Chapter 1

Manifolds and Dynamical Systems

In this chapter we shall introduce some basic topological and geometric concepts used in the studies of differentiable manifolds. In particular we shall present an intrinsic definition of vectors and tensors. These are the basic quantities for the formulation of classical mechanics and classical field theory. They also play an essential role in a geometric approach to quantization, providing a clear link between classical quantities and their quantized counterparts. The notations and materials, including many examples, are selected for their relevance to later physical applications. Mathematical technicality is kept to a minimum. More details on manifold theory are available in the references provided at the end of this chapter.

We shall denote the set of all real numbers by $I\!R$, i.e., $I\!R = (-\infty, \infty)$. The set of all ordered n-tuples of real numbers will be denoted by $I\!R^n$, i.e.,

$$I\!R^n = I\!R \times I\!R \times \cdots \times I\!R \tag{1.1}$$

with elements

$$\alpha = (\alpha^1, \alpha^2, \ldots, \alpha^n), \tag{1.2}$$

where α^j, $j = 1, 2, \ldots, n$, are real numbers. For $I\!R$ which corresponds to $n = 1$ we shall simply write $\alpha = \alpha^1$. Generally we call a set endowed with some kind of geometric or algebraic structure a *space*. There are two well-known structures which can be built on $I\!R^n$ making it respectively into a topological space and a Euclidean space.

1.1 Topological Spaces and Topological Equivalence

1.1.1 Basic concepts and definitions

We shall start by looking into structures arising from within $I\!R^n$. A simple structure is the family of subsets of $I\!R^n$.[1] For simplicity let us take the set $I\!R$ to begin with. Not all subsets of $I\!R$ are the same in nature. First we have the open intervals (a, b), $a, b \in I\!R$ defined by[2]

$$(a, b) = \{\alpha \in I\!R : a < \alpha < b\}. \tag{1.3}$$

Every element in (a, b) is contained in an open interval which is itself contained in (a, b). This enables us to generalize the notion of open intervals to that of open sets in $I\!R$.

Definition 1.1.1(1) Open sets in $I\!R$

- A subset Λ of $I\!R$ is an *open set* in $I\!R$ if every element α in Λ is contained in an open interval Λ_α inside Λ, i.e., $\alpha \in \Lambda \Rightarrow \alpha \in \Lambda_\alpha \subset \Lambda$.
- The complement of a subset Λ of $I\!R$ is the subset which contains all the elements of $I\!R$ that are not in Λ.
- A subset of $I\!R$ is a *closed set* in $I\!R$ if its complement is open.

Comments 1.1.1(1) Properties of open and closed sets

C1 The notation $\Lambda_1 \subset \Lambda_2$ means that $\alpha \in \Lambda_1 \Rightarrow \alpha \in \Lambda_2$, e.g., Λ_1 could be equal to Λ_2.

C2 Open intervals are open sets. Closed intervals are closed sets. A subset containing a single number $\{a\}$ is also closed. Open sets are generalizations of open intervals, e.g., $(1, 2) \cup (3, 4)$ is an open set but it is not an interval. Similarly closed sets are a generalization of closed intervals.

C3 There are sets which are neither open nor closed, e.g., a semi-open interval $(a, b]$.

---

[1] $I\!R^n$ itself is formally included as a subset, and so is the empty set $\emptyset$, which contains no elements.

[2] We adopt the standard notation that $[a, b]$ represents a closed interval and $(a, b]$ a semi-open interval, i.e.,

$$[a, b] = \{\alpha \in I\!R : a \leq \alpha \leq b\}, \quad (a, b] = \{\alpha \in I\!R : a < \alpha \leq b\}.$$

C4 Regarded as a subset $\mathbb{R}$ itself is clearly an open set. The empty set $\emptyset$ formally satisfies the requirement of an open set, since there is no element in $\emptyset$ which is not contained in an open interval in $\emptyset$. So, the empty set shall be regarded as an open set in $\mathbb{R}$. It then follows that $\mathbb{R}$ and $\emptyset$ also formally satisfy the requirement of a closed set, since their respective complements are open. Therefore, $\mathbb{R}$ and $\emptyset$ shall also be regarded as closed sets. These two are the only subsets of $\mathbb{R}$ which are both open and closed.

C5 The set of positive real numbers $(0, \infty)$, denoted by $\mathbb{R}^+$ hereafter, is an open set, and so is the set of negative real numbers $(-\infty, 0)$, denoted hereafter by $\mathbb{R}^-$.

C6 A fundamental property of closed sets is that its points cannot get arbitrarily close to any point outside it, e.g., points in $[a, b]$ cannot get arbitrarily close any point outside $[a, b]$. This is in sharp contrast to the situation for open sets. For example points in (a, b) can get arbitrarily close to a and b which lie outside (a, b).

C7 The fundamental structural differences between open sets and closed sets in $\mathbb{R}$ are:

1. The union of any, possibly *infinite*, number of open sets is an open set, and the intersection of any *finite* number of open sets is also open.

2. The union of any *finite* number of closed sets is closed, and the intersection of any, possibly *infinite*, number of closed sets is also closed.

The fact that the intersection of an infinite number of open sets is not necessarily open is exemplified as follows:

$$(-1, 1) \cap \left(-\frac{1}{2}, \frac{1}{2}\right) \cap \left(-\frac{1}{3}, \frac{1}{3}\right) \cap \left(-\frac{1}{4}, \frac{1}{4}\right) \cap \cdots = \{0\} \quad \text{which is closed.}$$

The fact that the union of an infinite number of closed sets is not always closed is seen in the following example:

$$\{1\} \cup \left\{\frac{1}{2}\right\} \cup \left\{\frac{1}{3}\right\} \cup \left\{\frac{1}{4}\right\} \cup \cdots .$$

This union is not closed because points in the union can get arbitrarily close to 0, which lies outside the union.

C8 Every open set in $\mathbb{R}$ can be shown to be a union of open intervals.

Definition 1.1.1(2) Topological structure and topological spaces[3]

- Let $\mathcal{T}$ be a non-empty set. A collection $\mathbf{C}$ of subsets of $\mathcal{T}$ is called a *topological structure* or a *topology* on $\mathcal{T}$ if:
    1. Both $\mathcal{T}$ and the empty set $\emptyset$ belong to $\mathbf{C}$.
    2. The union of any, possibly infinite, number of sets in $\mathbf{C}$ again belongs to $\mathbf{C}$.
    3. The intersection of any finite number of sets in $\mathbf{C}$ again belongs to $\mathbf{C}$.
- Members of $\mathbf{C}$ are called *open sets* and the set $\mathcal{T}$ together with a topological structure is called a *topological space*.
- A subset of $\mathcal{T}$ is said to be *closed* if its complement is open.

Comments 1.1.1(2) Neighbourhoods, closure and dense sets

C1 The open sets in $I\!R$ introduced earlier in terms of open intervals constitute a topological structure, known as the *standard topology* on $I\!R$. With this topology $I\!R$ becomes a topological space. The structure of open sets in a topological space defined above is a generalization of that of open sets in $I\!R$. From now on we shall always adopt the standard topology on $I\!R$.

C2 Closed sets in a topological space possess the properties of closed sets in $I\!R$ stated in §1.1.1C(1) C7. Also $\mathcal{T}$ and the empty set $\emptyset$ are both closed.

C3 An arbitrary open interval $(a, b) \subset I\!R$ is a topological space in its own right with the standard topology formed by the open sets of $I\!R$ that are subsets of (a, b).

C4 Generally it is possible to single out a different collection of subsets to form a different topological structure; this will result in a different topological space. A trivial example is to take the class $\mathbf{C}$ to consist of just the empty set $\emptyset$ and the entire set $\mathcal{T}$.

C5 An element of a topological space is often referred to as a *point* in the space. A useful concept is that of a *neighbourhood* of a point which is defined to be an open set containing the point.

C6 The terms *closure* and *denseness* are topological concepts useful also in the context of Hilbert spaces:

[3] See Lipschutz (1965), Simmons (1963) and Sutherland (1995) for an introduction to topology as a subject.

- The *closure* of a subset Λ of a toplogical space $\mathcal{T}$, denoted by $\bar{\Lambda}$, is the smallest closed set in $\mathcal{T}$ containing Λ. In other words $\bar{\Lambda}$ is the intersection of all closed sets in $\mathcal{T}$ containing Λ.

- A set Λ is *dense* in $\mathcal{T}$ if its closure $\bar{\Lambda}$ coincides with $\mathcal{T}$.

As an example we can see that the closure of an open interval (a, b) in $I\!R$ is the closed interval $[a, b]$. The set of all rational numbers is a dense set in $I\!R$. We know that rational numbers permeate every part of $I\!R$. Intuitively this is precisely the property of a dense set, i.e., a dense subset Λ of $\mathcal{T}$ is a subset which permeates every part of $\mathcal{T}$ so that the only closed set containing Λ is the space $\mathcal{T}$ itself.

C7 If a and b are two distinct points in $I\!R$ then there exists a neighbourhood of a and a neighbourhood of b such that these neighbourhoods are disjoint.[4] This property is not shared by all topological spaces. We can formalize this property as follows:

- A topological space is called a *Hausdorff space* if any two distinct points have disjoint neighbourhoods.

All the topological spaces we shall encounter in this book are Hausdorff.

Definition 1.1.1(3) Open sets in $I\!R^n$

- Let $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ be n open intervals in $I\!R$. Then the set $\Lambda_{rec} = \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_n$ is called an *open rectangle* in $I\!R^n$.

- A subset Λ of $I\!R^n$ is called an open set in $I\!R^n$ if each point $\alpha \in \Lambda$ is contained in an open rectangle Λ_{rec} inside Λ, i.e., $\alpha \in \Lambda \;\Rightarrow\; \alpha \in \Lambda_{rec} \subset \Lambda$.

Comments 1.1.1(3) Standard topology on $I\!R^n$

C1 For $I\!R^3$, the simplest example of an open rectangle is an *open cube* of width $w \in I\!R^+$, e.g.,

$$\Lambda_{cu}(w) = \Big\{\alpha \in I\!R^3 : -\frac{w}{2} < \alpha^j < \frac{w}{2},\, j = 1, 2, 3\Big\}. \tag{1.4}$$

Its closure is simply the closed cube

$$\bar{\Lambda}_{cu}(w) = \Big\{\alpha \in I\!R^3 : -\frac{w}{2} \leq \alpha^j \leq \frac{w}{2},\, j = 1, 2, 3\Big\}. \tag{1.5}$$

[4] In this chapter the term *disjointness* means having no common elements.

C2 The family of open sets defined above constitutes a topological structure, known as the *standard topology* on $I\!R^3$. A similar topology can be constructed in $I\!R^n$ which renders $I\!R^n$ a topological space.

C3 An open cube is a topological space in its own right, with the topology formed by the open sets of $I\!R^n$ that are subsets of the cube.

C4 We can also define closed rectangles and closed sets in $I\!R^n$ in terms of the openness of their complements as we did in $I\!R$.

Definition 1.1.1(4) Functions on $I\!R^n$

- A mapping $f : I\!R^n \mapsto I\!R$, assigning a real number $f(\alpha)$ to each point $\alpha \in I\!R^n$, is called a *function* on $I\!R^n$. The set of values $\mathcal{R} = \{f(\alpha) : \alpha \in I\!R^n\}$ is called the *range* of the function.

Comments 1.1.1(4) Domain, support and smoothness of functions

C1 We can also define *local functions*. Let $\mathcal{D}$ be an open subset of $I\!R^n$. A map $f : \mathcal{D} \mapsto I\!R$ is called a *local function* and $\mathcal{D}$ is called the *domain* of the function. By allowing $\mathcal{D} = I\!R^n$ all functions defined so far may be called local. Conversely a local function f can be extended to the entire space $I\!R^n$ by setting, for example $f(\alpha) = 0$ for every $\alpha \notin \mathcal{D}$. But from now on, a function means a function defined on the domain $\mathcal{D} = I\!R^n$, unless stated otherwise.

C2 The closure of the set of all points at which a function f is not zero is called the *support* of the function, to be denoted by $supp(f)$, i.e.,

$$supp(f) \text{ is the closure of the set } \{\alpha \in I\!R^n : f(\alpha) \neq 0\}.$$

C3 A function f on $I\!R^n$ is a function of n real variables α^j. We often write

$$f(\alpha) = f(\alpha^1, \alpha^2, \ldots, \alpha^n). \tag{1.6}$$

A function f is said to be of *class* C^ℓ if for all integers

$$\ell_k \geq 0,\ k = 1, 2, \ldots, n \tag{1.7}$$

such that

$$\ell_1 + \ell_2 + \cdots + \ell_n = \ell \tag{1.8}$$

the partial derivatives

$$\frac{\partial^\ell f}{(\partial\alpha^1)^{\ell_1}(\partial\alpha^2)^{\ell_2}\cdots(\partial\alpha^n)^{\ell_n}} \tag{1.9}$$

exist and are continuous. If f is of class C^ℓ for all positive integers ℓ we call f a C^∞ function. Such a function is also called *infinitely differentiable* or *smooth* for short.

C4 The following four groups of smooth functions prove to be particularly useful later on:

- $C^\infty(\mathbb{R}^n)$: The collection of all smooth functions on $\mathbb{R}^n$.
- $C_0^\infty(\mathbb{R}^n)$: The set of all smooth functions defined on $\mathbb{R}^n$ with bounded support, i.e., each of these functions vanishes outside some bounded rectangle in $\mathbb{R}^n$. These functions are generally known as smooth *functions of compact support*.[5]
- $C^\infty(\mathcal{D})$: Smooth local functions with domain $\mathcal{D}$.
- $C^\infty(\alpha)$: The set of all smooth local functions on $\mathbb{R}^n$ whose domains contain the point α in $\mathbb{R}^n$.

1.1.2 Topological equivalence

It is important to be able to compare and relate different topological spaces. Let $\mathcal{T}$ and $\mathcal{T}'$ be two topological spaces. To relate them we would need to map $\mathcal{T}$ to $\mathcal{T}'$, and it is the nature of this mapping which enables us to compare the two spaces. We shall confine ourselves to one-to-one mappings F of $\mathcal{T}$ onto $\mathcal{T}'$ unless otherwise is stated.[6] Clearly for such a mapping the inverse F^{-1}, which maps $\mathcal{T}'$ one-to-one onto $\mathcal{T}$, exists. When the mapping F is not one-to-one, the inverse mapping does not exist.[7]

Given a function F the *image* $F(\Lambda)$ of any subset Λ of $\mathcal{T}$ consists of all $\tau' \in \mathcal{T}'$ such that $\tau' = F(\tau)$ for some $\tau \in \Lambda$. We can also define the *inverse image* $F^{-1}(\Lambda')$ of any subset Λ' of $\mathcal{T}'$ to be the subset $\Lambda \in \mathcal{T}$ consisting of all $\tau \in \mathcal{T}$ such that $F(\tau) \in \Lambda'$, i.e.,

$$F(\Lambda) = \{\tau' = F(\tau) : \tau \in \Lambda\}, \tag{1.10}$$

$$F^{-1}(\Lambda') = \{\tau : \tau \in \mathcal{T}, F(\tau) \in \Lambda'\}. \tag{1.11}$$

---

[5] A rectangle $\Lambda_{rec} = \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_n$ is bounded if all the intervals $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ are bounded. Generally a subset in $\mathbb{R}^n$ is bounded if it is contained in a bounded rectangle Λ_{rec}. We have avoided introducing the notion of *compact sets* in the context of a general topological space. In $\mathbb{R}^n$ with the standard topology a compact set is just a closed and bounded set.

[6] Simmons (1963).

[7] A mapping $F : A \mapsto B$ is one-to-one if distinct elements a_1 and a_2 in A are mapped to distinct images in B, i.e., $F(a_1,) \neq F(a_2)$ if $a_1 \neq a_2$. It is an *onto mapping* if every element in B is the image of an element in A, i.e., given any $b \in B$ there is $a \in A$ such that $b = F(a)$, otherwise it is called an *into* mapping.

Definition 1.1.2(1) Open functions and continuous functions

- F is said to be an *open function* if the image of every open set Λ of $\mathcal{T}$ is an open set Λ' in $\mathcal{T}'$.
- F is said to be a *continuous function* if the inverse image of every open set Λ' of $\mathcal{T}'$ is an open set Λ in $\mathcal{T}$.

Comments 1.1.2(1) Agreement with usual notion of continuity

C1 The concepts of images and inverse images are defined whether the function is one-to-one or not. So, the above definitions of open and continuous functions apply to functions which are not necessarily one-to-one.

C2 Let $\mathcal{T} = \mathcal{T}' = I\!R$. Then, a mapping F of $\mathcal{T}$ onto $\mathcal{T}'$ is identifiable with a real-valued function of a real variable. The above definition of continuity can be shown to agree with the usual concept of continuity of real-valued functions used in elementary calculus. At first sight it may appear natural to define continuity in terms of open functions. However, this turns out to be wrong. Consider a constant function on $\mathcal{T} = I\!R$ which maps every point $\alpha \in \mathcal{T}$ to the value 1. This is not an open function since it maps every open set of $\mathcal{T}$ to the closed set $\{1\}$, but it is continuous in the usual sense. Open functions thus cannot serve to define the usual notion of continuity.

Definition 1.1.2(2) Homeomorphism and topological equivalence

- Two topological spaces $\mathcal{T}$ and $\mathcal{T}'$ are said to be *topologically equivalent* or *homeomorphic* if there is a one-to-one mapping F of $\mathcal{T}$ onto $\mathcal{T}'$ such that both F and F^{-1} are continuous. Such a mapping F is said to be a *homeomorphism* or to be *bicontinuous*.

Comments 1.1.2(2) Topological equivalence in $I\!R$

C1 Let $\mathcal{T} = (-\frac{1}{2}\pi, \frac{1}{2}\pi) \subset I\!R$ and $\mathcal{T}'$ be the real line $I\!R$. Then

$$F(x) = \tan x$$

regarded as a function from $\mathcal{T}$ to $\mathcal{T}'$ is one-to-one, onto and continuous. Its inverse is also continuous, making F a homeomorphism. Hence $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ is topologically equivalent to $I\!R$. Conversely $I\!R$ is also topologically equivalent to the open interval $(-\frac{1}{2}\pi, -\frac{1}{2}\pi)$. In fact the real line is topologically equivalent to any finite open interval.

