\big) \det\left(\frac{\partial x^i}{\partial y^j}\right) dy^1 \wedge \cdots \wedge dy^n.$$

If we had the absolute value of the Jacobian in the integral, the two sides would be equal. So, all we can say at this point is

$$\int_{\mathbb{R}^n, y} \boldsymbol{\omega} = \pm \int_{\mathbb{R}^n, x} \boldsymbol{\omega}.$$

This discussion is analogous to our discussion of orientation in vector spaces (see Sect. 26.3.1).

We therefore distinguish between two kinds of coordinate transformations: If the Jacobian determinant is positive, we say that the coordinate transformation is **orientation preserving**. Otherwise, the transformation is called **orientation reversing**.

Our ability to integrate functions on $\mathbb{R}^n$ depends crucially on the fact that volume elements do not change sign at any point of $\mathbb{R}^n$. If this were not so, we could find a finite (albeit small) region of space—in the vicinity of the point at which the volume element changes sign—whose volume would be zero. This property of $\mathbb{R}^n$ is the content of the following:

orientable manifolds

Definition 28.6.1 A manifold M of dimension n is called **orientable** if it has a nowhere vanishing n-form.

Any two nonvanishing n-forms $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ on an orientable manifold are related by a nowhere-vanishing function: $\boldsymbol{\omega}' = h\boldsymbol{\omega}$. Clearly, h has to be either positive or negative everywhere. $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ are said to be *equivalent* if h is positive. Thus, the nonvanishing n-forms on an orientable manifold fall into two classes, all members of each class being equivalent to one another, and a member of one class being related to a member of the other class via a negative function. Each class is called an **orientation** on M.

Given an orientation, an n-form $\boldsymbol{\omega}$, and a chart $\{U_\alpha, \phi_\alpha\}$ on M, we define

$$\int_M \boldsymbol{\omega} \equiv \sum_\alpha \int_{\mathbb{R}^n} \big(\phi_\alpha^{-1}\big)^*(\boldsymbol{\omega}|_\alpha), \tag{28.47}$$

where $(\phi_\alpha^{-1})^*$ is the pullback of $\phi_\alpha^{-1} : \mathbb{R}^n \to M$, so that it maps n-forms on M to n-forms on $\mathbb{R}^n$; $\boldsymbol{\omega}|_\alpha$ is the restriction of $\boldsymbol{\omega}$ to U_α, and the sum over α is assumed to exist. This amounts to saying that the region in M on which $\boldsymbol{\omega}$ is defined is finite, or that $\boldsymbol{\omega}$ has **compact support**.

Compact support

We note that the RHS of Eq. (28.47) is an integration on $\mathbb{R}^n$ that appears to depend on the choice of coordinate functions. However, it can be shown that the integral is independent of such choice. In practice, one chooses a coordinate patch and transfers the integration to $\mathbb{R}^n$, where the process is familiar.

"volume" of a manifold

If we choose a coordinate patch $\{x^i\}_{i=1}^n$ and integrate $dx^1 \wedge \cdots \wedge dx^n$ according to Eq. (28.47), we obtain the "volume" of the manifold M. If M is compact, this volume will be finite.[11]

[11] Recall from Chap. 17 that a subset of $\mathbb{R}^n$ is compact iff it is closed and bounded. It is a good idea to keep this in mind as a paradigm of compact spaces.

Theorem 28.6.2 (Stokes' theorem) *Let M be an oriented n-manifold. Let $\boldsymbol{\omega}$ be an $(n-1)$-form with compact support. Then* Stokes' Theorem

$$\int_M d\boldsymbol{\omega} = 0$$

Proof From Eq. (28.47), we have

$$\int_M d\boldsymbol{\omega} = \sum_\alpha \int_{\mathbb{R}^n} \big(\phi_\alpha^{-1}\big)^* (d\boldsymbol{\omega}|_\alpha) = \sum_\alpha \int_{\mathbb{R}^n} d\big(\big(\phi_\alpha^{-1}\big)^* \boldsymbol{\omega}|_\alpha\big)$$

where in the last equality, we used item 5 of Theorem 28.5.3. Now $(\phi_\alpha^{-1})^*\boldsymbol{\omega}|_\alpha$ is an $(n-1)$-from on $\mathbb{R}^n$. If $\boldsymbol{\beta}$ is such a form, it can be written as

$$\boldsymbol{\beta} = \beta_i dx^1 \wedge \cdots \wedge d\hat{x}^i \wedge \cdots \wedge dx^n$$

and $d\boldsymbol{\beta} = \sum_{i=1}^n (-1)^{i-1} \partial_i \beta_i dx^1 \wedge \cdots \wedge dx^n$. Therefore,

$$\begin{aligned}
\int_{\mathbb{R}^n} d\boldsymbol{\beta} &= \sum_{i=1}^n \int_{\mathbb{R}^n} (-1)^{i-1} \partial_i \beta_i dx^1 \wedge \cdots \wedge dx^n \\
&\equiv \sum_{i=1}^n (-1)^{i-1} \int_{\mathbb{R}^n} \frac{\partial \beta_i}{\partial x^i} dx^1 \dots dx^n \\
&= \sum_{i=1}^n (-1)^{i-1} \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} \frac{\partial \beta_i}{\partial x^i} dx^i \right) dx^1 \dots d\hat{x}^i \dots dx^n \\
&= \sum_{i=1}^n (-1)^{i-1} \int_{\mathbb{R}^{n-1}} \big(\beta_i \big|_{x^i=-\infty}^{x^i=\infty}\big) dx^1 \dots d\hat{x}^i \dots dx^n .
\end{aligned}$$

The term in parentheses is zero because $\boldsymbol{\beta}$ has compact support and all its components must vanish at infinity. □

A manifold may have a **boundary** ∂M, which is an $(n-1)$-dimensional submanifold of M, every point of which has a coordinate neighborhood in which one of the coordinates is zero. For example, the xy-plane is the boundary of the lower space on which $z = 0$. As another example, consider an open set U of an n-manifold M. Then U is also an n-manifold, and its boundary ∂U is an $(n-1)$-dimensional manifold. If M (and therefore, U) is oriented, then ∂U inherits an orientation from U. There is another version of the Stokes' Theorem for manifolds with boundary, which we state without proof.

Theorem 28.6.3 *Let U be an oriented n-manifold with boundary ∂U. Let $\boldsymbol{\omega}$ be an $(n-1)$-form with compact support. Then* Stokes' Theorem for manifolds with boundary

$$\int_U d\boldsymbol{\omega} = \int_{\partial U} \boldsymbol{\omega}$$

Combining the exterior differential with the Hodge star operator, we get a useful quantity.

Definition 28.6.4 The **codifferential** δ is a map $\delta : \Lambda^p(M) \to \Lambda^{p-1}(M)$ given by

Codifferential defined

$$\delta\boldsymbol{\omega} = (-1)^{\nu+1}(-1)^{n(p+1)} * d * \boldsymbol{\omega}$$

where n is the dimension of M. If $\boldsymbol{\omega}$ is a 0-form, i.e., a function f, then the definition leads to $\delta f = 0$. Furthermore, since $** = \pm 1$, $\delta^2 \equiv 0$.

If M has a metric, i.e., a nondegenerate symmetric bilinear form $\mathbf{g}$ defined smoothly on each point of M, and $\mathbf{g}$ does not vary over M, then we have the following:

Proposition 28.6.5 *If $\boldsymbol{\omega} = \frac{1}{p!}\omega_{i_1\dots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p}$ is a p-form on an n-manifold M with constant metric $\mathbf{g}$, then*

$$\delta\boldsymbol{\omega} = \frac{(-1)^p}{(p-1)!}\partial_i \omega_{i_1\dots i_{p-1}}{}^{i}\, dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}$$

where $\partial_i \omega_{i_1\dots i_{p-1}}{}^{i} = g^{ii_p}\partial_i \omega_{i_1\dots i_{p-1}i_p} = \partial_i(g^{ii_p}\omega_{i_1\dots i_{p-1}i_p})$.

Proof Start with the definition of δ and the Hodge star operator:

$$\begin{aligned}\delta\boldsymbol{\omega} &= (-1)^{\nu+1}(-1)^{n(p+1)} * d * \boldsymbol{\omega}\\ &= \frac{(-1)^{\nu+1}(-1)^{n(p+1)}}{p!(n-p)!} * d\big(\omega^{i_1\dots i_p}\epsilon_{i_1\dots i_n} dx^{i_{p+1}} \wedge \cdots \wedge dx^{i_n}\big).\end{aligned}$$

Now note that d differentiates only the function $\omega^{i_1\dots i_p}$ because $d^2 = 0$. Differentiating and applying $*$ afterwards, we get

$$\begin{aligned}\delta\boldsymbol{\omega} &= \frac{(-1)^{\nu+1}(-1)^{n(p+1)}}{p!(n-p)!}\epsilon_{i_1\dots i_n}\partial_i\omega^{i_1\dots i_p} * \big(dx^i \wedge dx^{i_{p+1}} \wedge \cdots \wedge dx^{i_n}\big)\\ &= \frac{(-1)^{\nu+1}(-1)^{n(p+1)}}{p!(n-p)!}\epsilon_{i_1\dots i_n}\partial_i\omega^{i_1\dots i_p}\\ &\quad\times \frac{1}{(p-1)!} g^{ij_p} g^{i_{p+1}j_{p+1}} \dots g^{i_n j_n} \epsilon_{j_p\dots j_n j_1\dots j_{p-1}} dx^{j_1} \wedge \cdots \wedge dx^{j_{p-1}}.\end{aligned}$$

Rearranging the indices of the ϵ,

$$\epsilon_{j_p\dots j_n j_1\dots j_{p-1}} = (-1)^{(n-p+1)(p-1)}\epsilon_{j_1\dots j_n},$$

manipulating the powers of -1, noting that $A^k B_k = A_k B^k$, and using the fact that g^{lm} are constant,[12] we rewrite the last expression as

[12]Reader, see where this fact is used!

$$\begin{aligned}\delta\boldsymbol{\omega} &= \frac{(-1)^{\nu+1}(-1)^{(p-1)^2}}{p!(n-p)!(p-1)!}\epsilon^{i_1\dots i_n}\epsilon_{j_1\dots j_n}\partial_i\omega_{i_1\dots i_p}g^{ij_p}g^{j_{p+1}}_{i_{p+1}}\dots g^{j_n}_{i_n}dx^{j_1}\wedge\cdots\\&\quad\wedge dx^{j_{p-1}}\\&= -\frac{(-1)^{p-1}}{p!(n-p)!(p-1)!}\delta^{i_1\dots i_n}_{j_1\dots j_n}\partial_i\omega_{i_1\dots i_p}g^{ij_p}\delta^{j_{p+1}}_{i_{p+1}}\dots\delta^{j_n}_{i_n}dx^{j_1}\wedge\cdots\\&\quad\wedge dx^{j_{p-1}},\end{aligned}$$