C2 Topologically equivalent spaces have common topological properties.

Definition 1.1.2(3) Connectivity

- A topological space $\mathcal{T}$ is said to be *disconnected* if it is the union of two open, non-empty and disjoint subsets, i.e., there are non-empty open subsets $\mathcal{M}_1$ and $\mathcal{M}_2$ of $\mathcal{T}$ such that

$$\mathcal{M}_1 \cup \mathcal{M}_2 = \mathcal{T} \quad \text{and} \quad \mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset. \tag{1.12}$$

- A topological space $\mathcal{T}$ is said to be *connected* if it is not disconnected.

Comments 1.1.2(3) Connectivity

C1 The symbol $\emptyset$ shall always denote the *empty set*.

C2 Consider the spaces $\mathcal{T}_1 = I\!R$ and $\mathcal{T}_2 = I\!R^- \cup I\!R^+$. Then $\mathcal{T}_1$ is connected and $\mathcal{T}_2$ is disconnected as we would intuitively expect. We shall have occasions to employ disconnected spaces in later chapters. However, unless it is stated otherwise all the topological spaces we shall consider are assumed to be connected.

1.2 Euclidean Spaces

1.2.1 Basic concepts and definitions

We can set up a real vector space structure in $I\!R^n$ by introducing scalar multiplication and addition.[8] Let a be a real number then $a\alpha$ is an element in $I\!R^n$ defined by

$$a\alpha = (a\alpha^1, a\alpha^2, \ldots, a\alpha^n), \tag{1.13}$$

and if $\beta = (\beta^1, \beta^2, \ldots, \beta^n)$ is another element in $I\!R^n$, then the sum $\alpha + \beta$ is defined to be the element in $I\!R^n$ given by

$$\alpha + \beta = (\alpha^1 + \beta^1, \alpha^2 + \beta^2, \ldots, \alpha^n + \beta^n). \tag{1.14}$$

The resulting vector space is clearly of dimension n with the zero element $(0, 0, \ldots, 0)$ which will simply be denoted by 0. We can go further to introduce a *norm* $||\alpha||$ to each element α by

$$||\alpha|| = \left(\sum_{j=1}^{n} \alpha^j \alpha^j \right)^{1/2}. \tag{1.15}$$

This norm possesses the following characteristic properties:

[8] Simmons (1963) §14 pp. 80-81.

1. $||a\alpha|| = |a|\,||\alpha||$.
2. $||\alpha|| \geq 0$, and $||\alpha|| = 0$ if and only if $\alpha = 0$.
3. $||\alpha + \beta|| \leq ||\alpha|| + ||\beta||$.

Another useful property is

$$\Big|\, ||\alpha|| - ||\beta|| \,\Big| \leq ||\alpha - \beta||.$$

Definition 1.2.1(1) Euclidean spaces

- The set $I\!R^n$ endowed with a real vector space structure and a norm, defined by Eqs. (1.13), (1.14) and (1.15), is called an n-dimensional *Euclidean space*.

Comments 1.2.1(1) Distance, metric and scalar product

C1 To emphasize the vector space structure we shall adopt the notation $I\!E^n$ to denote the above Euclidean space. The zero element is also referred to as the *origin* of the space $I\!E^n$. It is important to distinguish the space $I\!E^n$ from $I\!R^n$ which has no vector space structure. The one-dimensional Euclidean space is denoted simply by $I\!E$.

C2 The norm in $I\!E^n$ induces a *distance function* or a *metric* between any two elements α and β given by

$$||\alpha - \beta|| = \left(\sum_{j=1}^{n} (\alpha^j - \beta^j)^2 \right)^{1/2}. \tag{1.16}$$

C3 The Euclidean space $I\!E^n$ endowed with this metric becomes a *metric space*. With a concept of distance we can introduce the concept of *boundedness* of a set in $I\!E^n$. A subset Λ of $I\!E^n$ is *bounded* if there is a point $\alpha_0 \in \Lambda$ and a number $a \in I\!R$ such that $||\alpha_0 - \alpha|| < a$ for all $\alpha \in \Lambda$. This agrees with the boundedness definition in footnote 5 to §1.1.1C(4) C4.

C4 Another useful structure in $I\!E^n$ is the *scalar product* which assigns a real number $\langle \alpha \mid \beta \rangle$ to any two elements α and β in $I\!E^n$ by

$$\langle \alpha \mid \beta \rangle = \sum_{j=1}^{n} \alpha^j \beta^j. \tag{1.17}$$

This scalar product satisfies the following Schwarz's inequality:

$$|\langle \alpha \mid \beta \rangle| \leq \|\alpha\| \, \|\beta\|. \tag{1.18}$$

C5 By inheriting the standard topology of $\mathbb{R}^n$ the Euclidean space $\mathbb{E}^n$ is also a topological space. An alternative way to introduce a topological structure is to make use of the metric to define open sets. First, we define an *open sphere* $\mathcal{S}_r(\alpha_0)$ with centre α_0 and radius r in $\mathbb{E}^n$ to be the subset

$$\mathcal{S}_r(\alpha_0) = \Big\{\alpha \in \mathbb{E}^n : \ ||\alpha - \alpha_0|| < r\Big\}. \tag{1.19}$$

One can then define an open set Λ to be a subset such that every element α in Λ is contained in an open sphere $\mathcal{S}_r(\alpha_0)$ inside Λ, i.e.,

$$\alpha \in \Lambda \quad \Rightarrow \quad \alpha \in \mathcal{S}_r(\alpha_0) \subset \Lambda. \tag{1.20}$$

This results in the same standard topology inherited from $\mathbb{R}^n$. We shall adopt this topology for $\mathbb{E}^n$ from now on.

C6 Subsets of a topological space can be given a topology, known as a *relative topology*, in a natural manner. An example is that of a circle $\mathcal{C}$ in $\mathbb{E}^2$. Such a circle is a topological space with a topology consisting of open sets defined by the intersections of $\mathcal{C}$ with all the open sets in $\mathbb{E}^2$. Similarly a sphere in $\mathbb{E}^3$ is also a topological space.

1.2.2 Coordinate systems and coordinate transformations

A *coordinate system* on the Euclidean space $\mathbb{E}^n$ is an assignment of n real numbers, referred to as *coordinates*, to specify points in the space. The original n-tuples of real numbers $(\alpha^1, \alpha^2, \ldots, \alpha^n)$ can serve as a coordinate system with $\alpha^1, \alpha^2, \ldots, \alpha^n$ as coordinates.

Let $f^1, f^2, \ldots, f^n$ be n smooth functions on $\mathbb{R}^n$ which induce a one-to-one map of $\mathbb{R}^n$ onto itself by[9]

$$(\alpha^1, \alpha^2, \ldots, \alpha^n) \mapsto \Big(f^1(\alpha), f^2(\alpha), \ldots, f^n(\alpha)\Big). \tag{1.21}$$

Let

$$x^1 = f^1(\alpha),\ x^2 = f^2(\alpha), \ldots,\ x^n = f^n(\alpha). \tag{1.22}$$

Then every point $\alpha \in \mathbb{E}^n$ has two n-tuples of numbers associated with it, $(\alpha^1, \alpha^2, \ldots, \alpha^n)$ and $(x^1, x^2, \ldots, x^n)$. To specify points in $\mathbb{E}^n$ we can use the new n-tuple $(x^1, x^2, \ldots, x^n)$, leading to a new coordinatization with coordinates $x^1, x^2, \ldots, x^n$. This shows that $\mathbb{E}^n$ admits many different coordinate systems. We shall denote both the coordinate system and the coordinates by x^j.

[9]Smooth functions are members of $C^\infty(\mathbb{R}^n)$ introduced in §1.1.1C(4) C4.

To avoid confusion we emphasize here that the mapping defined by Eq. (1.21) is not a mapping of $\mathbb{E}^n$ onto itself; it is a mapping of the n-tuples of real numbers $\mathbb{R}^n$ onto itself. For each point $\alpha \in \mathbb{E}^n$ specified by the original coordinates α^j this mapping leads to a new set of numbers x^j associated with the same point α in $\mathbb{E}^n$. We call such a mapping a *coordinate transformation.* We shall confine ourselves to coordinate systems which relate to the original system $(\alpha^1, \alpha^2, \ldots, \alpha^n)$ in a smooth manner as described by Eq. (1.21).

A simple example is a *linear coordinate transformation* of the form

$$x^j = f^j(\alpha) = \sum_{k=1}^{n} M^j_{\ k}\alpha^k, \quad M^j_{\ k} \in \mathbb{R} \tag{1.23}$$

where the constants $M^j_{\ k}$ satisfy

$$\sum_{j=1}^{n} M^j_{\ i} M^j_{\ k} = \delta_{ik}, \tag{1.24}$$

where δ_{ik} is the *Kronecker delta* which takes the value 1 if $i = k$ and vanishes otherwise. The Kronecker delta is often denoted also by the symbol δ^i_k. This transformation, known as a *homogeneous orthogonal coordinate transformation*, is characterized by the preservation of the expression for the norm, i.e., we have

$$||\alpha||^2 = \sum_{j=1}^{n} \alpha^j\alpha^j = \sum_{j=1}^{n} x^j x^j. \tag{1.25}$$

The original coordinate system α^j and all other coordinate systems related to it in this way are called *rectangular Cartesian* coordinate systems.

We can define real-valued functions on $\mathbb{E}^n$ as mappings of $\mathbb{E}^n$ to $\mathbb{R}$, i.e.,

$$f : \mathbb{E}^n \mapsto \mathbb{R} \quad \text{by} \quad \alpha \mapsto f(\alpha) \in \mathbb{R}. \tag{1.26}$$

Such mappings are independent of any coordinate system. In a given coordinate system x^j a function f on $\mathbb{E}^n$ can be described as a function of the coordinates. The expression for the function will be different in different coordinate systems, but the function f itself has not changed in that the same point α in $\mathbb{E}^n$ is mapped to the same value $f(\alpha)$. A function on $\mathbb{E}^n$ is said to be *smooth* or *infinitely differentiable* if expressed as a function of the original coordinates α^j it is a smooth function on $\mathbb{R}^n$ in the sense described in §1.1.1C(4) C3. Since all other coordinates are smooth functions of α^j we can define smooth functions on $\mathbb{E}^n$ in terms of any coordinate system. We can also define local functions in $\mathbb{E}^n$ in the same way as we do in $\mathbb{R}^n$. As in §1.1.1C(4) C4 we can define four sets of functions:

- $C^\infty(\mathbb{E}^n)$, the set of all smooth functions on $\mathbb{E}^n$.
- $C_0^\infty(\mathbb{E}^n)$, the set of all smooth functions on $\mathbb{E}^n$ of compact support.
- $C_0^\infty(\mathcal{D})$, the set of all smooth local functions with domain $\mathcal{D}$.
- $C^\infty(\alpha)$, the set of all smooth local functions whose domains contain the point $\alpha \in \mathbb{E}^n$.

It is often useful to introduce coordinate systems which are not rectangular Cartesian or even global. A well-known example is the spherical coordinate system in the 3-dimensional Euclidean space $\mathbb{E}^3$. At the origin spherical coordinates are not defined. Generally let $\mathcal{D}$ be an arbitrary connected open subset of $\mathbb{E}^n$. Then the original coordinates α^j confined to $\mathcal{D}$ form a coordinate system on $\mathcal{D}$. We can introduce a new coordinate system on $\mathcal{D}$ in terms of n smooth functions f^j defined on $\mathcal{D}$ which maps $\mathcal{D}$ one-to-one onto a connected open set of $\mathbb{R}^n$. These functions then define new coordinates x^j on $\mathcal{D}$, forming what is known as a *local coordinate chart* on $\mathcal{D}$. Let $\mathcal{D}'$ be another connected open set of $\mathbb{E}^n$, and let x'^j be a local coordinate chart on $\mathcal{D}'$. On the intersection $\mathcal{D} \cap \mathcal{D}'$ the coordinates x^j and x'^j are smooth functions of each other due to their smooth relation with α^j. So, it does not matter which coordinate chart we use.[10]

Local coordinate charts are of fundamental importance in topological spaces which cannot be covered by a single global coordinate system. They have to be covered by overlapping local coordinate charts. We shall return to this problem of local coordinate charts later.

1.2.3 Contravariant and covariant vectors in $\mathbb{E}^n$

Coordinate systems and the way they transform are instrumental in establishing the notion of vectors in physics. The term *vector* has been used to refer to a wide variety of mathematical objects, most commonly to elements of vector spaces. What we are interested in here is the notion of vectors as geometric objects appearing in differential geometry, mechanics, Special and General Relativity. We come across the term *vector* in elementary mechanics describing quantities which possess a direction as well as a magnitude. Intuitively we can see that a direction has to be relative to something, e.g., relative to some coordinate axes. So, in more advanced texts, the rather qualitative notion of direction is sharpened in terms of a multi-component object in a given coordinate system, namely a set of n numbers X^j, $j = 1, 2, \ldots, n$, representing the components along some chosen coordinate axes. However, simply listing a set of n numbers in a given coordinate system is not enough. One has to know

[10] See Benn and Tucker (1987) p. 131 for a diagramatic illustration.

what would happen to these components when one goes to a new coordinate system. There are two kinds of n-component objects traditionally referred to as vectors, depending on how the components change on a transformation of coordinates.[11]

Definition 1.2.3(1) Contravariant and covariant vectors

- An n-tuple of numbers X^j, $j = 1, 2, \ldots, n$, associated with a point α in $\mathbb{E}^n$ are said to be the components of a *contravariant vector* at the point α in a given coordinate system x^j if they transform, on a change of coordinate system to x'^j, according to

$$X'^k = \sum_{j=1}^{n} \left(\frac{\partial x'^k}{\partial x^j} \right)_\alpha X^j, \tag{1.27}$$

  where the subscript α indicates that the derivatives are evaluated at α. The components of the vector in the new coordinate system become X'^k.

- An n-tuple of numbers Y_j, $j = 1, 2, \ldots, n$, associated with a point α in $\mathbb{E}^n$ are said to be the components of a *covariant vector* at the point α in a given coordinate system x^j if they transform, on a change of coordinate system to x'^j, according to

$$Y'_k = \sum_{j=1}^{n} \left(\frac{\partial x^j}{\partial x'^k} \right)_\alpha Y_j. \tag{1.28}$$

  The components of the vector in the new coordinate system become Y'_k.

Comments 1.2.3(1) Coordinate independence and examples

C1 The set of all contravariant vectors at a given point $\alpha \in \mathbb{E}^n$ form a vector space V_α under the usual component-wise addition and scalar multiplication rules for contravariant vectors. The same is true for covariant vectors.

C2 The above description of vectors appears explicitly coordinate dependent, in sharp contrast to the definition of functions. One would guess that such a description is merely a numerical representation of vectors in various coordinate systems, and that there should be an intrinsic definition which is not explicitly dependent on coordinates. This is indeed the case and we shall introduce such an intrinsic definition in the next section, where we will also

[11] Synge and Schild (1966).

see why the components of the vectors should transform in the way they do on a change of coordinates.

C3 In classical mechanics the state of a particle moving in $\mathbb{E}^3$ can be determined by its position specified by rectangular Cartesian coordinates x^j and its momentum p_j canonically conjugate to x^j. Associated with the particle's motion we have the following vectors:

1. Displacement vector The particle's coordinate displacement Δx^j form the components of a contravariant vector since a simple differentiation gives

$$\Delta x'^j = \sum_{k=1}^{n} \frac{\partial x'^j}{\partial x^k} \Delta x^k, \tag{1.29}$$

which is the transformation law for contravariant vectors.

2. Momentum vector We know from mechanics that a particle's momentum is a covariant vector with components denoted by p_j. This knowledge enables us to calculate, for example, the components of the momentum in different coordinate systems. Consider the motion in $\mathbb{E}^3$ with the usual rectangular Cartesian coordinates (x, y, z) and the corresponding momenta (p_x, p_y, p_z). In the usual spherical coordinates (r, θ, φ) we shall denote the corresponding momenta conjugate to (r, θ, φ) by $(p_r, p_\theta, p_\varphi)$. Using the covariant vector transformation law we obtain:

$$\begin{aligned} p_r &= \frac{\partial x}{\partial r} p_x + \frac{\partial y}{\partial r} p_y + \frac{\partial z}{\partial r} p_z \\ &= \frac{x}{\sqrt{x^2+y^2+z^2}} p_x + \frac{y}{\sqrt{x^2+y^2+z^2}} p_y + \frac{z}{\sqrt{x^2+y^2+z^2}} p_z. \end{aligned} \tag{1.30}$$

$$p_\theta = \frac{\partial x}{\partial \theta} p_x + \frac{\partial y}{\partial \theta} p_y + \frac{\partial z}{\partial \theta} p_z \tag{1.31}$$

$$= \frac{xz}{\sqrt{x^2+y^2}} p_x + \frac{yz}{\sqrt{x^2+y^2}} p_y - \sqrt{x^2+y^2}\, p_z. \tag{1.32}$$

$$\begin{aligned} p_\varphi &= \frac{\partial x}{\partial \varphi} p_x + \frac{\partial y}{\partial \varphi} p_y + \frac{\partial z}{\partial \varphi} p_z \\ &= -y p_x + x p_y. \end{aligned} \tag{1.33}$$

1.2.4 Contravariant, covariant and mixed tensors

Definition 1.2.4(1) Different types of tensors

- A set of n^2 numbers C^{jk}, $j, k = 1, 2, \ldots, n$, associated with a point $\alpha \in \mathbb{E}^n$ are said to be the components of a *contravariant tensor* of the second order at the point α in a given coordinate system x^j if they transform, on a change of coordinate system to x'^j, according to

$$C'^{rs} = \sum_{j,k=1}^{n} \left(\frac{\partial x'^r}{\partial x^j}\right)_\alpha \left(\frac{\partial x'^s}{\partial x^k}\right)_\alpha C^{jk}, \tag{1.34}$$

where the derivatives are evaluated at α. The components of the tensor in the new coordinate system become C'^{rs}.