where we used Eq. (26.44), the fact that $(-1)^{m^2}=(-1)^m$ for any integer m and that $g^i_j=\delta^i_j$. The last expression is now reduced to

$$\begin{aligned}\delta\boldsymbol{\omega} &= -\frac{(-1)^{p-1}}{p!(n-p)!(p-1)!}\delta^{i_1\dots i_p i_{p+1}\dots i_n}_{j_1\dots j_p i_{p+1}\dots i_n}\partial_i\omega_{i_1\dots i_p}g^{ij_p}dx^{j_1}\wedge\cdots\wedge dx^{j_{p-1}}\\&= -\frac{(-1)^{p-1}}{p!(p-1)!}\delta^{i_1\dots i_p}_{j_1\dots j_p}\partial_i\omega_{i_1\dots i_p}g^{ij_p}dx^{j_1}\wedge\cdots\wedge dx^{j_{p-1}}\\&= -\frac{(-1)^{p-1}}{(p-1)!}\partial_i\omega_{j_1\dots j_p}g^{ij_p}dx^{j_1}\wedge\cdots\wedge dx^{j_{p-1}},\end{aligned}$$

where we used Eq. (26.47) in the first equality and Eq. (26.50) in the last. □

If M has a metric, then a metric $\tilde{\mathbf{g}}$ can be defined on $\Lambda^p(M)$ in exact analogy with Definition 26.3.13, and we have the following:

Theorem 28.6.6 *Let M be an oriented n-manifold with volume element $\boldsymbol{\mu}$ and metric $\mathbf{g}$. Let $\boldsymbol{\alpha}\in\Lambda^p(M)$ and $\boldsymbol{\beta}\in\Lambda^{p+1}(M)$ such that $\boldsymbol{\alpha}\wedge*\boldsymbol{\beta}$ has compact support. Then*

$$\int_M\tilde{\mathbf{g}}(\boldsymbol{\alpha},\delta\boldsymbol{\beta})\boldsymbol{\mu}=\int_M\tilde{\mathbf{g}}(d\boldsymbol{\alpha},\boldsymbol{\beta})\boldsymbol{\mu}$$

Proof From Theorem 26.6.4, we have

$$\begin{aligned}\tilde{\mathbf{g}}(\boldsymbol{\alpha},\delta\boldsymbol{\beta})\boldsymbol{\mu} &= \boldsymbol{\alpha}\wedge*\delta\boldsymbol{\beta}=\boldsymbol{\alpha}\wedge*\big((-1)^{\nu+1}(-1)^{n(p+2)}*d*\boldsymbol{\beta}\big)\\&= (-1)^{\nu+1}(-1)^{n(p+2)}(-1)^{\nu}(-1)^{p(n-p)}\boldsymbol{\alpha}\wedge d*\boldsymbol{\beta}\\&= -(-1)^p\boldsymbol{\alpha}\wedge d*\boldsymbol{\beta}\end{aligned}$$

because $(-1)^{p^2}=(-1)^{-p^2}=(-1)^p$ for any integer p. Hence,

$$\tilde{\mathbf{g}}(d\boldsymbol{\alpha},\boldsymbol{\beta})\boldsymbol{\mu}-\tilde{\mathbf{g}}(\boldsymbol{\alpha},\delta\boldsymbol{\beta})\boldsymbol{\mu}=d\boldsymbol{\alpha}\wedge*\beta+(-1)^p\boldsymbol{\alpha}\wedge d*\boldsymbol{\beta}=d(\boldsymbol{\alpha}\wedge*\beta)$$

and the integral over M of the right-hand side is zero by Stokes' Theorem. □

28.7 Symplectic Geometry

Mechanics stimulated a great deal of dialogue between physics and mathematics in the latter part of the nineteenth century and the beginning of the twentieth. The branch of mathematics that benefited the most out of this dialog is the theory of differentiable manifolds, whose tribute back to mechanics has been the most beautiful language in which the latter can express itself, the language of symplectic geometry. All the discussion of symplectic vector spaces of the last chapter can be carried over to the tangent spaces of a manifold and patched together by the differentiable structure of the manifold.

symplectic form, symplectic structure, and symplectic manifold defined

Definition 28.7.1 A **symplectic form** (or a **symplectic structure**) on a manifold M is a nondegenerate, closed 2-form $\boldsymbol{\omega}$ on M. A **symplectic manifold** $(M, \boldsymbol{\omega})$ is a manifold M together with a symplectic form $\boldsymbol{\omega}$ on M. We define the map $\flat : \mathcal{X}(M) \to \mathcal{X}^*(M)$ by

$$\flat(\mathbf{X}) \equiv \mathbf{X}^\flat = i_\mathbf{X}\boldsymbol{\omega} = \boldsymbol{\omega}^\flat(\mathbf{X})$$

and the map $\sharp : \mathcal{X}^*(M) \to \mathcal{X}(M)$ as the inverse of $\flat$.