- A set of n^2 numbers $C_{jk},\ j,k = 1,2,\ldots,n$, associated with a point α in $\mathbb{E}^n$ are said to be the components of a *covariant tensor* of the second order at the point α in a given coordinate system x^j if they transform, on a change of coordinate system to x'^j, according to

$$C'_{rs} = \sum_{j,k=1}^{n} \left(\frac{\partial x^j}{\partial x'^r}\right)_\alpha \left(\frac{\partial x^k}{\partial x'^s}\right)_\alpha C_{jk}. \tag{1.35}$$

  The components of the tensor in the new coordinate system become C'_{rs}.

- A set of n^2 numbers $C^j_k,\ j,k = 1,2,\ldots,n$, associated with a point α in $\mathbb{E}^n$ are said to be the components of a *mixed tensor* of the second order at the point α in a given coordinate system x^j if they transform, on a change of coordinate system to x'^j, according to

$$C'^r_s = \sum_{j,k=1}^{n} \left(\frac{\partial x'^r}{\partial x^j}\right)_\alpha \left(\frac{\partial x^k}{\partial x'^s}\right)_\alpha C^j_k. \tag{1.36}$$

  The components of the tensor in the new coordinate system become C'^r_s.

Comments 1.2.4(1) Outer products, contraction and interior product, symmetric tensors

C1 The above definitions are a generalization of contravariant and covariant vectors. They can be extended further in a straightforward manner to define tensors of higher orders. In view of the similarity of these definitions, we can refer to a vector as a tensor of first order. We can also formally call a coordinate independent number associated with a point α a tensor of zero order at α.

C2 A second order covariant tensor C_{jk} is called *symmetric* if $C_{jk} = C_{kj}$ and *anti-symmetric* if $C_{kj} = -C_{jk}$, and likewise for contravariant tensors. Anti-symmetric tensors will be seen to play a crucial role in formulating Hamiltonian mechanics.

C3 Vectors and tensors of different types may be combined together to produce tensors of higher orders. For example, a covariant vector Y_i and a contravariant vector X^j can be combined to form a mixed tensor with components $C_i^j = Y_i X^j$. Such a combination is known as an *outer product*. The outer product of a covariant vector of components Y_i and a contravariant tensor of the second order of components C^{jk} produces a mixed tensor of the third order of components $C_i^{jk} = Y_i C^{jk}$.

C4 The order of a mixed tensor can be reduced by two by summing over a superscript and a matching subscript. For example a third order mixed tensor C_i^{jk} is reduced to a first order tensor, i.e., a contravariant vector with components X^j, by

$$X^j = \sum_{k=1}^{n} C_k^{jk}. \tag{1.37}$$

Such a process is known as a *contraction*.

C5 Applying contraction to the outer product $C_j^k = Y_j X^k$ produces a tensor of zero order:

$$\sum_{k=1}^{n} C_k^k = \sum_{k=1}^{n} Y_k X^k. \tag{1.38}$$

This number is the same in different coordinate systems, i.e.,

$$\sum_{j=1}^{n} Y_j X^j = \sum_{r=1}^{n} Y_r' X'^r, \tag{1.39}$$

since

$$\begin{aligned}
\sum_{r=1}^{n} Y_r' X'^r &= \sum_{r=1}^{n} \left(\sum_{k=1}^{n} \frac{\partial x^k}{\partial x'^r} Y_k \sum_{j=1}^{n} \frac{\partial x'^r}{\partial x^j} X^j \right) \\
&= \sum_{j,k=1}^{n} Y_k \left(\sum_{r=1}^{n} \frac{\partial x^k}{\partial x'^r} \frac{\partial x'^r}{\partial x^j} \right) X^j \\
&= \sum_{j=1}^{n} Y_j X^j,
\end{aligned} \tag{1.40}$$

since

$$\sum_{r=1}^{n} \frac{\partial x^k}{\partial x'^r} \frac{\partial x'^r}{\partial x^j} = \delta_j^k. \tag{1.41}$$

We call such a number a *scalar* or an *invariant*.

C6 Another process which produces new tensors is a combination of outer product and contraction. For example, given a second rank covariant tensor C_{jk} and a contravariant vector X^r their outer product is a third rank mixed tensor $C^r_{jk} = C_{jk}X^r$. One can then carry out a contraction to produce a covariant vector

$$Y_j = \sum_{k=1}^{n} C^k_{jk} = \sum_{k=1}^{n} C_{jk}X^k. \tag{1.42}$$

This process enables us to relate a contravariant vector X^r to a covariant vector Y_j through a covariant tensor C_{jk}. We shall give a formal discussion of this procedure later in §1.4.4 under the heading of *interior product*.

C7 A covariant vector Y_j at a point α generates a mapping of the set of all contravariant vectors at α to the reals by contraction, i.e., the covariant vector Y_j maps a contravariant vector X^j to the number

$$\sum_{j=1}^{n} Y_j X^j \in \mathbb{R}. \tag{1.43}$$

We can put this mapping on a more formal footing. We know that the set of all contravariant vectors at α form a vector space V_α. A covariant vector Y_j at α then induces a mapping $\mathcal{Y}_\alpha$ of V_α to $\mathbb{R}$ by contraction:

$$\mathcal{Y}_\alpha : V_\alpha \mapsto \mathbb{R} \quad \text{by} \quad X^j \mapsto \mathcal{Y}_\alpha(X^j) = \sum_{j=1}^{n} Y_j X^j. \tag{1.44}$$

This mapping is linear. We shall see later that it is possible to turn this process around to define a covariant vector at α as a linear mapping of V_α to $\mathbb{R}$.

C8 We can introduce the concept of a *vector field* by assigning a vector of the same type at each point in $\mathbb{E}^n$, and similarly for *tensor fields*. Most of the operations introduced for vectors carry over for vector fields, e.g., contraction between contravariant and covariant vector fields.

1.3 Differential Operators, Vectors and Fields

In view of the non-uniqueness of coordinate systems it would be highly desirable to define geometric quantities without direct reference to coordinates. The idea is to start with functions defined on $\mathbb{E}^n$ as primary objects since they are coordinate-independent. We can then introduce other geometric objects in terms of their action on these functions. So, our starting point is the set $C^\infty(\mathbb{E}^n)$ of smooth functions on $\mathbb{E}^n$. Since all the functions are defined on the same domain, i.e., $\mathbb{E}^n$, we can carry out all usual algebraic operations, e.g., we can add and multiply these functions, without having to impose any conditions. On the other extreme we have the set $C^\infty(\alpha)$ of smooth functions each defined on a neighbourhood of the point $\alpha \in \mathbb{E}^n$. Let f and g be members of $C^\infty(\alpha)$ with f defined on domain $\mathcal{D}(f)$ and g defined on domain $\mathcal{D}(g)$. Since both $\mathcal{D}(f)$ and $\mathcal{D}(g)$ contain the point α we would generally have

$$\mathcal{D}(f) \cap \mathcal{D}(g) \neq \emptyset, \quad \mathcal{D}(f) \neq \mathcal{D}(g). \tag{1.45}$$

We can define the sum $f+g$ and the product fg as functions on the domain $\mathcal{D}(f) \cap \mathcal{D}(g)$ according to:

$$f+g : \mathcal{D}(f) \cap \mathcal{D}(g) \mapsto \mathbb{R} \quad \text{by} \quad (f+g)(\alpha') = f(\alpha') + g(\alpha'), \tag{1.46}$$

$$fg : \mathcal{D}(f) \cap \mathcal{D}(g) \mapsto \mathbb{R} \quad \text{by} \quad (fg)(\alpha') = f(\alpha')g(\alpha'). \tag{1.47}$$

Here $\alpha' \in \mathcal{D}(f) \cap \mathcal{D}(g)$. With this introduction we can now define further operations on the functions in $C^\infty(\alpha)$ and in $C^\infty(\mathbb{E}^n)$.

1.3.1 Differential operators and derivations

Definition 1.3.1(1) Linear operators acting on $C^\infty(\alpha)$ and $C^\infty(\mathbb{E}^n)$

- A mapping $\widehat{A}_\alpha$ of $C^\infty(\alpha)$ into $\mathbb{R}$ with the property

  $$\widehat{A}_\alpha(af + b\,g) = a\,\widehat{A}_\alpha(f) + b\,\widehat{A}_\alpha(g) \quad \forall f, g \in C^\infty(\alpha), \tag{1.48}$$

  where $a, b \in \mathbb{R}$, is called a *linear operator at a point* $\alpha \in \mathbb{E}^n$ on the set of smooth functions $C^\infty(\alpha)$.

- A mapping $\widehat{A}$ of $C^\infty(\mathbb{E}^n)$ into $C^\infty(\mathbb{E}^n)$ with the property

  $$\widehat{A}(af + b\,g) = a\,\widehat{A}(f) + b\,\widehat{A}(g) \quad \forall f, g \in C^\infty(\mathbb{E}^n), \tag{1.49}$$

  where $a, b \in \mathbb{R}$, is called a *linear operator* on the set of smooth functions $C^\infty(\mathbb{E}^n)$.

Comments 1.3.1(1) Properties

C1 A linear operator $\widehat{A}_\alpha$ gives rise to a real number $\widehat{A}_\alpha(f)$ for any $f \in C^\infty(\alpha)$. If f is a zero function then $\widehat{A}_\alpha(f) = 0$. In contrast a linear operator $\widehat{A}$ on $C^\infty(\mathbb{E}^n)$ gives rise to a new smooth function $\widehat{A}(f)$ on $\mathbb{E}^n$.

C2 A linear operator $\widehat{A}_\alpha$ at a point α depends only on the behaviour of local functions in the neighbourhood of α. Consequently, if two local functions

f and g agree on some neighbourhood of α, then $\widehat{A}_\alpha(f) = \widehat{A}_\alpha(g)$. This is not true for a linear operator $\widehat{A}$ on $C^\infty(I\!\!E^n)$ in general.

Definition 1.3.1(2) Differential operators

- A linear operator $\widehat{A}_\alpha$ on $C^\infty(\alpha)$ at a point α is called a *differential operator of the first order* at α if there is a coordinate system x^j in a neighbourhood of α such that

$$\widehat{A}_\alpha(f) = \sum_{j=1}^{n} A_\alpha^j \left(\frac{\partial f}{\partial x^j}\right)_\alpha \qquad \forall f \in C^\infty(\alpha), \tag{1.50}$$

  where $A_\alpha^j \in I\!\!R$ are independent of f but are dependent on the coordinate system used.

- A linear operator $\widehat{A}$ on $C^\infty(I\!\!E^n)$ is called a *smooth differential operator of the first order* on $C^\infty(I\!\!E^n)$ if there is a coordinate system x^j such that

$$\widehat{A}(f) = \sum_{j=1}^{n} A^j \left(\frac{\partial f}{\partial x^j}\right) \qquad \forall f \in C^\infty(I\!\!E^n), \tag{1.51}$$

  where $A^j \in C^\infty(I\!\!E^n)$ are independent of f, but are dependent on the coordinate system used.

Comments 1.3.1(2) Properties of differential operators

C1 We often express $\widehat{A}_\alpha$ and $\widehat{A}$ respectively as

$$\widehat{A}_\alpha = \sum_{j=1}^{n} A_\alpha^j \left(\frac{\partial}{\partial x^j}\right)_\alpha \quad \text{and} \quad \widehat{A} = \sum_{j=1}^{n} A^j \left(\frac{\partial}{\partial x^j}\right). \tag{1.52}$$

A differential operator $\widehat{A}$ is not smooth if A^j is not smooth.

C2 These operators satisfy the following properties:

$$\widehat{A}_\alpha(fg) = \widehat{A}_\alpha(f)\, g(\alpha) + f(\alpha)\, \widehat{A}_\alpha(g) \quad \forall f,\ g \in C^\infty(\alpha), \tag{1.53}$$

$$\widehat{A}(fg) = \widehat{A}(f)\, g + f\, \widehat{A}(g) \quad \forall f,\ g \in C^\infty(I\!\!E^n). \tag{1.54}$$

These properties reflect the product rule of differentiation and are therefore characteristic of a differential operator. A linear operator which does not possess these properties is not a differential operator, e.g., a multiplication operator is a linear operator but not a differential operator. As we shall see later Eqs. (1.53) and (1.54) can lead to an intrinsic definition of differential operators.

C3 Some authors would include an additive term in the definition of differential operators, i.e., they would regard operators of the form

$$\widehat{C}(f) = \sum_{j=1}^{n} A^j \left(\frac{\partial f}{\partial x^j} \right) + Bf, \tag{1.55}$$

where B is a function on $\mathbb{E}^n$, as differential operators. We shall not adopt this definition since operators of this form do not obey the product rule of differentiation shown in Eqs. (1.53) and (1.54) unless $B = 0$.

C4 Although defined explicitly through a specific coordinate system differential operators are in fact coordinate independent. Let us examine what would happen when we move to a new coordinate system x'^j. The function f is now a function of x'^j. An application of normal rules of differentiation yields:

$$\widehat{A}_\alpha(f) = \sum_{j=1}^{n} A^j_\alpha \left(\frac{\partial f}{\partial x^j} \right)_\alpha \tag{1.56}$$

$$= \sum_{j=1}^{n} A^j_\alpha \left\{ \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x'^k} \right)_\alpha \left(\frac{\partial x'^k}{\partial x^j} \right)_\alpha \right\}. \tag{1.57}$$

It follows that we can express $\widehat{A}_\alpha(f)$ in coordinates x'^j as

$$\widehat{A}_\alpha(f) = \sum_{k=1}^{n} A'^k_\alpha \left(\frac{\partial f}{\partial x'^k} \right)_\alpha, \tag{1.58}$$

where

$$A'^k_\alpha = \sum_{j=1}^{n} \left(\frac{\partial x'^k}{\partial x^j} \right)_\alpha A^j_\alpha. \tag{1.59}$$

We conclude that $\widehat{A}_\alpha$ is a coordinate independent quantity in the sense that:

1. $\widehat{A}_\alpha$ maps every function $f \in C^\infty(\alpha)$ to a real number $\widehat{A}_\alpha(f)$, and this value $\widehat{A}_\alpha(f)$ is the same in all coordinate systems.

2. $\widehat{A}_\alpha$ can be written down in the same form in any coordinate system, i.e., we have

$$\widehat{A}_\alpha = \sum_{j=1}^{n} A^j_\alpha \left(\frac{\partial}{\partial x^j} \right)_\alpha = \sum_{k=1}^{n} A'^k_\alpha \left(\frac{\partial}{\partial x'^k} \right)_\alpha, \tag{1.60}$$

   where A'^k_α to A^j_α are related by Eq. (1.59).

What has been said above applies to $\widehat{A}$. We have

$$\widehat{A}(f) = \sum_{j=1}^{n} A^j \left(\frac{\partial f}{\partial x^j}\right) = \sum_{k=1}^{n} A'^k \left(\frac{\partial f}{\partial x'^k}\right), \tag{1.61}$$

where

$$A'^k = \sum_{j=1}^{n} \left(\frac{\partial x'^k}{\partial x^j}\right) A^j. \tag{1.62}$$

In fact, differential operators can be introduced in a more abstract manner without explicit reference to any coordinate system. The idea is to employ the characteristic property of differentiation shown in Eqs. (1.53) and (1.54) to define differential operators.

C5 We shall proceed to use differential operators to establish an intrinsic definition of vectors and vector fields and other geometric objects.[12]

Definition 1.3.1(3) Derivations

- A *derivation on* $C^\infty(\alpha)$ at a point α is a linear operator $\widehat{X}_\alpha$ at the point α such that

$$\widehat{X}_\alpha(fg) = \widehat{X}_\alpha(f)g(\alpha) + f(\alpha)\widehat{X}_\alpha(g) \quad \forall f,\ g \in C^\infty(\alpha). \tag{1.63}$$

- A *derivation on* $C^\infty(\mathbb{E}^n)$ is a linear operator $\widehat{X}$ on $C^\infty(\mathbb{E})$ such that

$$\widehat{X}(fg) = \widehat{X}(f)g + f\widehat{X}(g) \quad \forall f,\ g \in C^\infty(\mathbb{E}^n). \tag{1.64}$$

Comments 1.3.1(3) Differential operators, derivations and contravariant vectors

C1 It can be shown that given a derivation $\widehat{X}_\alpha$ at α there exists a coordinate system x^j and n real numbers X^j_α such that

$$\widehat{X}_\alpha(f) = \sum_{j=1}^{n} X^j_\alpha \left(\frac{\partial f}{\partial x^j}\right)_\alpha \quad \forall f \in C^\infty(\alpha), \tag{1.65}$$

where X^j_α are independent of f. It follows that we can express $\widehat{X}_\alpha$ as a differential operator, i.e., we can simply write

$$\widehat{X}_\alpha = \sum_{j=1}^{n} X^j_\alpha \left(\frac{\partial}{\partial x^j}\right)_\alpha. \tag{1.66}$$

[12] Brickell and Clark (1970), Isham (1989), Darling (1994).

It should not be so surprising that derivations are related to differential operators since they satisfy the distintive product rule of differentiation.

C2 Given a non-zero derivation $\widehat{X}_\alpha$, i.e., one with $X^j_\alpha \neq 0$ for some j, it is possible to transform to new coordinates x'^j in which $\widehat{X}_\alpha$ takes the simple form, i.e.,

$$\widehat{X}_\alpha = \frac{\partial}{\partial x'^1}. \tag{1.67}$$

C3 A derivation $\widehat{X}$ on $C^\infty(\mathbb{E}^n)$ can also be shown to be expressible as a differential operator on $C^\infty(\mathbb{E}^n)$, i.e., we have[13]

$$\widehat{X} = \sum_{j=1}^{n} X^j \left(\frac{\partial}{\partial x^j}\right), \quad X^j \in C^\infty(\mathbb{E}^n). \tag{1.68}$$

In other words smooth differential operators of the first order are identifiable with derivations.