Chapter 26 identified some special basis, the canonical basis, in which the symplectic form of a symplectic vector space took on a simple expression. The analogue of such a basis exists in a symplectic manifold. The reader should keep in mind that this existence is not automatic, because although one can find such bases at every point of the manifold, the smooth patching up of all such bases to cover the entire manifold is not trivial and is the content of the following important theorem, which we state without proof (see [Abra 85, p. 175]):

Darboux theorem

Theorem 28.7.2 (Darboux) *Suppose $\boldsymbol{\omega}$ is a 2-form on a $2n$-dimensional manifold M. Then $d\boldsymbol{\omega} = 0$ if and only if there is a chart (U, φ) at each $P \in M$ such that $\varphi(P) = \mathbf{0}$ and*

$$\boldsymbol{\omega} = \sum_{i=1}^{n} dx^i \wedge dy^i,$$

where $x^1, \ldots, x^n, y^1, \ldots, y^n$ are coordinates on U. Furthermore, on such a chart, the volume element μ_ω is

$$\mu_\omega = dx^1 \wedge \cdots \wedge dx^n \wedge dy^1 \wedge \cdots \wedge dy^n.$$

symplectic charts, canonical coordinates, and canonical transformations

Definition 28.7.3 The charts guaranteed by Darboux's theorem are called **symplectic charts**, and the coordinates x^i, y^i are called **canonical coordinates**. If $(M, \boldsymbol{\omega})$ and $(N, \boldsymbol{\rho})$ are symplectic manifolds, then a $\mathcal{C}^\infty$ map $f : M \to N$ is called **symplectic**, or a **canonical transformation**, if $f^*\boldsymbol{\rho} = \boldsymbol{\omega}$.

Coordinate representation of sharp and flat maps

Example 28.7.4 In this example, we derive a formula that gives the action of $\boldsymbol{\omega}^\flat$ and $\boldsymbol{\omega}^\sharp$ in terms of components of vectors and 1-forms in canonical coordinates. Let

$$\mathbf{Z} \equiv X^i \frac{\partial}{\partial x^i} + Y^i \frac{\partial}{\partial y^i}$$

be a vector field. When $\boldsymbol{\omega}^\flat$ acts on $\mathbf{Z}$, it gives a 1-form, which we write as $\boldsymbol{\omega}^\flat(\mathbf{Z}) \equiv U_k dx^k + W_k dy^k$. To find the unknowns U_k and W_k, we let both sides act on coordinate basis vectors. For the RHS, we get

$$\left(U_k dx^k + W_k dy^k\right)\left(\frac{\partial}{\partial x^j}\right) = U_k \underbrace{dx^k\left(\frac{\partial}{\partial x^j}\right)}_{=\delta^k_j} + W_k \underbrace{dy^k\left(\frac{\partial}{\partial x^j}\right)}_{=0} = U_j$$

and

$$\left(U_k dx^k + W_k dy^k\right)\left(\frac{\partial}{\partial y^j}\right) = W_j.$$

For the LHS, we obtain

$$\begin{aligned}
&\left[\boldsymbol{\omega}^\flat\left(X^i \frac{\partial}{\partial x^i} + Y^i \frac{\partial}{\partial y^i}\right)\right]\left(\frac{\partial}{\partial x^j}\right) \\
&\quad = X^i\left[\boldsymbol{\omega}^\flat\left(\frac{\partial}{\partial x^i}\right)\right]\left(\frac{\partial}{\partial x^j}\right) + Y^i\left[\boldsymbol{\omega}^\flat\left(\frac{\partial}{\partial y^i}\right)\right]\left(\frac{\partial}{\partial x^j}\right) \\
&\quad = X^i \boldsymbol{\omega}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) + Y^i \boldsymbol{\omega}\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial x^j}\right).
\end{aligned}$$

But

$$\begin{aligned}
\boldsymbol{\omega}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) &= \left(\sum_{k=1}^n dx^k \wedge dy^k\right)\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \\
&= \sum_{k=1}^n dx^k\left(\frac{\partial}{\partial x^i}\right) \underbrace{dy^k\left(\frac{\partial}{\partial x^j}\right)}_{=0} \\
&\quad - \sum_{k=1}^n dx^k\left(\frac{\partial}{\partial x^j}\right) \underbrace{dy^k\left(\frac{\partial}{\partial x^i}\right)}_{=0} = 0
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\omega}\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial x^j}\right) &= \left(\sum_{k=1}^n dx^k \wedge dy^k\right)\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial x^j}\right) \\
&= \sum_{k=1}^n \underbrace{dx^k\left(\frac{\partial}{\partial y^i}\right)}_{=0} \underbrace{dy^k\left(\frac{\partial}{\partial x^j}\right)}_{=0} \\
&\quad - \sum_{k=1}^n \underbrace{dx^k\left(\frac{\partial}{\partial x^j}\right)}_{=\delta^k_j} \underbrace{dy^k\left(\frac{\partial}{\partial y^i}\right)}_{=\delta^k_i} = -\delta^i_j.
\end{aligned}$$

It follows that

$$\left[\boldsymbol{\omega}^\flat\left(X^i\frac{\partial}{\partial x^i}+Y^i\frac{\partial}{\partial y^i}\right)\right]\left(\frac{\partial}{\partial x^j}\right)=-Y^j.$$

Similarly,

$$\left[\boldsymbol{\omega}^\flat\left(X^i\frac{\partial}{\partial x^i}+Y^i\frac{\partial}{\partial y^i}\right)\right]\left(\frac{\partial}{\partial y^j}\right)=X^j.$$

Therefore,

$$\boldsymbol{\omega}^\flat\left(X^i\frac{\partial}{\partial x^i}+Y^i\frac{\partial}{\partial y^i}\right)=-Y^j dx^j+X^j dy^j, \tag{28.48}$$

where a summation over repeated indices is understood.

If we multiply both sides of this equation by $\boldsymbol{\omega}^\sharp$ on the left and recall that $\boldsymbol{\omega}^\sharp\boldsymbol{\omega}^\flat=\mathbf{1}$, we obtain the following equation for the action of $\boldsymbol{\omega}^\sharp$:

$$\boldsymbol{\omega}^\sharp\left(X^j dx^j+Y^j dy^j\right)=Y^i\frac{\partial}{\partial x^i}-X^i\frac{\partial}{\partial y^i}. \tag{28.49}$$

Equations (28.48) and (28.49) are very useful in Hamiltonian mechanics.

Our discussion of symplectic transformations of symplectic vector spaces showed that such maps are necessarily isomorphisms. Applied to the present situation, this means that if $f:M\to N$ is symplectic, then $f_*:\mathcal{T}_P(M)\to\mathcal{T}_{f(P)}(N)$ is an isomorphism. Theorem 28.3.2, the inverse mapping theorem, now gives the following theorem.

Theorem 28.7.5 *If $f:M\to N$ is symplectic, then it is a local diffeomorphism.*

Example 28.7.6 Hamiltonian mechanics takes place in the phase space of a system. The phase space is derived from the configuration space as follows. Let $(q_1,\dots,q_n)$ be the generalized coordinates of a mechanical system. They describe an n-dimensional manifold N. The dynamics of the system is described by the (time-independent) Lagrangian L, which is a function of $(q^i,\dot{q}^i)$. But $\dot{q}^i$ are the components of a vector at $(q_1,\dots,q_n)$ [see Eq. (28.13) and replace γ^i with x^i]. Thus, in the language of manifold theory, a Lagrangian is a function on the tangent bundle, $L:T(N)\to\mathbb{R}$.

from Lagrangian to Hamiltonian in the language of differential forms

The Hamiltonian is obtained from the Lagrangian by a Legendre transformation: $H=\sum_{i=1}^n p_i\dot{q}^i-L$. The first term can be thought of as a pairing of an element of the tangent space with its dual. In fact, if P has coordinates $(q_1,\dots,q_n)$, then $\dot{\mathbf{q}}\equiv\dot{q}^i\partial_i\in\mathcal{T}_P(N)$ (with the Einstein summation convention enforced), and if we pair this with the dual vector $p_j dx^j\in\mathcal{T}_P^*(N)$, we obtain the first term in the definition of the Hamiltonian. The effect of the Legendre transformation is to replace $\dot{q}^i$ by p_i as the second set of independent variables. This has the effect of replacing $T(N)$ with $T^*(N)$. Thus

Box 28.7.7 *The manifold of Hamiltonian dynamics, or the phase space, is $T^*(N)$, with coordinates (q^i, p_i) on which the Hamiltonian $H : T^*(N) \to \mathbb{R}$ is defined.*

$T^*(N)$ is $2n$-dimensional; so it has the potential of becoming a symplectic manifold. In fact, it can be shown that[13] the 2-form suggested by Darboux's theorem,

symplectic 2-form of $T^*(N)$

$$\boldsymbol{\omega} \equiv \sum_{i=1}^{n} dq^i \wedge dp_i, \tag{28.50}$$

is nondegenerate, and therefore a symplectic form for $T^*(N)$.

The phase space, equipped with a symplectic form, turns into a geometric arena in which Hamiltonian mechanics unfolds. We saw in the above example that a Hamiltonian is a function on the phase space. More generally, if $(M, \boldsymbol{\omega})$ is a symplectic manifold, a Hamiltonian H is a real-valued function, $H : M \to \mathbb{R}$. Given a Hamiltonian, one can define a vector field as follows. Consider $dH \in T^*(M)$. For a symplectic manifold, there is a natural isomorphism between $T^*(M)$ and $T(M)$, namely, $\boldsymbol{\omega}^\sharp$. The unique vector field $\mathbf{X}_H$ associated with dH is the vector field we are after.

Hamiltonian vector field and Hamiltonian systems defined

Definition 28.7.8 Let $(M, \boldsymbol{\omega})$ be a symplectic manifold and $H : M \to \mathbb{R}$ a real-valued function. The vector field

$$\mathbf{X}_H \equiv \boldsymbol{\omega}^\sharp(dH) \equiv (dH)^\sharp$$

is called the **Hamiltonian vector field** with **energy function** H. The triplet $(M, \boldsymbol{\omega}, \mathbf{X}_H)$ is called a **Hamiltonian system**.

The significance of the Hamiltonian vector field lies in its integral curve which turns out to be the path of evolution of the system in the phase space. This is shown in the following proposition.

Proposition 28.7.9 *If $(q^1, \dots, q^n, p_1, \dots, p_n)$ are canonical coordinates for $\boldsymbol{\omega}$—so $\boldsymbol{\omega} = \sum dq^i \wedge dp_i$—then, in these coordinates*

$$\mathbf{X}_H = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial x^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} \equiv \left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i}\right). \tag{28.51}$$

[13] Here, we are assuming that the mechanical system in question is nonsingular, by which is meant that there are precisely n independent p_i's. There are systems of considerable importance that happen to be singular. Such systems, among which are included all gauge theories such as the general theory of relativity, are called **constrained systems** and are characterized by the fact that $\boldsymbol{\omega}$ is degenerate. Although of great interest and currently under intense study, we shall not discuss constrained systems in this book.