C4 Physicists are more used to defining things explicitly or constructively. However, it is often more desirable to define a quantity in terms of its characteristic features. Such a definition enables us to appreciate the concept better, often leading to extension of the concept to new and more general situations. In the present case one can appreciate the coordinate independent nature of differential operators in terms of the concept of derivations which are manifestly coordinate independent.

C5 In a given coordinate system x^j the operator $\widehat{X}_\alpha$ is characterized by n numbers, X^j_α. On a coordinate transformation from x^j to x'^j we have, following Eqs. (1.59) and (1.60),

$$\widehat{X}_\alpha = \sum_{j=1}^{n} X^j_\alpha \left(\frac{\partial}{\partial x^j}\right)_\alpha \tag{1.69}$$

$$= \sum_{k=1}^{n} X'^k_\alpha \left(\frac{\partial}{\partial x'^k}\right)_\alpha, \tag{1.70}$$

where

$$X'^k_\alpha = \sum_{j=1}^{n} \left(\frac{\partial x'^k}{\partial x^j}\right)_\alpha X^j_\alpha. \tag{1.71}$$

Compared with Eq. (1.27) we can see that X^j_α change like the components of a contravariant vector on a coordinate transformation. Further studies reveal

[13] Matsushima (1972) p. 73.

that differential operators are in fact the intrinsic and coordinate independent definition of contravariant vectors. This being the case we can now appreciate why the numerical components of a contravariant vector transform the way they do.

C6 We shall see later that covariant vectors are definable in a coordinate independent manner in terms of differentials of functions.

C7 While physicists are more used to the term contravariant vectors, differential geometers prefer to use the term *tangent vectors*. So, we have in effect the same quantity being referred to by four different names, i.e., **contravariant vectors, tangent vectors, derivations and differential operators**. When we want to concentrate on its numerical components we would employ the terms contravariant vectors. In other situations we may prefer the terms tangent vectors or differential operators.

1.3.2 Tangent vectors, tangent vector fields and their integral curves

Definition 1.3.2(1) Tangent vectors, tangent spaces and tangent vector fields

- A *tangent vector* at a point α is a derivation $\widehat{X}_\alpha$ on $C^\infty(\alpha)$.
- The set $\widehat{T}_\alpha(\mathbb{E}^n)$ of all tangent vectors at a point α, endowed with a natural vector space structure under addition and scalar multiplication defined by

$$\Big(\widehat{X}_\alpha + \widehat{Y}_\alpha\Big)(f) = \widehat{X}_\alpha(f) + \widehat{Y}_\alpha(f), \quad \Big(a\widehat{X}_\alpha\Big)(f) = a\widehat{X}_\alpha(f), \qquad (1.72)$$

  where $a \in \mathbb{R}$ and $\widehat{X}_\alpha,\ \widehat{Y}_\alpha \in \widehat{T}_\alpha(\mathbb{E}^n)$, is called the *tangent space* at α.
- A derivation $\widehat{X}$ on $C^\infty(\mathbb{E}^n)$ is called a *tangent vector field* on $\mathbb{E}^n$, or simply a *vector field*. A point at which $\widehat{X}$ vanishes is called a *critical point* of the vector field.

Comments 1.3.2(1) Coordinate independence and terminology

C1 Tangent vectors and vector fields have been defined in a way which is manifestly independent of coordinates.

C2 From §1.3.1C(3) C1 we can see that:

1. Tangent vectors are expressible in the form of differential operators by Eq. (1.66).

2. Vector fields can be written in the form of smooth differential operators using Eq. (1.68). It follows from §1.3.1C(3) C2 that if a given vector field $\widehat{X}$ does not have a critical point in the neighbourhood of a given point α, then new coordinates x'^j exist in which the vector field takes the simple form
$$\widehat{X} = \frac{\partial}{\partial x'^1} \tag{1.73}$$
in that neighbourhood. Generally $\widehat{X}$ may not be expressible in such a simple form globally over the entire $I\!E^n$.

C3 A vector field $\widehat{X}$ gives rise to a tangent vector $\widehat{X}_\alpha$ at each point $\alpha \in I\!E^n$ by
$$\widehat{X}_\alpha(f) = \Big(\widehat{X}(f)\Big)(\alpha). \tag{1.74}$$
Explicitly we have
$$\widehat{X} = \sum_{j=1}^{n} X^j \left(\frac{\partial}{\partial x^j}\right) \quad \Rightarrow \quad \widehat{X}_\alpha = \sum_{j=1}^{n} X^j_\alpha \left(\frac{\partial}{\partial x^j}\right)_\alpha . \tag{1.75}$$
Conversely a vector field $\widehat{X}$ may be regarded as a *smooth* assignment of tangent vectors throughout $I\!E^n$, with one tangent vector $\widehat{X}_\alpha$ at each point α. Here, *smoothness* means that the function defined by $\alpha \mapsto \widehat{X}_\alpha(f)$ is smooth for every $f \in C^\infty(I\!E^n)$.

C4 Differential operators at α are closely related to the tangents at α to curves passing through α. This is why $\widehat{X}_\alpha$ is called a tangent vector. Before discussing this in detail we have to introduce the concept of differentiable curves first.

Definition 1.3.2(2) Curves in $I\!E^n$

- Let J be an open interval of $I\!R$ containing the origin 0. A mapping
$$\sigma : J \mapsto I\!E^n, \tag{1.76}$$
which associates each real number $\tau \in J$ to a point $\alpha(\tau)$ in $I\!E^n$, is called a *curve* in $I\!E^n$ starting from the point $\alpha(0)$. The interval J is called the *domain of the curve*.

Comments 1.3.2(2) Curves, their tangents and operators

C1 As the parameter τ varies, the map σ traces out a set of points $\alpha(\tau)$ in $I\!E^n$ with coordinates $x^j(\alpha(\tau))$. The curve is said to be *smooth* or *differentiable* if $x^j(\alpha(\tau))$ are smooth functions of τ. From now on a curve means a smooth curve, unless stated otherwise.

C2 Traditionally the *tangent to a curve* at a point $\alpha(\tau)$ on the curve is defined to be a contravariant vector with components

$$X_\alpha^j = \frac{dx^j(\alpha(\tau))}{d\tau} \tag{1.77}$$

in a given coordinate system x^j. In our present set-up we can construct a differential operator $\widehat{X}_\alpha$ at $\alpha(\tau)$ on the curve by

$$\widehat{X}_\alpha = \sum_{j=1}^n X_\alpha^j \left(\frac{\partial}{\partial x^j}\right)_\alpha, \quad \text{with} \quad X_\alpha^j = \frac{dx^j(\alpha(\tau))}{d\tau} \tag{1.78}$$

and call this differential operator the *tangent vector to the curve* σ at $\alpha(\tau)$. A curve generates a family of tangent vectors along itself. A family of curves covering the entire space $\mathbb{E}^n$ will generate a vector field on $\mathbb{E}^n$. More interesting is the converse, i.e., whether a given vector field can generate curves in $\mathbb{E}^n$.

Definition 1.3.2(3) Maximal integral curves of a vector field

- The operator $\widehat{X}_\alpha$ at $\alpha(\tau)$ on a given curve σ defined by

$$\widehat{X}_\alpha = \sum_{j=1}^n X_\alpha^j \left(\frac{\partial}{\partial x^j}\right)_\alpha, \quad \text{with} \quad X_\alpha^j = \frac{dx^j(\alpha(\tau))}{d\tau} \tag{1.79}$$

  is called the *tangent vector to the curve* σ at $\alpha(\tau)$.

- Let σ be a curve. If at every point $\alpha(\tau)$ on the curve the tangent vector to the curve coincides with the tangent vector at $\alpha(\tau)$ given rise by a vector field $\widehat{X}$ according to Eqs. (1.74) and (1.75) the curve σ is called an *integral curve* of the vector field $\widehat{X}$ starting from $\alpha(0)$.

- The integral curve starting from $\alpha(0)$ defined on the union of the domains of all the integral curves of $\widehat{X}$ starting from $\alpha(0)$ is called the *maximal integral curve* of the vector field $\widehat{X}$ starting from $\alpha(0)$.

- A vector field is said to be *complete* if the maximal integral curve starting from every point in $\mathbb{E}^n$ has the entire real line $\mathbb{R}$ as its domain. A vector field which is not complete is said to be *incomplete*.

Comments 1.3.2(3) Completeness of vector fields

C1 Intuitively the relationship between a vector field and its integral curves resembles that of a velocity field of a liquid in flow and its flow lines or an electric field and its field lines. The maximal integral curve starting from $\alpha(0)$ is the integral curve starting from $\alpha(0)$ defined on the largest domain. This is to contrast with other "shorter" integral curves starting from $\alpha(0)$. From now on an integral curve means a maximal integral curve.

C2 Based on the properties of differential equations we can show that there exists one and only one (maximal) integral curve of a vector field starting from any given point. Let σ be the integral curve of the vector field

$$\widehat{X} = \sum_{j=1}^{n} X^j \frac{\partial}{\partial x^j}, \tag{1.80}$$

starting from the point $\alpha(0)$. Let $x^j(\alpha(\tau))$ be the coordinates of the point $\alpha(\tau)$ on the curve. Then $x^j(\alpha(\tau))$ and $X^j(\alpha(\tau))$ are smooth functions of τ satisfying the following first order differential equations:

$$\frac{dx^j(\alpha(\tau))}{d\tau} = X^j(\alpha(\tau)). \tag{1.81}$$

Given $X^j(\alpha(\tau))$ one can solve for $x^j(\alpha(\tau))$ as functions of τ. If all these functions $x^j(\alpha(\tau))$ corresponding to integral curves starting from every point in $I\!E^n$ are defined for all values of $\tau \in (-\infty, \infty)$, then the vector field is complete. In §1.3.2E(1) below we shall examine a number of vector fields and their integral curves explicitly.

C3 Sometimes we want to confine ourselves to an open subset Λ of $I\!E^n$, i.e., we desire to define vector fields and examine their integral curves within a subset Λ. Clearly we can do this and everything introduced on $I\!E^n$ can be carried over on Λ.

C4 A Euclidean space structure is not necessary for the introduction of vector fields and their integral curves. A similar statement applies to the introduction of covariant vectors.

Examples 1.3.2(1) Vector fields, integral curves and completeness

E1 A rectangular Cartesian coordinate system in $I\!E^3$ is often denoted by (x, y, z). A corresponding Cartesian coordinates in $I\!E^2$ is denoted by (x, y). A vector field on $I\!E^3$ is of the form

$$\widehat{X} = X^1 \frac{\partial}{\partial x} + X^2 \frac{\partial}{\partial y} + X^3 \frac{\partial}{\partial z}, \tag{1.82}$$

and its integral curves satisfy equations of the form

$$\frac{dx}{d\tau} = X^1, \quad \frac{dy}{d\tau} = X^2, \quad \frac{dz}{d\tau} = X^3. \tag{1.83}$$

We shall adopt this notation in the ensuing examples many of which will have direct physical applications in later chapters.

E2 Consider the vector field

$$\widehat{X} = \frac{\partial}{\partial x} \tag{1.84}$$

in $\mathbb{E}^3$. The integral curve starting from any point (x_0, y_0, z_0) satisfies the following equations:

$$\frac{dx}{d\tau} = 1, \quad \frac{dy}{d\tau} = 0, \quad \frac{dz}{d\tau} = 0. \tag{1.85}$$

The solution is

$$x(\tau) = \tau + x_0, \quad y(\tau) = y_0, \quad z(\tau) = z_0 \tag{1.86}$$

valid for $\tau \in (-\infty, \infty)$. This is true for any initial point (x_0, y_0, z_0) in $\mathbb{E}^3$. This vector field is therefore complete.

E3 Consider the vector field

$$\widehat{X} = x\,\frac{\partial}{\partial x} + y\,\frac{\partial}{\partial y} \tag{1.87}$$

in $\mathbb{E}^3$. An integral curve starting from a point (x_0, y_0, z_0) is a solution of

$$\frac{dx}{d\tau} = x, \quad \frac{dy}{d\tau} = y, \quad \frac{dz}{d\tau} = 0. \tag{1.88}$$

The solution satisfying initial conditions $x(0) = x_0,\ y(0) = y_0,\ z(0) = z_0$ is

$$x(\tau) = x_0\, e^{\tau}, \quad y(\tau) = y_0\, e^{\tau}, \quad z(\tau) = z_0, \quad \text{for} \quad \tau \in (-\infty, \infty). \tag{1.89}$$

This vector field vanishes along the z-axis so that the integral curve starting from any point $(0, 0, z_0)$ on the z-axis is simply the point itself, i.e.,

$$x(\tau) = 0, \quad y(\tau) = 0, \quad z(\tau) = z_0, \quad \text{for} \quad \tau \in (-\infty, \infty). \tag{1.90}$$

This vector field is complete.

E4 Consider the vector field

$$\widehat{X} = y\,\frac{\partial}{\partial x} - x\,\frac{\partial}{\partial y} \tag{1.91}$$

in $I\!\!E^3$. An integral curve starting from a point (x_0, y_0, z_0) is a solution of

$$\frac{dx}{d\tau} = y, \quad \frac{dy}{d\tau} = -x, \quad \frac{dz}{d\tau} = 0 \tag{1.92}$$

satisfying initial conditions $x(0) = x_0, y(0) = y_0, z(0) = z_0$. The solution can be written down conveniently in terms of the usual cylindrical coordinates (r, θ, z) where

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(y/x). \tag{1.93}$$

The solution is

$$x(\tau) = r\sin(\tau + \theta), \quad y(\tau) = r\cos(\tau + \theta), \quad z(\tau) = z_0, \tag{1.94}$$

where

$$x_0 = r\sin\theta, \quad y_0 = r\cos\theta \quad \text{and} \quad \tau \in (-\infty, \infty). \tag{1.95}$$

The integral curves defined by Eq. (1.94) are circles lying on a plane parallel the x-y planes and centered at a point on the z-axis. This vector field vanishes along the z-axis so that the integral curve starting from any point $(0, 0, z_0)$ on the z-axis is simply the point itself. This vector field is again complete.

This example can be compared with a related vector field defined on a circle given in §1.5.1E(1) E5 below.

E5 A vector field is said to have a *compact support* if it vanishes outside a closed and bounded cube in $I\!\!E^n$. As an illustration, consider a vector field of the form

$$\widehat{X} = \xi\, \frac{d}{dx} \tag{1.96}$$

in $I\!\!E$, where $\xi = \xi(x)$ is a smooth function of compact support $[a, b]$ on $I\!\!E$ and $\xi(x) > 0$ for every $x \in (a, b)$. An integral curve σ starting from $x_0 \in (a, b)$ satisfies

$$\frac{dx}{d\tau} = \xi(x), \tag{1.97}$$

or

$$\int_{x_0}^{x} \frac{dx}{\xi(x)} = \tau. \tag{1.98}$$

We can appreciate that the integral tends to $-\infty$ as x approaches a and to ∞ as x approaches b since $\xi(x)$ tends to zero. It follows that τ can range from $-\infty$ to ∞. In other words such a vector field is complete. This result is also true in $I\!\!E^n$, i.e., generally a vector field of compact support in $I\!\!E^n$ is complete.[14]

[14]Abraham and Marsden (1978) Corollary 2.1.19 on p. 70.

E6 Consider vector fields of the form

$$\widehat{X}^{(k)} = x^k \frac{d}{dx}, \; k = 1, 2, \ldots \tag{1.99}$$

in $I\!\!E$. All these vector fields have a single critical point at $x = 0$. There are two distinct cases:

1. When $k = 1$ we have a vector field $\widehat{X}^{(1)}$ whose integral curves satisfy equation $dx/d\tau = x$. The integral curve starting from x_0 is given by

$$x(\tau) = x_0 e^\tau, \quad \tau \in (-\infty, \infty). \tag{1.100}$$

   The integral curve from $x = 0$ is the point itself. The vector field is therefore complete.

2. When $k > 1$ we have vector fields $\widehat{X}^{(2)}, \widehat{X}^{(3)}, \ldots$ which are all incomplete. As an illustration consider the case with $k = 2$, i.e., the vector field is

$$\widehat{X}^{(2)} = x^2 \frac{d}{dx}. \tag{1.101}$$

   We have the equation $dx/d\tau = x^2$ for integral curves. The integral curve starting from a point $x_0 > 0$ is given by

$$x(\tau) = \frac{x_0}{1 - x_0 \tau}. \tag{1.102}$$

   This solution is not valid for $\tau = 1/x_0$, i.e., the domain of the curve is not the entire real line $I\!\!R$. The vector field is therefore incomplete.

E7 The sum of two complete vector fields is not necessarily complete.[15] Let us illustrate this by considering the following three vector fields on $I\!\!E^2$:

1. The vector field

$$\widehat{X}_1 = y \frac{\partial}{\partial x} \tag{1.103}$$

   is complete with its integral curve starting from the point (x_0, y_0) being

$$x(\tau) = y_0 \tau + x_0, \quad y(\tau) = y_0. \tag{1.104}$$

2. The vector field

$$\widehat{X}_2 = x^2 \frac{\partial}{\partial y} \tag{1.105}$$

   is complete with its integral curves starting from a point (x_0, y_0) given by

$$x(\tau) = x_0, \quad y(\tau) = x_0^2 \tau + y_0. \tag{1.106}$$

[15] For another example see Abraham and Marsden (1978) Exercise 2.2H (i) on p. 99.

3. Now consider
$$\widehat{X}_3 = \widehat{X}_1 + \widehat{X}_2 = y\,\frac{\partial}{\partial x} + x^2\,\frac{\partial}{\partial y}. \tag{1.107}$$
Its integral curves are given by solutions of
$$\frac{dx}{d\tau} = y, \quad \frac{dy}{d\tau} = x^2. \tag{1.108}$$
The integral curve starting from
$$x_0 = 6/a^2, \quad y_0 = 12/a^3, \quad a \in \mathbb{R}^+ \tag{1.109}$$
is given by
$$x(\tau) = \frac{6}{(a-\tau)^2}, \quad y(\tau) = \frac{12}{(a-\tau)^3}. \tag{1.110}$$
This curve is not defined at $\tau = a$, rendering $\widehat{X}_3$ incomplete.