Therefore, $(q(t), p(t))$ *is an integral curve of* $\mathbf{X}_H$ *iff Hamilton's equations hold*:

$$\frac{\partial q^i}{\partial t} = \frac{\partial H}{\partial p_i}, \qquad \frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial q^i}, \quad i = 1, \dots, n. \tag{28.52}$$

Proof The first part of the proposition follows from

$$dH = \frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial p_i} dp_i,$$

from the definition of $\mathbf{X}_H$ in terms of dH, and from Eq. (28.49). The second part follows from the definition of integral curve and Eq. (28.19). □

We called H the energy function; this is for good reason:

conservation of energy in the language of symplectic geometry

Theorem 28.7.10 *Let* $(M, \boldsymbol{\omega}, \mathbf{X}_H)$ *be a Hamiltonian system and* $\gamma(t)$ *an integral curve of* $\mathbf{X}_H$. *Then* $H(\gamma(t))$ *is constant in* t.

Proof We show that the time-derivative of $H(\gamma(t))$ is zero:

$$\begin{aligned} \frac{d}{dt} H\big(\gamma(t)\big) &= \gamma_{*t}(H) && \text{by Proposition 28.2.4} \\ &= dH(\gamma_{*t}) && \text{by Eq. (28.14)} \\ &= dH\big(\mathbf{X}_H\big(\gamma(t)\big)\big) && \text{by definition of integral curve} \\ &= \big[\boldsymbol{\omega}^\flat\big(\mathbf{X}_H\big(\gamma(t)\big)\big)\big]\big(\mathbf{X}_H\big(\gamma(t)\big)\big) && \text{by definition of } \mathbf{X}_H\big(\gamma(t)\big) \\ &= \boldsymbol{\omega}\big(\mathbf{X}_H\big(\gamma(t)\big), \mathbf{X}_H\big(\gamma(t)\big)\big) && \text{by the definition of } \boldsymbol{\omega}^\flat \\ &= 0 && \text{because } \boldsymbol{\omega} \text{ is skew-symmetric} \end{aligned}$$

□

Theorem 28.7.10 is the statement of the conservation of energy.

Historical Notes

Sir William Rowan Hamilton 1805–1865

Sir William Rowan Hamilton (1805–1865), the fourth of nine children, was mostly raised by an uncle, who quickly realized the extraordinary nature of his young nephew. By the age of five, Hamilton spoke Latin, Greek, and Hebrew, and by the age of nine had added more than a half dozen languages to that list. He was also quite famous for his skill at rapid calculation. Hamilton's introduction to mathematics came at the age of 13, when he studied Clairaut's Algebra, a task made somewhat easier as Hamilton was fluent in French by this time. At age 15 he started studying Newton, whose *Principia* spawned an interest in astronomy that would provide a great influence in Hamilton's early career.

In 1822, at the age of 18, Hamilton entered Trinity College, Dublin, and in his first year he obtained the top mark in classics. He divided his studies equally between classics and mathematics and in his second year he received the top award in mathematical physics. Hamilton discovered an error in Laplace's *Méchanique céleste*, and as a result, he came to the attention of John Brinkley, the Astronomer Royal of Ireland, who said: "This young man, I do not say will be, but is, the first mathematician of his age." While in his final year as an undergraduate, he presented a memoir entitled *Theory of Systems of Rays* to the Royal Irish Academy in which he planted the seeds of **symplectic geometry**.

Hamilton's personal life was marked at first by despondency. Rejected by a college friend's sister, he became ill and nearly suicidal. He was rejected a few years later by another friend's sister and wound up marrying a very timid woman prone to ill health. Hamilton's own personality was much more energetic and humorous, and he easily acquired friends among the literati. His own attempts at poetry, which he himself fancied, were generally considered quite poor. No less an authority than Wordsworth attempted to convince him that his true calling was mathematics, not poetry. Nevertheless, Hamilton maintained close connection with the worlds of literature and philosophy, insisting that the ideas to be gleaned from them were integral parts of his life's work. While Hamilton is best known in physics for his work in dynamics, more of his time was spent on studies in optics and the theory of quaternions. In optics, he derived a function of the initial and final coordinates of a ray and termed it the "characteristic function," claiming that it contained "the whole of mathematical optics." Interestingly, his approach shed no new light on the wave/corpuscular debate (being independent of which view was taken), another appearance of Hamilton's quest for ultimate generality.

In 1833 Hamilton published a study of *vectors* as ordered pairs. He used algebra to study dynamics in *On a General Method in Dynamics* in 1834. The theory of quaternions, on which he spent most of his time, grew from his dissatisfaction with the current state of the theoretical foundation of algebra. He was aware of the description of complex numbers as points in a plane and wondered if any other geometrical representation was possible or if there existed some hypercomplex number that could be represented by three-dimensional points in space. If the latter supposition were true, it would entail a natural algebraic representation of ordinary space. To his surprise, Hamilton found that in order to create a hypercomplex number algebra for which the modulus of a product equaled the product of the two moduli, *four components* were required—hence, quaternions.

Hamilton felt that this discovery would revolutionize mathematical physics, and he spent the rest of his life working on quaternions, including publication of a book entitled *Elements of Quaternions*, which he estimated would be 400 pages long and take two years to write. The title suggests that Hamilton modeled his work on Euclid's *Elements* and indeed this was the case. The book ended up double its intended length and took seven years to write. In fact, the final chapter was incomplete when Hamilton died, and the book was finally published with a preface by his son, William Edwin Hamilton. While quaternions themselves turned out to be of no such monumental importance, their appearance as the first noncommutative algebra opened the door for much research in this field, including much of vector and matrix analysis. (As a side note, the "del" operator, named later by Gibbs, was introduced by Hamilton in his papers on quaternions.)

In dynamics, Hamilton extended his characteristic function from optics to the classical action for a system moving between two points in configuration space. A simple transformation of this function gives the quantity (the time integral of the Lagrangian) whose variation equals zero in what we now call Hamilton's principle. Jacobi later simplified the application of Hamilton's idea to mechanics, and it is the Hamilton–Jacobi equation that is most often used in such problems. Hamiltonian dynamics was rescued from what could have become historical obscurity with the advent of quantum mechanics, in which its close association with ideas in optics found fertile application in the wave mechanics of de Broglie and Schrödinger. Hamilton's later life was unhappy, and he became addicted to alcohol. He died from a severe attack of gout shortly after receiving the news that he had been elected the first foreign member of the National Academy of Sciences of the USA.

In the theoretical development of mechanics, canonical transformations play a central role. The following proposition shows that the flows of a Hamiltonian system are such transformations:

Flow of Hamiltonian vector field is canonical transformation of mechanics.

Proposition 28.7.11 *Let* $(M, \boldsymbol{\omega}, \mathbf{X}_H)$ *be a Hamiltonian system, and* F_t *the flow of* $\mathbf{X}_H$. *Then for each* t, $F_t^*\boldsymbol{\omega} = \boldsymbol{\omega}$, *i.e.*, F_t *is symplectic.*

Proof We have

$$
\begin{aligned}
\frac{d}{dt} F_t^* \boldsymbol{\omega} &= F_t^* L_{\mathbf{X}_H} \boldsymbol{\omega} && \text{by Eq. (28.39)} \\
&= F_t^* (i_{\mathbf{X}_H} d\boldsymbol{\omega} + d i_{\mathbf{X}_H} \boldsymbol{\omega}) && \text{by Theorem 28.5.10} \\
&= F_t^* (0 + ddH) && \text{because } d\boldsymbol{\omega} = 0 \text{ and } i_{\mathbf{X}} \boldsymbol{\omega} = \boldsymbol{\omega}^\flat(\mathbf{X}) \\
&= 0 && \text{because } d^2 = 0.
\end{aligned}
$$

Thus, $F_t^* \boldsymbol{\omega}$ is constant in t. But $F_0^* = \text{id}$. Therefore, $F_t^* \boldsymbol{\omega} = \boldsymbol{\omega}$. □

The celebrated Liouville's theorem of mechanics, concerning the preservation of volume of the phase space, is a consequence of the proposition above:

Liouville's theorem

Corollary 28.7.12 (Liouville's theorem) *F_t preserves the phase volume μ_ω.*

Poisson brackets in the language of symplectic geometry

Definition 28.7.13 Let $(M, \boldsymbol{\omega})$ be a symplectic manifold. Let $f, g : M \to \mathbb{R}$ with $\mathbf{X}_f = (df)^\sharp$ and $\mathbf{X}_g = (dg)^\sharp$ their corresponding Hamiltonian vector fields. The **Poisson bracket** of f and g is the function

$$\{f, g\} \equiv \boldsymbol{\omega}(\mathbf{X}_f, \mathbf{X}_g) = i_{\mathbf{X}_g} i_{\mathbf{X}_f} \boldsymbol{\omega} = -i_{\mathbf{X}_f} i_{\mathbf{X}_g} \boldsymbol{\omega}.$$

We can immediately obtain the familiar expression for the Poisson bracket of two functions.

Proposition 28.7.14 *In canonical coordinates $(q^i, \dots, q^n, p_1, \dots, p_n)$, we have*

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} \right).$$

In particular,

$$\{q^i, q^j\} = 0, \qquad \{p_i, p_j\} = 0, \qquad \{q^i, p_j\} = \delta^i_j.$$

Proof From Eq. (28.51), we have

$$
\begin{aligned}
\boldsymbol{\omega}(\mathbf{X}_f, \mathbf{X}_g) &= \boldsymbol{\omega}\left(\frac{\partial f}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial f}{\partial q^i} \frac{\partial}{\partial p_i}, \frac{\partial g}{\partial p_j} \frac{\partial}{\partial q^j} - \frac{\partial g}{\partial q^j} \frac{\partial}{\partial p_j} \right) \\
&= \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial p_j} \underbrace{\boldsymbol{\omega}\left(\frac{\partial}{\partial q^i}, \frac{\partial}{\partial q^j} \right)}_{=0} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^j} \underbrace{\boldsymbol{\omega}\left(\frac{\partial}{\partial q^i}, \frac{\partial}{\partial p_j} \right)}_{=\delta^j_i} \right. \\
&\quad \left. - \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_j} \underbrace{\boldsymbol{\omega}\left(\frac{\partial}{\partial p_i}, \frac{\partial}{\partial q^j} \right)}_{=-\delta^i_j} + \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial q^j} \underbrace{\boldsymbol{\omega}\left(\frac{\partial}{\partial p_i}, \frac{\partial}{\partial p_j} \right)}_{=0} \right]
\end{aligned}
$$

$$= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} \right),$$

where we have assumed that $\boldsymbol{\omega} = \sum_{k=1}^{n} dq^k \wedge dp_k$. The other formulas follow immediately once we substitute p_i or q^i for f or g. □

28.8 Problems

28.1 Provide the details of the fact that a finite-dimensional vector space $\mathcal{V}$ is a manifold of dimension $\dim \mathcal{V}$.