E8 We may desire to restrict ourselves to a subset Λ of $\mathbb{E}^n$. Then whether a vector field is complete or not would depend also on Λ. As examples let us consider two cases:

1. For the open interval $\Lambda = (0,\pi) \subset \mathbb{E}$ with coordinate $x \in (0,\pi)$ we can define the vector field $\widehat{X} = d/dx$. The integral curve starting from, say, the point $x = \pi/2$ is given by $x = \tau + \pi/2$. Note that the domain of the curve is $(-\pi/2, \pi/2)$ instead of the entire real line, and hence $\widehat{X}$ is incomplete in Λ.

2. For the half line $\mathbb{E}^+ = (0,\infty) \subset \mathbb{E}$ with coordinate $x \in (0,\infty)$ we can again define the vector field $\widehat{X} = d/dx$. The integral curve starting from, say, the point $x = \pi/2$ is given by $x = \tau + \pi/2$. The domain of the curve is $(-\pi/2,\infty)$ instead of the entire real line, and hence $\widehat{X}$ is incomplete. Similarly we can introduce such a vector field in the half line $\mathbb{E}^- = (-\infty, 0)$.

E9 The completeness of vector fields is a concept which plays a fundamental role in quantum mechanics. The completeness or otherwise of vector fields is directly relevant to the problem of quantizability. The fact that the sum of two complete vector fields is not necessarily complete will have an important consequence in quantization. We shall return to some of the examples presented here when we investigate quantization in §3.3 in Chapter 3. Some of the seemingly trivial examples presented above turn out to be physically very important when it comes to quantization.

Definition 1.3.2(4) Lie brackets

- The Lie bracket of two vector fields $\widehat{X}, \widehat{Y}$ in $I\!\!E^n$ is the vector field $\widehat{Z}$ on $I\!\!E^n$ defined by

$$\widehat{Z}(f) = \widehat{X}\Big(\widehat{Y}(f)\Big) - \widehat{Y}\Big(\widehat{X}(f)\Big), \quad f \in C^\infty(I\!\!E^n). \tag{1.111}$$

Comments 1.3.2(4) Explicit expressions and properties

C1 Let

$$\widehat{X} = \sum_{j=1}^{n} X^j \frac{\partial}{\partial x^j}, \qquad \widehat{Y} = \sum_{j=1}^{n} Y^j \frac{\partial}{\partial x^j}, \tag{1.112}$$

then we have

$$\widehat{Z} = \sum_{k=1}^{n} Z^k \frac{\partial}{\partial x^k}, \quad Z^k = \sum_{j=1}^{n} \left(X^j \frac{\partial Y^k}{\partial x^j} - Y^j \frac{\partial X^k}{\partial x^j} \right). \tag{1.113}$$

The minus sign in the definition of the bracket eliminates terms with second-order derivatives so that the bracket contains only first-order derivatives. The Lie bracket is often written symbolically as

$$\widehat{Z} = [\widehat{X}, \widehat{Y}] = \widehat{X}\widehat{Y} - \widehat{Y}\widehat{X}. \tag{1.114}$$

C2 It is easy to verify that the Lie bracket is linear in $\widehat{X}$ and $\widehat{Y}$, and that

$$[\widehat{X}, \widehat{Y}] = -[\widehat{Y}, \widehat{X}], \tag{1.115}$$

$$\Big[[\widehat{X}, \widehat{Y}], \widehat{Z}\Big] + \Big[[\widehat{Z}, \widehat{X}], \widehat{Y}\Big] + \Big[[\widehat{Y}, \widehat{Z}], \widehat{X}\Big] = 0. \tag{1.116}$$

The set of all vector fields forms a Lie algebra under Lie bracket.[16]

C3 The Lie bracket of two complete vector fields is not necessarily complete.[17] An example is

$$\widehat{X} = y\,\frac{\partial}{\partial x}, \quad \widehat{Y} = x^2\,\frac{\partial}{\partial y} \tag{1.117}$$

in $I\!\!E^2$. We have

$$\widehat{Z} = [\widehat{X}, \widehat{Y}] = -x^2\,\frac{\partial}{\partial x} + 2xy\,\frac{\partial}{\partial y} \tag{1.118}$$

[16] Abraham and Marsden (1978), Martin (1991).

[17] For another example see Brickell and Clark (1970) Problems 8.2 on p. 139.

whose integral curve starting from $(x_0 > 0,\, y_0)$ is

$$x(\tau) = \frac{x_0}{1 + x_0\,\tau}, \quad y(\tau) = (1 + x_0\,\tau)^2\, y_0 \tag{1.119}$$

which is not defined for $\tau = -1/x_0$. So, $\widehat{Z}$ is incomplete.

1.3.3 Transformation groups and complete vector fields

Consider one-to-one mappings of $\mathbb{E}^n$ onto itself, i.e.,

$$T : \mathbb{E}^n \mapsto \mathbb{E}^n \quad \text{by} \quad \alpha \mapsto \bar{\alpha} = T\alpha. \tag{1.120}$$

In view of the one-to-one and onto nature of T the inverse map T^{-1} exists and is again one-to-one and onto in nature. Let the coordinates of α and $\bar{\alpha}$ in $\mathbb{E}^n$ be x^j and $\bar{x}^j$ respectively. Each coordinate $\bar{x}^k$ is a function of the coordinates x^j.

Definition 1.3.3(1) Differentiable mappings and diffeomorphisms

- The mapping T above is said to be *differentiable* if the coordinates $\bar{x}^k$ are smooth functions of x^j.
- The mapping T above is *diffeomorphic* and is a *diffeomorphism* if the inverse T^{-1} is also differentiable. A diffeomorphism is also known as a *transformation.*

Comments 1.3.3(1) Mappings of spaces of different dimensions

C1 We can define differentiable mappings of spaces of different dimensions. Consider the following mapping of $\mathbb{E}^n$ into $\mathbb{E}^m$:

$$M : \mathbb{E}^n \mapsto \mathbb{E}^m \quad \text{by} \quad \alpha \mapsto \beta = M\alpha. \tag{1.121}$$

Let x^j be the coordinates of $\alpha \in \mathbb{E}^n$ and y^k, $k = 1, 2, \ldots, m$, be the coordinates of $\beta = M\alpha \in \mathbb{E}^m$. The mapping is said to be *differentiable* if y^k are smooth functions of x^j. These mappings are not diffeomorphisms which map spaces of the same dimension.

C2 The notion of differentiable mappings can be extended to more general situations. Consider a mapping of $\mathbb{R} \times \mathbb{E}^n$ into $\mathbb{E}^n$:

$$T : \mathbb{R} \times \mathbb{E}^n \mapsto \mathbb{E}^n \quad \text{by} \quad (\tau, \alpha) \mapsto \bar{\alpha}(\tau) = T(\tau, \alpha), \tag{1.122}$$

where $\tau \in I\!R$ and $\alpha,\, \bar{\alpha}(\tau) \in I\!E^n$. For each τ we have a mapping of $I\!E^n$ onto itself:

$$T_\tau : I\!E^n \mapsto I\!E^n \quad \text{defined by} \quad \alpha \mapsto T_\tau \alpha = \bar{\alpha}(\tau) = T(\tau, \alpha). \tag{1.123}$$

In other words T corresponds to a family of such mappings. Conversely a family of such mappings can be assembled into a single mapping T from $I\!R \times I\!E^n$ to $I\!E^n$.

Let the coordinates of α and $\bar{\alpha}(\tau)$ be x^j and $\bar{x}^j(\tau)$ respectively. Mapping T is said to be differentiable if $\bar{x}^j(\tau)$ are smooth functions of x^j and τ.

Definition 1.3.3(2) One-parameter group of transformations

- Let $T = \{T_\tau : \tau \in I\!R\}$ be a family of transformations of $I\!E^n$ such that[18]
  1. $T_{\tau_2}(T_{\tau_1}\alpha) = T_{\tau_2+\tau_1}\alpha$.
  2. The mapping from $I\!R \times I\!E^n$ to $I\!E^n$ defined by $(\tau, \alpha) \mapsto T_\tau \alpha$ is differentiable.

  Then the family of transformations is called a *one-parameter group of transformations* of $I\!E^n$ or simply a *one-parameter transformation group of* $I\!E^n$.

Comments 1.3.3(2) Transformation groups and vector fields

C1 The group structure of the family of transformations manifests itself with the identity element T_0 corresponding to $\tau = 0$ and with the inverse T_τ^{-1} of T_τ equal to $T_{-\tau}$. Note that

$$(T_{\tau_2} T_{\tau_1})^{-1} = T_{\tau_1}^{-1} T_{\tau_2}^{-1}. \tag{1.124}$$

C2 Consider an example in $I\!E$ with rectangular Cartesian coordinate x. Let T_τ be a family of mappings of $I\!E$ onto itself defined by mapping every $\alpha \in I\!E$ to another point $T_\tau \alpha$ according to the following coordinate expression:

$$x(\alpha) \mapsto x(T_\tau \alpha) = x(\alpha) + \tau, \qquad \tau \in (-\infty, \infty). \tag{1.125}$$

One can see that T_τ form a one-parameter transformation group of $I\!E$, translating every point in $I\!E$ by an amount τ. This group generates a curve σ starting from each point $\alpha(0)$ by

$$x(\alpha(\tau)) = x(\alpha(0)) + \tau, \quad \tau \in (-\infty, \infty). \tag{1.126}$$

[18] Matsushima (1972) p. 79.

This curve in turn gives rise to a complete vector field

$$\widehat{X} = X\,\frac{d}{dx}, \quad \text{where} \quad X = \frac{dx(\alpha(\tau))}{d\tau} = 1 \quad \Rightarrow \quad \widehat{X} = \frac{d}{dx}. \tag{1.127}$$

C3 We can easily reverse the argument in C2 above. Given the complete vector field $\widehat{X} = d/dx$ in $\mathbb{E}$ we can obtain its integral curve σ starting from x. This enables us to define a transformation T_τ of $\mathbb{E}$ by

$$x \mapsto T_\tau x = \sigma(\tau) = x + \tau, \quad \text{for every } \tau \in (-\infty, \infty). \tag{1.128}$$

Then $T = \{T_\tau, \tau \in \mathbb{R}\}$ constitutes a one-parameter transformation group of $\mathbb{E}$. We can illustrate the situation with two more examples:

1. The vector field $\widehat{X} = \xi(x)d/dx$ in Eq. (1.96) is complete in $\mathbb{E}$. The transformation group is given by $x(\tau) = T_\tau x_0$ where $x(\tau)$ is related to x_0 and τ by Eq. (1.98).

2. Consider the vector field in Eq. (1.91). The integral curves are circles given by Eq. (1.94). The transformation group consists of rotations of $\mathbb{E}^3$ about the z-axis.

C4 Generally let $T : \mathbb{R} \times \mathbb{E}^n \mapsto \mathbb{E}^n$ be a one-parameter transformation group of $\mathbb{E}^n$. We can generate a curve σ starting from any $\alpha \in \mathbb{E}^n$ by

$$\alpha(\tau) = T_\tau \alpha, \quad \tau \in \mathbb{R}. \tag{1.129}$$

All these curves are defined on the domain $\mathbb{R}$. We can also generate a vector field with these curves as integral curves by assigning the vector

$$\widehat{X}_\alpha = \sum_{j=1}^{n} X_\alpha^j \frac{\partial}{\partial x^j}\bigg|_\alpha, \quad X_\alpha^j = \frac{dx^j(\alpha(\tau))}{d\tau}\bigg|_{\tau=0} \tag{1.130}$$

to every $\alpha \in \mathbb{E}^n$. Moreover, this is a smooth assignment in view of the differentiable nature of the mapping T. The resulting vector field is complete since its integral curves are defined on the domain $\mathbb{R}$. The converse statement is also true. Given a complete vector field $\widehat{X}$ we can construct a transformation group T of $\mathbb{E}^n$ by defining its element T_τ in terms of the integral curves σ of $\widehat{X}$ by $T_\tau \alpha = \sigma(\tau)$. This group T is called the *transformation group of the vector field* $\widehat{X}$.

C5 We can summarize our discussions above as follows:[19]

> *a one-parameter transformation group of $\mathbb{E}^n$ gives rise to a complete vector field on $\mathbb{E}^n$, and the converse is also true.*

[19] Choquet-Bruhat, de Witt-Morette with Dillard-Bleick (1989) p. 145.

Definition 1.3.3(3) Flow of vector fields and transformation groups

- The *flow* of a complete vector field is defined to be the transformation group of the vector field.

Comments 1.3.3(3) The concept of flow

C1 For a fluid in motion we have a *velocity field* to describe its motion. The fluid particles move along the integral curves of the velocity field. We have an intuitive notion of fluid flow to signify how far and fast the fluid particles move in time. This intuitive notion is now sharpened by a precise definition in terms of the transformation group of the velocity field. This makes sense since this group contains information on how fast and far the fluid moves.

C2 For an incomplete vector field $\widehat{X}$ we still have an integral curve σ starting from every point α. Generally the domain $J(\alpha)$ of the integral curve starting from α is dependent on α and it may not be the entire real line.

C3 It is possible to restrict one's attention to a neighbourhood $\mathcal{N}$ of α and consider the mappings of $\mathcal{N}$ onto new regions $\mathcal{N}'$ by moving every point $\alpha \in \mathcal{N}$ along its integral curve by a common parameter τ. This can be used to establish a concept of local transformations.

Definition 1.3.3(4) Lie derivative and derivative along an integral curve

- The *Lie derivative* of a function f with respect to a vector field $\widehat{X}$ is defined to be the function $\widehat{X}(f)$.
- Let σ be an integral curve of a given vector field $\widehat{X}$ starting from α. Then the *derivative of a function* f along this integral curve at α is defined to be

$$\frac{df(\alpha(\tau))}{d\tau}\bigg|_{\tau=0} = \lim_{\tau\mapsto 0} \frac{f(\alpha(\tau)) - f(\alpha(0))}{\tau}, \quad \alpha(0) = \alpha. \tag{1.131}$$

Comments 1.3.3(4) Relationship between the two derivatives

C1 Given a vector field $\widehat{X} = \sum_{j=1}^{n} X^j \, d/dx^j$ we have

$$\widehat{X}(f) = \sum_{j=1}^{n} X^j \frac{\partial f}{\partial x^j}. \tag{1.132}$$

Let σ be an integral curve of $\widehat{X}$ starting from α. Then

$$\frac{dx^i(\alpha(\tau))}{d\tau}\bigg|_{\tau=0} = X^j\bigg|_{\alpha}. \tag{1.133}$$

We have

$$\frac{df(\alpha(\tau))}{d\tau}\bigg|_{\tau=0} = \sum_{j=1}^{n}\left[\frac{\partial f(\alpha(\tau))}{\partial x^j(\alpha(\tau))}\,\frac{dx^j(\alpha(\tau))}{d\tau}\right]_{\tau=0} \tag{1.134}$$

$$= \sum_{j=1}^{n}\left[\frac{\partial f}{\partial x^j}\,X^j\right]_{\alpha}. \tag{1.135}$$

It follows from Eqs. (1.132) and (1.135) that

$$\frac{df(\alpha(\tau))}{d\tau}\bigg|_{\tau=0} = \widehat{X}_\alpha(f). \tag{1.136}$$

In other words the derivative of a function at a point along the integral curve is equal to the Lie derivative of the function with respect to the vector field at that point.

C2 We can also write down an explicit expression of a complete vector field $\widehat{X}$ in terms of the elements of its transformation group T as follows:

$$\widehat{X} = \sum_{j=1}^{n} X^j\frac{\partial}{\partial x^j}, \quad \text{where} \quad X^j(\alpha) = \frac{dx^j(T_\tau\alpha)}{d\tau}\bigg|_{\tau=0}. \tag{1.137}$$

We also have

$$\frac{df(T_\tau\alpha)}{d\tau}\bigg|_{\tau=0} = \widehat{X}_\alpha(f), \tag{1.138}$$

or

$$\frac{df(T_\tau^{-1}\alpha)}{d\tau}\bigg|_{\tau=0} = -\widehat{X}_\alpha(f). \tag{1.139}$$

These expressions are useful for later applications.

Definition 1.3.3(5) Gradient, divergence and the Laplacian

- Let f be a smooth function on $I\!E^n$ with rectangular Cartesian coordinates x^j. The *gradient* of f, denoted by ∇f, is the vector field

$$\nabla f = \sum_{j=1}^{n}\frac{\partial f}{\partial x^j}\,\frac{\partial}{\partial x^j}. \tag{1.140}$$

- Let $\widehat{X} = \sum_{j=1}^{n} X^j \partial/\partial x^j$ be a vector field on $\mathbb{E}^n$. The *divergence* of $\widehat{X}$, denoted by $\operatorname{div} \widehat{X}$, is the function

$$\operatorname{div} \widehat{X} = \sum_{j=1}^{n} \frac{\partial X^j}{\partial x^j}. \tag{1.141}$$

$\widehat{X}$ is said to be *divergence-free* if $\operatorname{div} \widehat{X}$ vanishes everywhere in $\mathbb{E}^n$.

- The *Laplacian* of a function f, denoted by $\nabla^2 f$, is the function

$$\nabla^2 f = \operatorname{div} \nabla f = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial (x^j)^2}. \tag{1.142}$$

Comments 1.3.3(5) Coordinate independence

C1 The above quantities are actually coordinate independent. We have chosen what appear to be coordinate dependent definitions for familiarity.

C2 As will be seen in Chapter 3 divergence-free vector fields have some simplifying properties in quantization.