28.2 Choose a different curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ whose tangent at $u = 0$ is still (a_x, a_y) of Example 28.2.2. For instance, you may choose

$$\gamma(u) = \left(\frac{a_x}{2}(u+1)^2, \frac{a_y}{3}(u-1)^3 \right).$$

Show that this curve gives the same relation between partials and unit vectors as obtained in that example. Can you find another curve doing the same job?

28.3 For every $\mathbf{t} \in \mathcal{T}_P(M)$ and every constant function $c \in F^\infty(P)$, show that $\mathbf{t}(c) = 0$. Hint: Use both parts of Definition 28.2.3 on the two functions $f = c$ and $g = 1$.

28.4 Find the coordinate vector field ∂_1 of Example 28.2.10.

28.5 Use the procedure of Example 28.2.10 to find a coordinate frame for S^2 corresponding to the stereographic projection charts (see Example 28.1.12).

28.6 Let (x^i) and (y^j) be coordinate systems on a subset U of a manifold M. Let X^i and Y^i be the components of a vector field with respect to the two coordinate systems. Show that $Y^i = X^j \partial y^i / \partial x^j$.

28.7 Show that if $\psi : M \to N$ is a local diffeomorphism at $P \in M$, then $\psi_{*P} : \mathcal{T}_P(M) \to \mathcal{T}_{\psi(P)}(N)$ is a vector space isomorphism.

28.8 Let $\mathbf{X}$ be a vector field on M and $\psi : M \to N$ a differentiable map. Then for any function f on N, $[\psi_* \mathbf{X}](f)$ is a function on N. Show that

$$\mathbf{X}(f \circ \psi) = \left\{ [\psi_* \mathbf{X}](f) \right\} \circ \psi.$$

28.9 Verify that the vector field $\mathbf{X} = -y\partial_x + x\partial_y$ has an integral curve through (x_0, y_0) given by

$$x = x_0 \cos t - y_0 \sin t,$$

$$y = x_0 \sin t + y_0 \cos t.$$

28.10 Show that the vector field $\mathbf{X} = x^2\partial_x + xy\partial_y$ has an integral curve through (x_0, y_0) given by

$$x(t) = \frac{x_0}{1 - x_0 t}, \qquad y(t) = \frac{y_0}{1 - x_0 t}.$$

28.11 Let $\mathbf{X}$ and $\mathbf{Y}$ be vector fields. Show that $\mathbf{X} \circ \mathbf{Y} - \mathbf{X} \circ \mathbf{Y}$ is also a vector field, i.e., it satisfies the derivation property.

28.12 Prove the remaining parts of Proposition 28.4.13.

28.13 Suppose that x^i are coordinate functions on a subset of M and $\boldsymbol{\omega}$ and $\mathbf{X}$ are a 1-form and a vector field there. Express $\boldsymbol{\omega}(\mathbf{X})$ in terms of component functions of $\boldsymbol{\omega}$ and $\mathbf{X}$.

28.14 Show that $d \circ L_{\mathbf{X}} = L_{\mathbf{X}} \circ d$. Hint: Use the definition of the Lie derivative for p-forms and the fact that d commutes with the pullback.

28.15 Let $M = \mathbb{R}^3$ and let f be a real-valued function. Let $\boldsymbol{\omega} = a_i dx^i$ be a one-form and $\boldsymbol{\eta} = b_1 dx^2 \wedge dx^3 + b_2 dx^3 \wedge dx^1 + b_3 dx^1 \wedge dx^2$ be a two-form on $\mathbb{R}^3$. Show that

(a) df gives the gradient of f,
(b) $d\boldsymbol{\eta}$ gives the divergence of the vector $\mathbf{B} = (b_1, b_2, b_3)$, and that
(c) $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) = 0$ and $\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{A}) = 0$ are consequences of $d^2 = 0$.

28.16 Show that $i_{\mathbf{X}}$ is an antiderivation with respect to the wedge product.

28.17 Given that $\mathsf{F} = \frac{1}{2} F_{\alpha\beta} dx^\alpha \wedge dx^\beta$, show that $\mathsf{F} \wedge (*\mathsf{F}) = |\mathbf{B}|^2 - |\mathbf{E}|^2$.

28.18 Use Eq. (28.41) to show that the zeroth component of the relativistic Lorentz force law gives the rate of change of energy due to the electric field, and that the magnetic field does not change the energy.

28.19 Derive Eq. (28.44).

28.20 Write the equation

$$F_{\alpha\beta} = A_{\beta,\alpha} - A_{\alpha,\beta} = \frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta}$$

in terms of $\mathbf{E}$, $\mathbf{B}$, and vector and scalar potentials.

28.21 With $\mathsf{F} = \frac{1}{2} F_{\alpha\beta} dx^\alpha \wedge dx^\beta$ and $\mathbf{J} = J_\gamma dx^\gamma$, show that $d * \mathsf{F} = 4\pi(*\mathbf{J})$ takes the following form in components:

$$\frac{\partial F^{\alpha\beta}}{\partial x^\beta} = 4\pi J^\alpha,$$

where indices are raised and lowered by $\mathrm{diag}(-1, -1, -1, 1)$.

28.22 Interpret Theorem 28.5.15 for $p = 1$ and $p = 2$ on $\mathbb{R}^3$.

28.23 Let f be a function on $\mathbb{R}^3$. Calculate $d * df$.

28.24 Show that current conservation is an automatic consequence of Maxwell's inhomogeneous equation $d * \mathbf{F} = 4\pi(*\mathbf{J})$.