1.4 Cotangent Vectors and Differential Forms

Given any vector space V we can define linear mappings of V to $\mathbb{R}$. Such mappings are known as *linear functionals* on the vector space. Under usual addition and scalar multiplication rules the set of all these linear functionals form a vector space called the *dual space* to V, denoted by V^*. As an example consider the set of contravariant vectors X^j at a given point $\alpha \in \mathbb{E}^n$. We know that:

1. This set forms a vector space V_α.[20]
2. A covariant vector Y_j at α induces a mapping $\mathcal{Y}_\alpha$ of V_α to $\mathbb{R}$ through a contraction operation.[21]

In other words a covariant vector gives rise to a linear functional on the space of contravariant vectors at α. It turns out that we can actually identify a covariant vector with such a linear functional, i.e., we can identify the mapping $\mathcal{Y}_\alpha$ with the covariant vector Y_j. This argument will now be used to formulate the notion of cotangent vectors as linear functionals on the space of tangent vectors. Cotangent vectors are the intrinsic definition of covariant vectors in the same way that tangent vectors are the intrinsic definition of contravariant vectors.

[20] See §1.2.3C(1) C1.
[21] See §1.2.4C(1) C7 and §1.2.4C(1) C4, C5.

1.4.1 Cotangent vectors, differentials and one-forms

Definition 1.4.1(1) Cotangent vectors

- A *cotangent vector* $\widehat{Y}^*_\alpha$ at a point α in $\mathbb{E}^n$ is a linear mapping of the tangent space $\widehat{T}_\alpha(\mathbb{E}^n)$ at α to $\mathbb{R}$.
- The set of all cotangent vectors at α endowed with the usual vector space structure of a dual space is called the *cotangent space* at α and is denoted by $\widehat{T}^*_\alpha(\mathbb{E}^n)$.

Comments 1.4.1(1) Representation of cotangent vectors

C1 We often denote cotangent vectors at α by a capital letter with an asterisk, e.g., by $\widehat{W}^*_\alpha$, $\widehat{Z}^*_\alpha$ to highlight the fact that cotangent vectors are linear functionals on the tangent space and so on. A cotangent vector $\widehat{W}^*_\alpha$ acts on a tangent vector $\widehat{X}_\alpha$ to produce a number $\widehat{W}^*_\alpha(\widehat{X}_\alpha)$.

C2 Similar to Eq. (1.72) the vector space structure of a cotangent space $\widehat{T}^*_\alpha(\mathbb{E}^n)$ is defined by

$$\left(\widehat{W}^*_\alpha + \widehat{Z}^*_\alpha\right)(\widehat{X}_\alpha) = \widehat{W}^*_\alpha(\widehat{X}_\alpha) + \widehat{Z}^*_\alpha(\widehat{X}_\alpha), \tag{1.143}$$

$$\left(a\,\widehat{W}^*_\alpha\right)(\widehat{X}_\alpha) = a\left(\widehat{W}^*_\alpha(\widehat{X}_\alpha)\right). \tag{1.144}$$

C3 A tangent vector $\widehat{X}_\alpha$ is a differential operator, i.e.,

$$\widehat{X}_\alpha = \sum_{j=1}^n X^j_\alpha\,(\partial/\partial x^j)_\alpha, \tag{1.145}$$

which acts on functions $f \in C^\infty(\alpha)$ to yield a value

$$\widehat{X}_\alpha(f) = \sum_{j=1}^n X^j_\alpha\left(\frac{\partial f}{\partial x^j}\right)_\alpha. \tag{1.146}$$

To appreciate the significance of this in our present context let us make the following observations:

1. We can define a contravariant coordinate displacement vector at a point $\alpha \in \mathbb{E}^n$ by a set of numerical components $(\Delta x)^1_\alpha, (\Delta x)^2_\alpha, \ldots, (\Delta x)^n_\alpha$ representing a displacement.[22] Let us denote such a displacement vector by $(\Delta x)^j_\alpha$.

[22] See §1.2.3C(1) C3.

2. In calculus the differential $(df)_\alpha$ of a function f at a point α along a displacement vector $(\Delta x)^j_\alpha$ is given by

$$(df)_\alpha = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x^j}\right)_\alpha (\Delta x)^j_\alpha. \tag{1.147}$$

   We can see that while at each point α the derivative $(\partial f/\partial x^j)_\alpha$ of f has a definite value determined by the function, the numerical value of the differential $(df)_\alpha$ is not determined by f alone. The displacement vector $(\Delta x)^j_\alpha$ is also involved, i.e., the differential $(df)_\alpha$ takes different values for different displacement vectors.

3. We know from §1.3.1C(3) C5 that X^j_α form a contravariant vector at α. If we associate and equate X^j_α with a displacement vector $(\Delta x)^j_\alpha$ we can, by comparing Eqs. (1.146) and (1.147), identify $\widehat{X}_\alpha(f)$ with the differential $(df)_\alpha$ of f along the direction specified by the contravariant displacement vector $X^j_\alpha = (\Delta x)^j_\alpha$. Since the value of the differential $(df)_\alpha$ is generally different for different displacement vectors X^j_α we can generate a mapping of the space V_α of contravariant displacement vectors to the reals by $X^j_\alpha \mapsto \widehat{X}_\alpha(f)$. We call this mapping the *differential* of f at α and denote it simply by $(df)_\alpha$, i.e., we have, for a given function f,

$$(df)_\alpha : V_\alpha \mapsto I\!R \quad \text{by} \quad X^j_\alpha \mapsto \widehat{X}_\alpha(f). \tag{1.148}$$

4. Equations (1.146), (1.147) and (1.148) are linear in f.

5. Since V_α is identifiable with the tangent space $\widehat{T}_\alpha(I\!E^n)$ we finally arrive at a mapping of $\widehat{T}_\alpha(I\!E^n)$ to $I\!R$ defined by Eq. (1.148). We also call this map the the *differential* of f at α. A formal definition of this is given below.

Definition 1.4.1(2) Differentials of functions on $I\!E^n$

- The *differential* $(df)_\alpha$ of a function $f \in C^\infty(\alpha)$ at a point α is a mapping of the tangent space $\widehat{T}_\alpha(I\!E^n)$ at α to the reals

$$(df)_\alpha : \widehat{T}_\alpha(I\!E^n) \mapsto I\!R \tag{1.149}$$

  defined by

$$\widehat{X}_\alpha \mapsto (df)_\alpha(\widehat{X}_\alpha) = \widehat{X}_\alpha(f) \in I\!R. \tag{1.150}$$

Comments 1.4.1(2) Differentials, cotangent and covariant vectors

C1 It should be emphasized that despite its familiar expression the differential of a function at a point α is not a number or a function. It must be considered as a mapping of $\widehat{T}_\alpha(\mathbb{E}^n)$ to $\mathbb{R}$. In other words it is a **cotangent vector**. In particular we have to distinguish a contravariant coordinate displacement vector $(\Delta x)^j_\alpha$ and the differential $(dx^j)_\alpha$ of a coordinate variable x^j regarded as a function in its own right on $\mathbb{E}^n$. The former consists of a set of numerical components while the latter is a cotangent vector mapping the tangent space $\widehat{T}_\alpha(\mathbb{E}^n)$ to the reals by

$$(dx^j)_\alpha : \widehat{T}_\alpha(\mathbb{E}^n) \mapsto \mathbb{R} \quad \text{by} \quad \widehat{X}_\alpha \mapsto (dx^j)_\alpha(\widehat{X}_\alpha) = \widehat{X}_\alpha(x^j) = X^j_\alpha. \tag{1.151}$$

As the dual space to $\widehat{T}_\alpha(\mathbb{E}^n)$ the cotangent space is also n-dimensional. Since there are n independent coordinate differentials $(dx^j)_\alpha$, $j = 1, 2, \ldots, n$, we can conclude that $(dx^j)_\alpha$ span the cotangent space $\widehat{T}^*_\alpha(\mathbb{E}^n)$. In other words a cotangent vector is generally of the form

$$\widehat{Y}^*_\alpha = \sum_{j=1}^n Y_{\alpha,j}\,(dx^j)_\alpha, \quad Y_{\alpha,j} \in \mathbb{R}. \tag{1.152}$$

As an immediate application we can write

$$(df)_\alpha \quad = \quad \sum_{j=1}^n \left(\frac{\partial f}{\partial x^j}\right)_\alpha (dx^j)_\alpha. \tag{1.153}$$

This is similar to a tangent space being spanned by derivations $(\partial/\partial x^j)_\alpha$.

C2 Let us examine how the differential $(df)_\alpha$ of a function f is related to a covariant vector. To do this we must examine how the expression for a differential changes under a coordinate transformation. On a transformation to a new coordinate system x'^j the expression for $(df)_\alpha$ in Eq. (1.153) can be written as

$$(df)_\alpha \quad = \quad \sum_{j=1}^n \left(\frac{\partial f}{\partial x^j}\right)_\alpha \left(\sum_{k=1}^n \left(\frac{\partial x^j}{\partial x'^k}\right)_\alpha (dx'^k)_\alpha\right). \tag{1.154}$$

It follows that

$$(df)_\alpha \quad = \quad \sum_{k=1}^n \left(\frac{\partial f}{\partial x'^k}\right)_\alpha (dx'^k)_\alpha, \tag{1.155}$$

where

$$\left(\frac{\partial f}{\partial x'^k}\right)_\alpha = \sum_{j=1}^{n} \left(\frac{\partial x^j}{\partial x'^k}\right)_\alpha \left(\frac{\partial f}{\partial x^j}\right)_\alpha . \tag{1.156}$$

We can now rewrite Eq. (1.153) as

$$(df)_\alpha = \sum_{j=1}^{n} Y_{\alpha,j}\,(dx^j)_\alpha \tag{1.157}$$

$$= \sum_{k=1}^{n} Y'_{\alpha,k}\,(dx'^k)_\alpha, \tag{1.158}$$

where

$$Y_{\alpha,j} = \left(\frac{\partial f}{\partial x^j}\right)_\alpha, \quad Y'_{\alpha,k} = \sum_{j=1}^{n} \left(\frac{\partial x^j}{\partial x'^k}\right)_\alpha Y_{\alpha,j}. \tag{1.159}$$

The coefficients $Y_{\alpha,j}$ of the cotangent vector $Y^*_\alpha = (df)_\alpha$ are seen to transform as the components of a covariant vector. Note that a coordinate differential $(dx^i)_\alpha$ corresponds to a covariant vector having components of the form of a Kronecker delta, i.e., $Y_{\alpha,j} = \delta_i^j$. The transformed components in the dashed coordinates are given, according to Eq. (1.159), by

$$Y'_{\alpha,k} = \sum_{i=1}^{n} \left(\frac{\partial x^i}{\partial x'^k}\right)_\alpha \delta_i^j = \left(\frac{\partial x^j}{\partial x'^k}\right)_\alpha . \tag{1.160}$$

So, generally the components of $\widehat{Y}^*_\alpha$ in Eq. (1.152) are seen to transform according to Eq. (1.159). This leads us to conclude that cotangent vectors are the intrinsic definition of covariant vectors. We can now appreciate why the numerical components of a covariant vector transform the way they do.

C3 To sum up, we have the following results:

1. A coordinate differential $(dx^k)_\alpha$ is a cotangent vector.
2. A partial derivative $(\partial/\partial x^j)_\alpha$ is a tangent vector.
3. When using $(dx^k)_\alpha$ to act on tangent vectors $\widehat{X}_\alpha$ we have two cases:

   (a) For $\widehat{X}_\alpha = (\partial/\partial x^j)_\alpha$ we get

$$(dx^k)_\alpha \left(\left(\frac{\partial}{\partial x^j}\right)_\alpha\right) = \left(\frac{\partial x^k}{\partial x^j}\right)_\alpha = \delta_j^k. \tag{1.161}$$

(b) For $\widehat{X}_\alpha = \sum_j X^j_\alpha \, (\partial/\partial x^j)_\alpha$ we get

$$(dx^k)_\alpha(\widehat{X}_\alpha) = X^k_\alpha. \tag{1.162}$$

4. When using a general cotangent vector

$$\widehat{Y}^*_\alpha = \sum_{k=1}^{n} Y_{\alpha,k}(dx^k)_\alpha \tag{1.163}$$

to act on a general tangent vector

$$\widehat{X}_\alpha = \sum_j X^j_\alpha \, (\partial/\partial x^j)_\alpha, \tag{1.164}$$

we get

$$\widehat{Y}^*_\alpha \left(\widehat{X}_\alpha\right) = \sum_{j=1}^{n} Y_{\alpha,j} X^j_\alpha. \tag{1.165}$$

This agrees with the previous contraction process.

C4 We can introduce cotangent vector fields by assigning a cotangent vector at each point in $\mathbb{E}^n$. One way to achieve this is to extend the concept of the differential of a function at a point α to the differential of a function on $\mathbb{E}^n$. The differential of a function f on $\mathbb{E}^n$, denoted by df, can be regarded as a cotangent vector field in that it assigns a cotangent vector $(df)_\alpha$ at each point α in $\mathbb{E}^n$. A cotangent vector field is commonly called a *differential form*.

Definition 1.4.1(3) One-forms

- A *one-form* $\Omega^{(1)}$ on $\mathbb{E}^n$ is a smooth assignment of cotangent vectors on $\mathbb{E}^n$ with a cotangent vector $\widehat{Y}^*_\alpha$ to each point $\alpha \in \mathbb{E}^n$. Here smoothness means that the function g on $\mathbb{E}^n$ defined by $g(\alpha) = \widehat{Y}^*_\alpha(\widehat{X}_\alpha)$ is smooth for any given (smooth) vector field $\widehat{X}$.

Comments 1.4.1(3) Explicit description of one-forms

C1 The cotangent vector assigned to a point α by a one-form is generally denoted by $\Omega^{(1)}_\alpha$.

C2 A simple example of one-forms is simply the differential of a function $f \in C^\infty(\mathbb{E}^n)$ on $\mathbb{E}^n$. We can write down an expression for df as

$$\Omega^{(1)} = df = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\, dx^j. \tag{1.166}$$

This one-form assigns a cotangent vector

$$\Omega^{(1)}_\alpha = (df)_\alpha = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x^j}\right)_\alpha (dx^j)_\alpha \tag{1.167}$$

to every point α. Coordinate differentials dx^j are also examples of one-forms.[23]

C3 One-forms dx^j and $\Omega^{(1)} = df$ act on vector fields $\widehat{X} = \sum_j X^j\, \partial/\partial x^j$ to yield

$$dx^j(\widehat{X}) = X^j, \tag{1.168}$$

$$\Omega^{(1)}(\widehat{X}) = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\, X^j. \tag{1.169}$$

The right-hand side of both of these equations is a smooth function on $\mathbb{E}^n$.

C4 Generally, a one-form is expressible as

$$\Omega^{(1)} = \sum_{j=1}^{n} u_j dx^j, \quad u_j \in C^\infty(\mathbb{E}^n), \tag{1.170}$$

with

$$\Omega^{(1)}(\widehat{X}) = \sum_{j=1}^{n} u_j X^j, \tag{1.171}$$

which is a smooth function on $\mathbb{E}^n$. But $\Omega^{(1)}$ is *not necessarily the differential of a function*, i.e., there may not exist a function f on $\mathbb{E}^n$ such that $u_j = \partial f/\partial x^j$ at every point in $\mathbb{E}^n$. For example, consider a one-form $\Omega^{(1)}$ in $\mathbb{E}^2$ given in rectangular Cartesian coordinates (x, y) by

$$\Omega^{(1)} = y\, dx + x^2\, dy. \tag{1.172}$$

There does not exist a function f such that

$$\frac{\partial f}{\partial x} = y, \quad \frac{\partial f}{\partial y} = x^2 \quad \text{since} \quad \frac{\partial^2 f}{\partial y \partial x} \neq \frac{\partial^2 f}{\partial x \partial y}. \tag{1.173}$$

[23] Loomis and Sternberg (1968) pp. 393-397.

1.4.2 Tensor fields and two-forms

At any point $\alpha \in I\!\!E^n$ we have two vector spaces, the tangent space $\widehat{T}_\alpha(I\!\!E^n)$ and the cotangent space $\widehat{T}^*_\alpha(I\!\!E^n)$. We can form their tensor product spaces

$$\widehat{T}_\alpha(I\!\!E^n) \otimes \widehat{T}_\alpha(I\!\!E^n), \quad \widehat{T}^*_\alpha(I\!\!E^n) \otimes \widehat{T}^*_\alpha(I\!\!E^n), \quad \widehat{T}_\alpha(I\!\!E^n) \otimes \widehat{T}^*_\alpha(I\!\!E^n) \tag{1.174}$$

which lead to the definition of various types of tensors.[24]

Definition 1.4.2(1) Different types of tensors

- Elements of $\widehat{T}_\alpha(I\!\!E^n) \otimes \widehat{T}_\alpha(I\!\!E^n)$ are called contravariant tensors of the second order at the point α.
- Elements of $\widehat{T}^*_\alpha(I\!\!E^n) \otimes \widehat{T}^*_\alpha(I\!\!E^n)$ are called covariant tensors of the second order at the point α.
- Elements of $\widehat{T}_\alpha(I\!\!E^n) \otimes \widehat{T}^*_\alpha(I\!\!E^n)$ are called mixed tensors of the second order at the point α.

Comments 1.4.2(1) Anti-symmetric tensors and wedge products

C1 We can write down the expression for a contravariant tensor at a point α as

$$\widehat{C}^{\otimes}_\alpha = \sum_{j,k=1}^n C^{jk}_\alpha \left(\frac{\partial}{\partial x^j}\right)_\alpha \otimes \left(\frac{\partial}{\partial x^k}\right)_\alpha, \tag{1.175}$$

where C^{jk}_α are the numerical components of the tensor in coordinate system x^j. Similar expressions for covariant and mixed tensors are given respectively by

$$\widehat{C}^{\otimes *}_\alpha = \sum_{j,k=1}^n C_{\alpha,jk} \left(dx^j\right)_\alpha \otimes \left(dx^k\right)_\alpha, \tag{1.176}$$

$$\widehat{M}^{\otimes}_\alpha = \sum_{j,k=1}^n M^j_{\alpha,k} \left(\frac{\partial}{\partial x^j}\right)_\alpha \otimes \left(dx^k\right)_\alpha. \tag{1.177}$$

These definitions can be extended to define tensors of higher orders.

It is easy to verify that upon a coordinate transformation, the numerical components of a tensor transform according to one of the Eqs. (1.34), (1.35) and (1.36) for tensor transformations. Indeed we can say that these transformation equations come from the above intrinsic definitions of tensors.

[24] For an intuitive definition of tensor product of vector spaces see Darling (1994). See also §2.2.3 later for a definition of tensor product of Hilbert spaces.

C2 The relation between cotangent vectors and tangent vectors carries over to contravariant and covariant tensors. A second order covariant tensor $\widehat{C}_\alpha^{\otimes *}$ may be defined as a *bilinear functional* on the Cartesian product $\widehat{T}_\alpha(\mathbb{E}^n) \times \widehat{T}_\alpha(\mathbb{E}^n)$. In other words we have

$$\widehat{C}_\alpha^{\otimes *} : \widehat{T}_\alpha(\mathbb{E}^n) \times \widehat{T}_\alpha(\mathbb{E}^n) \mapsto \mathbb{R} \tag{1.178}$$

by

$$\big(\widehat{X}_\alpha, \widehat{Y}_\alpha\big) \mapsto \widehat{C}_\alpha^{\otimes *}\big(\widehat{X}_\alpha, \widehat{Y}_\alpha\big) \in \mathbb{R} \tag{1.179}$$

with properties:

$$\widehat{C}_\alpha^{\otimes *}\big(a\widehat{X}_\alpha + b\widehat{Y}_\alpha, \widehat{Z}_\alpha\big) = a\,\widehat{C}_\alpha^{\otimes *}\big(\widehat{X}_\alpha, \widehat{Z}_\alpha\big) + b\,\widehat{C}_\alpha^{\otimes *}\big(\widehat{Y}_\alpha, \widehat{Z}_\alpha\big), \tag{1.180}$$

$$\widehat{C}_\alpha^{\otimes *}\big(\widehat{Z}_\alpha, a\widehat{X}_\alpha + b\widehat{Y}_\alpha\big) = a\,\widehat{C}_\alpha^{\otimes *}\big(\widehat{Z}_\alpha, \widehat{X}_\alpha\big) + b\,\widehat{C}_\alpha^{\otimes *}\big(\widehat{Z}_\alpha, \widehat{Y}_\alpha\big). \tag{1.181}$$

C3 For a general covariant tensor $\widehat{C}_\alpha^{\otimes *}$ in Eq. (1.176) we have the following explicit expression:

$$\widehat{C}_\alpha^{\otimes *}\Big(\widehat{X}_\alpha, \widehat{Y}_\alpha\Big) = \sum_{j,k=1}^{n} C_{\alpha,jk}\, X_\alpha^j Y_\alpha^k. \tag{1.182}$$

This reduces to

$$\Big((dx^j)_\alpha \otimes (dx^k)_\alpha\Big)\Big((\partial/\partial x^r)_\alpha, (\partial/\partial x^s)_\alpha \Big) = \delta_r^j\,\delta_s^k, \tag{1.183}$$

$$\Big((dx^j)_\alpha \otimes (dx^k)_\alpha\Big)\Big(\widehat{X}_\alpha, \widehat{Y}_\alpha\Big) = X_\alpha^j Y_\alpha^k \tag{1.184}$$

when $\widehat{C}_\alpha^{\otimes *} = \big(dx^j\big)_\alpha \otimes (dx^k)_\alpha$.

Likewise, we can regard a contravariant tensor as a bilinear functional on the Cartesian product $\widehat{T}_\alpha^*(\mathbb{E}^n) \times \widehat{T}_\alpha^*(\mathbb{E}^n)$ of cotangent vectors.

C4 A second order covariant tensor $\widehat{C}_\alpha^{\otimes *}$ is said to be *anti-symmetric* if

$$\widehat{C}_\alpha^{\otimes *}(\widehat{X}_\alpha, \widehat{Y}_\alpha) = -\,\widehat{C}_\alpha^{\otimes *}(\widehat{Y}_\alpha, \widehat{X}_\alpha). \tag{1.185}$$

This is consistent with the definition in §1.2.4C(1) C2, i.e., for $\widehat{C}_\alpha^{\otimes *}$ to be anti-symmetric it must satisfy the condition $C_{\alpha,jk} = -\,C_{\alpha,kj}$. Symmetric covariant tensors are similarly defined. Many tensors of physical significance, such as the electromagnetic tensor, are anti-symmetric. A simple example is

$$\widehat{C}_\alpha^{\otimes *} = \big(dx^j\big)_\alpha \otimes \big(dx^k\big)_\alpha - \big(dx^k\big)_\alpha \otimes \big(dx^j\big)_\alpha \tag{1.186}$$

with

$$\widehat{C}_\alpha^{\otimes *}\Big(\widehat{X}_\alpha, \widehat{Y}_\alpha\Big) = X_\alpha^j Y_\alpha^k - X_\alpha^k Y_\alpha^j \tag{1.187}$$

clearly showing anti-symmetry. We can build up other anti-symmetric covariant tensors in terms of these simple ones. To simplify the notation we introduce the concept of an *exterior product*, denoted symbolically by a wedge, hence the alternative name *wedge product*:

$$(dx^j)_\alpha \wedge (dx^k)_\alpha = (dx^j)_\alpha \otimes (dx^k)_\alpha - (dx^k)_\alpha \otimes (dx^j)_\alpha. \tag{1.188}$$

We have

$$\Big((dx^j)_\alpha \wedge (dx^k)_\alpha\Big)\Big(\widehat{X}_\alpha, \widehat{Y}_\alpha\Big) = X^j_\alpha Y^k_\alpha - X^k_\alpha Y^j_\alpha. \tag{1.189}$$

In $I\!E^3$ there are only 3 such wedge products at each point:

$$(dx^1)_\alpha \wedge (dx^2)_\alpha, \quad (dx^2)_\alpha \wedge (dx^3)_\alpha, \quad (dx^1)_\alpha \wedge (dx^3)_\alpha \tag{1.190}$$

or equivalently

$$(dx)_\alpha \wedge (dy)_\alpha, \quad (dy)_\alpha \wedge (dz)_\alpha, \quad (dx)_\alpha \wedge (dz)_\alpha. \tag{1.191}$$

These products form the basis spanning the space of second order anti-symmetric covariant tensors at α, i.e., an anti-symmetric covariant tensor $\widehat{C}^{\otimes *}_\alpha$ is expressible as

$$\begin{aligned}\widehat{C}^{\otimes *}_\alpha &= C_{\alpha,12}\,(dx^1)_\alpha \wedge (dx^2)_\alpha \\ &+ C_{\alpha,23}\,(dx^2)_\alpha \wedge (dx^3)_\alpha + C_{\alpha,13}\,(dx^1)_\alpha \wedge (dx^3)_\alpha,\end{aligned} \tag{1.192}$$

or more succinctly

$$\widehat{C}^{\otimes *}_\alpha = \sum_{1 \le j < k}^{3} C_{\alpha,jk}\,(dx^j)_\alpha \wedge (dx^k)_\alpha. \tag{1.193}$$

C5 Let

$$\widehat{X} = x\,\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}, \tag{1.194}$$

$$\widehat{Y} = \frac{\partial}{\partial x} + 2y\,\frac{\partial}{\partial y} + z\,\frac{\partial}{\partial z} \tag{1.195}$$

be two vector fields on $I\!E^3$ expressed in usual Cartesian coordinates (x, y, z). At a point $\alpha = (1, 1, 3)$ we have two tangent vectors

$$\widehat{X}_\alpha = \left(\frac{\partial}{\partial x}\right)_\alpha + \left(\frac{\partial}{\partial y}\right)_\alpha + \left(\frac{\partial}{\partial z}\right)_\alpha, \tag{1.196}$$

$$\widehat{Y}_\alpha = \left(\frac{\partial}{\partial x}\right)_\alpha + 2\left(\frac{\partial}{\partial y}\right)_\alpha + 3\left(\frac{\partial}{\partial z}\right)_\alpha. \tag{1.197}$$

Let

$$\widehat{C}^{\otimes *}_{\alpha} = (dy)_{\alpha} \wedge (dz)_{\alpha}. \tag{1.198}$$

Using Eq. (1.189) we get

$$\widehat{C}^{\otimes *}_{\alpha}\big(\widehat{X}_{\alpha}, \widehat{Y}_{\alpha}\big) = 3 - 2 = 1. \tag{1.199}$$

C6 The wedge product is linear so that

$$(dx^1)_{\alpha} \wedge \Big(a\,(dx^2)_{\alpha} + b\,(dx^3)_{\alpha}\Big) = a\,(dx^1)_{\alpha} \wedge (dx^2)_{\alpha} + b\,(dx^1)_{\alpha} \wedge (dx^3)_{\alpha}. \tag{1.200}$$

Definition 1.4.2(2) Covariant tensor fields of second order

- A covariant tensor field of second order $\widehat{C}^{\otimes *}$ is a smooth assignment of a second order covariant tensor to each point $\alpha \in I\!\!E^n$, smoothness being defined in a comment below.

Comments 1.4.2(2) Smoothness and coordinate expressions

C1 Given two vector fields $\widehat{X}$ and $\widehat{Y}$ any assignment of covariant tensors throughout $I\!\!E^n$, with a tensor $\widehat{C}^{\otimes *}_{\alpha}$ at each point α, gives rise to a function g on $I\!\!E^n$ defined by

$$g(\alpha) = \widehat{C}^{\otimes *}_{\alpha}(\widehat{X}_{\alpha}, \widehat{Y}_{\alpha}). \tag{1.201}$$

The assignment is said to be *smooth* if g is a smooth function for every pair of vector fields $\widehat{X}$ and $\widehat{Y}$. A covariant vector field is expressible as

$$\widehat{C}^{\otimes *} = \sum_{j,k=1}^{n} C_{jk}\, dx^j \otimes dx^k, \tag{1.202}$$

where the components $C_{jk} \in C^{\infty}(I\!\!E^n)$. We can see that the function g defined by

$$g = \widehat{C}^{\otimes *}(\widehat{X}, \widehat{Y}) = \sum_{j,k=1}^{n} C_{jk}\, X^j Y^k \tag{1.203}$$

is smooth.

C2 A covariant tensor field $\widehat{C}^{\otimes *}$ is said to be *symmetric* if

$$\widehat{C}^{\otimes *}\big(\widehat{X}, \widehat{Y}\big) = \widehat{C}^{\otimes *}\big(\widehat{Y}, \widehat{X}\big), \tag{1.204}$$

and *anti-symmetric* if

$$\widehat{C}^{\otimes *}\big(\widehat{X}, \widehat{Y}\big) = -\,\widehat{C}^{\otimes *}\big(\widehat{Y}, \widehat{X}\big). \tag{1.205}$$

Explicitly we have

$$\widehat{C}^{\otimes *} = \sum_{j,k=1}^{n} C_{jk}\, dx^j \otimes dx^k \tag{1.206}$$

which is symmetric if $C_{jk} = C_{kj}$ and is anti-symmetric if $C_{jk} = -C_{kj}$. Anti-symmetry implies $C_{jj} = -C_{jj} = 0$. We can also make use of wedge product to express an anti-symmetric tensor as

$$\Omega^{(2)} = \sum_{1 \le j < k}^{n} \omega_{jk}\, dx^j \wedge dx^k, \quad \omega_{jk} \in C^{\infty}(\mathbb{E}^n). \tag{1.207}$$

In $\mathbb{E}^3$ in the usual Cartesian coordinates (x, y, z) an example is

$$\Omega^{(2)} = x^2\, dx \wedge dy. \tag{1.208}$$

Acting on the pair of vector fields $\widehat{X} = \partial/\partial x$ and $\widehat{Y} = y^2\, \partial/\partial y$ gives

$$\Omega^{(2)}\big(\widehat{X}, \widehat{Y}\big) = x^2 y^2. \tag{1.209}$$

Anti-symmetric tensor fields of second order are of such fundamental importance in physics and mathematics that a special name is to be given to them. We shall also use a separate notation, i.e., $\Omega^{(2)}$, to denote a *second order anti-symmetric covariant tensor.*

Definition 1.4.2(3) Two-forms

- A *two-form* $\Omega^{(2)}$ is an anti-symmetric covariant tensor field of second order.

1.4.3 Exterior differentiation

A differential operation on a function gives rise to a one-form, and a two-form is generated by the wedge product of two one-forms, and so on. Let us call a function a *zero-form*, then we can build up one-forms and two-forms from zero-forms. For uniformity of notation let $\Omega^{(0)}$ denote a zero-form, i.e., $\Omega^{(0)}$ is just a function f. We can set up a formal procedure for generating higher forms in terms of an *exterior differential operator*. An exterior differential operator $\widehat{d}_e$ is a mapping which maps a p-form $\Omega^{(p)}$ into a $p+1$ form $\Omega^{(p+1)}$ satisfying the following properties:[25]

[25] Frankel (1997) p. 73.

1. Given a zero-form $\Omega^{(0)} = f$ the one-form $\Omega^{(1)} = \widehat{d}_e f$ is given by

$$\Omega^{(1)} = \widehat{d}_e \Omega^{(0)} = df = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\, dx^j. \tag{1.210}$$

   In particular we have, treating a coordinate variable x^j as a function,

$$\widehat{d}_e x^j = dx^j. \tag{1.211}$$

   In other words the exterior differential is the same as the differential when acting on functions. This is why we employed the term exterior differential rather than the more commonly used term of *exterior derivative*. It is when applied to a p-form that the exterior differential operator really comes into its own.

2. Given any p-form we have $\widehat{d}_e(\widehat{d}_e\Omega^{(p)}) = 0$. It follows that $\widehat{d}_e(df) = \widehat{d}_e(\widehat{d}_e\Omega^{(0)}) = 0$. In particular we have $\widehat{d}_e(dx^j) = 0$. This does not mean that the exterior differentials of all one-forms vanish since not every one-form is the differential of a zero form, as pointed out in §1.4.1C(3) C4.

3. The exterior differential operator $\widehat{d}_e$ is linear, i.e., given two p-forms $\Omega_1^{(p)}$ and $\Omega_2^{(p)}$ we have

$$\widehat{d}_e(\Omega^{(p)} + \Omega^{(p)}) = \widehat{d}_e\Omega_1^{(p)} + \widehat{d}_e\Omega_2^{(p)}. \tag{1.212}$$

4. Given the wedge product of a p-form $\Omega^{(p)}$ and a q-form $\Omega^{(q)}$ we have

$$\widehat{d}_e\left(\Omega^{(p)} \wedge \Omega^{(q)}\right) = \left(\widehat{d}_e\Omega^{(p)}\right) \wedge \Omega^{(q)} + (-1)^p\, \Omega^{(p)} \wedge \left(\widehat{d}_e\Omega^{(q)}\right), \tag{1.213}$$

   with the understanding that, when $\Omega^{(p)}$ is a zero-form $\Omega^{(0)} = f$, the expression $\Omega^{(p)} \wedge \Omega^{(q)}$ is taken to be $f\,\Omega^{(q)}$ so that

$$\widehat{d}_e(f\,\Omega^{(q)}) = df \wedge \Omega^{(q)} + f\left(\widehat{d}_e\Omega^{(q)}\right). \tag{1.214}$$

Examples 1.4.3(1) Exterior differentiation and forms

E1 The above prescription is sufficient to define the operator $\widehat{d}_e$. To see how it works in practice let us consider the exterior differential of a general one-form

$$\Omega^{(1)} = \sum_{k=1}^{n} u_k\, dx^k, \quad u_k \in C^\infty(\mathbb{E}^n). \tag{1.215}$$

We can work out $\widehat{d}_e\Omega^{(1)}$ as follows:

$$\begin{aligned}\Omega^{(2)} &= \widehat{d}_e\Omega^{(1)} = \sum_{k=1}^{n} \widehat{d}_e(u_k dx^k) \\ &= \sum_{k=1}^{n} \Big((\widehat{d}_e u_k) \wedge dx^k + u_k\, \widehat{d}_e(dx^k) \Big) \\ &= \sum_{j,k}^{n} \frac{\partial u_k}{\partial x^j}\, dx^j \wedge dx^k, \end{aligned} \tag{1.216}$$

since

$$\widehat{d}_e u_k = \sum_{1=j}^{n} \frac{\partial u_k}{\partial x^j}\, dx^j \quad \text{and} \quad \widehat{d}_e(dx^k) = 0. \tag{1.217}$$

E2 The highest form we can get in $I\!\!E^n$ is an n-form. For example, in $I\!\!E^3$ we can have a 3-form

$$\Omega^{(3)} = f(x, y, z)\, dx \wedge dy \wedge dz. \tag{1.218}$$

It is obvious that any attempt to create a higher form, for example by applying the operator $\widehat{d}_e$ to $\Omega^{(3)}$, would annihilate $\Omega^{(3)}$ and hence would fail to produce a 4-form.

We have discussed several ways of increasing the order of a tensor field, in particular by using the exterior product and exterior differentiation. On the other hand, we can reduce the order of tensors by contraction. Another important method is through a process known as an *interior product* which we shall discuss in the next section.

1.4.4 Interior products, closed and exact forms

An *interior product* is a product between a vector field and a p-form resulting in a $(p-1)$-form. The generally adopted notation is

$$\widehat{X} \lrcorner\, \Omega^{(p)} = \Omega^{(p-1)}. \tag{1.219}$$

However, we shall not go into a general definition or a general prescription for arbitrary forms. It is sufficient for our purposes to see how this interior product works for one-forms and two-forms. For these cases the interior product satisfies the following properties:

1. $\widehat{X} \lrcorner\, \Omega^{(0)} = 0.$

2. $\widehat{X} \lrcorner \Omega^{(1)} = \Omega^{(0)} = \Omega^{(1)}(\widehat{X})$.

3. $\widehat{X} \lrcorner \Omega^{(2)} = \Omega^{(1)}$, where $\Omega^{(1)}$ is determined by

$$\Omega^{(1)}(\widehat{Y}) = \Omega^{(2)}(\widehat{X}, \widehat{Y}) \quad \text{for all} \quad \widehat{Y}. \tag{1.220}$$

4. The interior product obeys the following rule:

$$\widehat{X} \lrcorner (\Omega^{(p)} \wedge \Omega^{(q)}) = (\widehat{X} \lrcorner \Omega^{(p)}) \wedge \Omega^{(q)} + (-1)^p \, \Omega^{(p)} \wedge (\widehat{X} \lrcorner \Omega^{(q)}). \tag{1.221}$$

Note that:

(a) If $\Omega^{(p)}$ is a 0-form, i.e., a function f, then

$$\widehat{X} \lrcorner \left(f \, \Omega^{(q)} \right) = f \left(\widehat{X} \lrcorner \Omega^{(q)} \right), \tag{1.222}$$

since $(\widehat{X} \lrcorner \Omega^{(p)}) = 0$ and $1^{-p} = 1$ when $p = 0$. In particular we have

$$\begin{aligned} & \widehat{X} \lrcorner \left(\sum_{1=j<k}^{n} \omega_{jk} \, dx^j \wedge dx^k \right) \\ = & \sum_{1=j<k}^{n} \omega_{jk} \left(\widehat{X} \lrcorner \, (dx^j \wedge dx^k) \right) \qquad (1.223) \\ = & \sum_{1=j<k}^{n} \omega_{jk} \, (X^j dx^k - X^k dx^j). \qquad (1.224) \end{aligned}$$

(b) If $\Omega^{(p)} = \Omega^{(1)}$ is a one-form then $(\widehat{X} \lrcorner \Omega^{(1)})$ is a zero-form, i.e., $(\widehat{X} \lrcorner \Omega^{(1)}) = f$ which is a function. We should take $(\widehat{X} \lrcorner \Omega^{(1)}) \wedge \Omega^{(q)}$ to be $f \, \Omega^{(q)}$.

Examples 1.4.4(1) Interior products

E1 For a zero-form $\Omega^{(0)} = f$ we have $\widehat{X} \lrcorner f = 0$.

E2 For a one-form $\Omega^{(1)} = \sum_j u_j dx^j$ we have

$$\widehat{X} \lrcorner \left(\sum_{j=1}^{n} u_j \, dx^j \right) = \left(\sum_{j=1}^{n} u_j \, dx^j \right) (\widehat{X}) = \sum_{j=1}^{n} u_j \, X^j. \tag{1.225}$$

Here we have used a result from §1.4.1C(3) C4. In particular we have

$$\widehat{X} \lrcorner dx^j = dx^j(\widehat{X}) = X^j. \tag{1.226}$$

An interior product is similar to the contraction operation between a covariant vector field and a contravariant vector field discussed in §1.2.4C(1) C7.

E3 To appreciate property 3 embodied in Eq. (1.220) let us consider a two-form

$$\Omega^{(2)} = \sum_{1 \leq j < k} \omega_{jk} \, dx^j \wedge dx^k. \tag{1.227}$$

Then

$$\Omega^{(1)} = \widehat{X} \lrcorner \, \Omega^{(2)} = \widehat{X} \lrcorner \left(\sum_{1=j<k}^{n} \omega_{jk} \, dx^j \wedge dx^k \right) \tag{1.228}$$

is a one-form given by Eq. (1.224). When this one-form $\Omega^{(1)}$ acts on an arbitrary vector field $\widehat{Y}$ we get

$$\Omega^{(1)}\big(\widehat{Y}\big) = \sum_{1=j<k}^{n} \omega_{jk} \left(X^j dx^k - X^k dx^j \right)\big(\widehat{Y}\big) \tag{1.229}$$

$$= \sum_{1=j<k}^{n} \omega_{jk} \left(X^j Y^k - X^k Y^j \right). \tag{1.230}$$

Using Eq. (1.189) we also find that

$$\Omega^{(2)}\big(\widehat{X}, \widehat{Y}\big) = \sum_{1=j<k}^{n} \omega_{jk} \, (X^j Y^k - X^k Y^j). \tag{1.231}$$

It follows that

$$\Omega^{(1)}\big(\widehat{Y}\big) = \Omega^{(2)}\big(\widehat{X}, \widehat{Y}\big) \tag{1.232}$$

as stated in Eq. (1.220). This enables us to relate a one-form $\Omega^{(1)}$ to a vector field $\widehat{X}$ through a chosen two-form $\Omega^{(2)}$. We shall return to make use of this later when we formulate Hamiltonian mechanics.

Definition 1.4.4(1) Closed, exact and non-degenerate two-forms

- A p-form $\Omega^{(p)}$ is said to be *closed* if $\widehat{d}_e \Omega^{(p)} = 0$.
- $\Omega^{(p)}$ is said to be *exact* if there exists a $(p-1)$-form $\Omega^{(p-1)}$ such that $\Omega^{(p)} = \widehat{d}_e \Omega^{(p-1)}$.
- A two-form $\Omega^{(2)}$ is said to be *non-degenerate* if $\Omega^{(2)}(\widehat{X}, \widehat{Y}) = 0$ for all $\widehat{Y}$ implies $\widehat{X} = 0$.

Examples 1.4.4(2) Degenerate and non-degenerate forms

E1 An exact form is obtained from a lower form by exterior differentiation. Generally a closed form is not necessarily exact, but an exact form is also closed, e.g., $\Omega^{(1)} = df$ is exact and closed. The one-form in Eq. (1.172) is neither closed nor exact.

E2 $\Omega^{(2)} = dx^1 \wedge dx^2$ in $\mathbb{E}^2$ is closed and non-degenerate. In $\mathbb{E}^3$ the two form $\Omega^{(2)} = dx^1 \wedge dx^2$ is closed but degenerate since

$$\Omega^{(2)}\big(\widehat{X}, \widehat{Y}\big) = X^1 Y^2 - Y^1 X^2 = 0 \quad \text{for all } \widehat{Y} \tag{1.233}$$

implies

$$X^1 = X^2 = 0 \quad \text{but leaving } X^3 \text{ unrestricted, i.e., } \widehat{X} \neq 0. \tag{1.234}$$

1.5 Differentiable Manifolds

1.5.1 Definition and examples

Although Euclidean spaces are most useful in physical applications there are many occasions when we have to deal with non-Euclidean spaces either in principle or for practical convenience. The simplest examples are the circle and the sphere in $\mathbb{E}^3$. These spaces, known as differentiable manifolds, are non-Euclidean because they are not topologically equivalent to $\mathbb{E}^n$.

Basically a differentiable manifold of dimension ℓ is a Hausdorff space which is locally topologically equivalent to $\mathbb{R}^\ell$, i.e., it locally resembles an open cube of $\mathbb{R}^\ell$. This intuitive notion helps us to appreciate the concept of differentiable manifolds. The set of points forming a circle in $\mathbb{R}^2$ is a manifold since any small section of the circle resembles an interval of $\mathbb{R}$; a circle is a one-dimensional manifold. Note that the resemblance is local only. The circle as a whole (globally) certainly does not look anything like $\mathbb{R}$. A sphere is not topologically equivalent to $\mathbb{R}^2$. However it is locally like an open rectangle of $\mathbb{R}^2$, making it a two-dimensional manifold. An electrical circuit containing branches joined up at some points, referred to as *branch points*, does not have the geometry of a manifold; branch points with several wires coming in and out do not resemble any open set of $\mathbb{R}^n$.

Definition 1.5.1(1) Differentiable manifolds

- An ℓ-dimensional differentiable manifold $\mathbb{M}^\ell$ is a Hausdorff space with the following properties:

1. Every point $m \in I\!M^{\ell}$ has a neighbourhood $\mathcal{N}_m$ which is topologically equivalent to an open cube Λ of $I\!R^{\ell}$. The homeomorphism F_m of $\mathcal{N}_m$ onto Λ which maps $m \in \mathcal{N}_m$ to $\alpha \in \Lambda$ defines a local coordinate chart on $\mathcal{N}_m$ by assigning the coordinates x^j of α to m.
2. Let $m' \in I\!M$ be another point which has a neighbourhood $\mathcal{N}'_{m'}$ topologically equivalent to an open cube Λ' of $I\!R^{\ell}$. Let x'^j be the local coordinates assigned to points in $\mathcal{N}'_{m'}$. Then on the overlapped region $\mathcal{N}_m \cap \mathcal{N}'_{m'}$ the coordinates x^j and x'^j are smooth functions of one another.

Examples 1.5.1(1) Spheres and circles

E1 The Euclidean space $I\!E^n$ is clearly a differentiable manifold. In fact the space $I\!R^n$ itself is a differentiable manifold and so is any open cube of $I\!R^n$. For brevity we often refer to a differentiable manifold simply as a manifold. According to the definition above we can have a single coordinate system covering the entire manifold only if the manifold is topologically equivalent to an open cube of $I\!R^{\ell}$. A coordinate system covering the entire manifold is referred to as a *global coordinate system*.

E2 In the usual rectangular Cartesian coordinates (x, y) in $I\!E^2$ a circle $\mathcal{C}$ of radius a and centered at the origin is specified by the set of points $x^2 + y^2 = a^2$. With the topology induced from that of $I\!E^2$ defined in §1.2.1C(1) C6 such a circle is a one-dimensional differentiable manifold. In the usual polar coordinates (r, θ) we have[26]

$$\mathcal{C} = \{(r, \theta) : r = a,\ \theta \in [0, 2\pi]\}, \tag{1.235}$$

where $\theta = 0$ and $\theta = 2\pi$ refer to the same point. Although the parameter θ does cover the entire circle it is not a global coordinate in $\mathcal{C}$. By Definition 1.5.1(1) a coordinate variable must be single-valued and takes values in an open cube of $I\!E^n$. In the case of the circle $\mathcal{C}$ a coordinate must take values in an open interval of $I\!E$. However, it is not possible to find a coordinate variable which takes values in an open interval of $I\!E$ to cover the entire circle. In other words $\mathcal{C}$ is not topologically equivalent to an open interval of $I\!E$. A coordinatization of $\mathcal{C}$ requires two overlapping local coordinate charts. Let

1. $\mathcal{C}_c = \{(r, \theta) : r = a,\ \theta \in (0, 2\pi)\}$, i.e., $\mathcal{C}_c$ is the circle with the point $\theta = 0$ cut away.

[26]We follow a standard practice to denote the polar angle in $I\!E^2$ by θ although the angle variable θ here is quite different from the angle variable also denoted by θ in spherical coordinates in $I\!E^3$.

2. $\mathcal{C}_\pi = \{(r,\theta) : r = a,\ \theta \in [0,\pi) \cup (\pi, 2\pi]\}$, i.e., $\mathcal{C}_\pi$ is the circle with the point $\theta = \pi$ removed.

We can define global coordinates in $\mathcal{C}_c$ and $\mathcal{C}_\pi$ as follows:

1. Introduce a new variable ϑ defined on $\mathcal{C}_c$ by

$$\vartheta = \theta \qquad \text{for} \quad \theta \in (0, 2\pi). \tag{1.236}$$

   Then ϑ takes values in the open interval $(0, 2\pi)$ and hence constitutes a coordinate covering $\mathcal{C}_c$.

2. Introduce a new variable ϑ_π defined on $\mathcal{C}_\pi$ by

$$\vartheta_\pi = \begin{cases} \theta & \text{if } \theta \in (0,\pi) \\ \theta - 2\pi & \text{if } \theta \in (\pi, 2\pi] \end{cases} . \tag{1.237}$$

   This new variable ϑ_π, which takes values in the open interval $(-\pi, \pi)$ and agrees with θ over the range $(0,\pi)$, forms a coordinate covering $\mathcal{C}_\pi$.

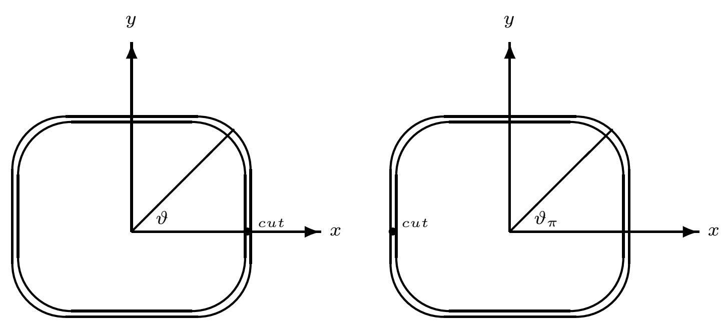

Fig. 1.5.1(1) Circle $\mathcal{C}_c$ with point $\theta = 0$ removed (left) and circle $\mathcal{C}_\pi$ with point $\theta = \pi$ removed (right).

We have the following results:

1. $\mathcal{C}_c \cup \mathcal{C}_\pi = \mathcal{C}$.

2. $\mathcal{C}_c \cap \mathcal{C}_\pi$ consists of two disjoint regions $\mathcal{C}_+$ and $\mathcal{C}_-$:

   (a) $\mathcal{C}_+$ is the overlapping region specified by $\vartheta \in (0,\pi)$.

   (b) $\mathcal{C}_-$ is the region specified by $\vartheta_\pi \in (-\pi, 0)$.

3. In $\mathcal{C}_+$ we have $\vartheta = \vartheta_\pi$, and in $\mathcal{C}_-$ we have $\theta_\pi = \vartheta - 2\pi$.
4. The two coordinates ϑ and ϑ_π are smoothly related in the overlapping regions and together they cover the entire circle $\mathcal{C}$.

E3 A sphere $\mathcal{S}^2$ of radius a in $I\!E^3$ is a two-dimensional differentiable manifold. Similar to the circle, a sphere is not coverable by a single coordinate chart. This is related to the fact that $\mathcal{S}^2$ is not topologically equivalent to any open cube of $I\!R^2$.

E4 It is possible to introduce non-global coordinates in a Euclidean space. The usual spherical coordinates (r, θ, φ) in $I\!E^3$ are examples. These spherical coordinates do not form a global coordinate system in $I\!E^3$. In addition to the angle variables not being global they are also undefined at the origin $r = 0$.

E5 Vector fields can be established in a differentiable manifold $I\!M^\ell$, despite the possible absence of global coordinates. Clearly we can do this since vector fields are definable independent of coordinates. In fact the definitions of vector fields, tensor fields and differential forms introduced in $I\!E^\ell$ carry over into $I\!M^\ell$. There is just one complication: we have to match quantities expressed in terms of coordinates defined on overlapping local coordinate charts. Taking again the example of the circle $\mathcal{C}$ above, we can first define a vector field in region $\mathcal{C}_c$ and one in $\mathcal{C}_\pi$, e.g.,

$$\widehat{X}_c = \frac{d}{d\vartheta} \quad \text{on} \quad \mathcal{C}_c \tag{1.238}$$

$$\widehat{X}_\pi = \frac{d}{d\vartheta_\pi} \quad \text{on} \quad \mathcal{C}_\pi. \tag{1.239}$$

In the intersection $\mathcal{C}_c \cap \mathcal{C}_\pi$ we have $\widehat{X}_c = \widehat{X}_\pi$, a result derived from the relationship between ϑ and ϑ_π spelled out earlier. It follows that we have a vector field $\widehat{X}$ defined on the entire circle $\mathcal{C}$ with

$$\widehat{X} = \begin{cases} \widehat{X}_c & \text{on } \mathcal{C}_c \\ \widehat{X}_\pi & \text{on } \mathcal{C}_\pi \end{cases}. \tag{1.240}$$

We shall have many occasions to return to the circle and to this vector field on $\mathcal{C}$. It is clearly very tedious to have to explicitly spell out all these details every time. For notational convenience we shall formally regard $\theta \in [0, 2\pi]$ as a coordinate covering $\mathcal{C}$ and denote the above vector field symbolically by

$$\widehat{X} = \frac{d}{d\theta}. \tag{1.241}$$

To see if this vector field is complete or not we need to examine its integral curves. The integral curve of $\widehat{X}$ starting from, say, the point $\theta = 0$ in $\mathcal{C}$ is given by

$$\theta(\tau) = \tau \quad \text{for} \quad \tau \in [0, 2\pi]. \tag{1.242}$$

We can extend the domain of the curve to the entire real line by writing down, for example,

$$\theta(\tau) = \tau - 2\pi \quad \text{for} \quad \tau \in [2\pi, 4\pi], \tag{1.243}$$

and so on. Intuitively we can see that this corresponds to translations along, and round and round, the circumference of $\mathcal{C}$. These translations form a one-parameter group of transformations of $\mathcal{C}$. The same applies to integral curves starting from all other points in $\mathcal{C}$. This vector field is therefore complete.[27] We can also consider the translations as rotations of $\mathcal{C}$ which are known to form a group.

1.5.2 Riemannian manifolds

Definition 1.5.2(1) Pseudo-Riemannian and Riemannian metric

- A *pseudo-Riemannian metric* on a manifold $I\!M^\ell$ is a symmetric and non-degenerate covariant tensor field $\widehat{G}^{\otimes *}$ of second order. A manifold with a pseudo-Riemannian metric is called a *pseudo-Riemannian manifold*.

- A pseudo-Riemannian metric $\widehat{G}^{\otimes *}$ on $I\!M^\ell$ is said to be *positive definite* if for all vector fields $\widehat{X}$ on $I\!M^\ell$ we have:

  1. $\widehat{G}^{\otimes *}_m(\widehat{X}_m, \widehat{X}_m) \geq 0 \quad \forall\, m \in \mathcal{M}^\ell$.
  2. $\widehat{G}^{\otimes *}_m(\widehat{X}_m, \widehat{X}_m) = 0 \quad$ if and only if $\widehat{X}_m = 0 \quad \forall\, m \in \mathcal{M}^\ell$.

- A positive definite pseudo-Riemannian metric on a manifold is called a *Riemannian metric* and a manifold with a Riemannian metric is called a *Riemannian manifold*.

Comments 1.5.2(1) Volume elements and integration

C1 A metric is an important geometric quantity. As a second order covariant tensor field it is able to act on vector fields, in accordance with §1.4.2C(2) C1, to give rise to smooth functions on the manifold. This opens the way to the introduction of a host of other useful quantities some of which will now be discussed.

[27] See also §1.3.2E(1) E4.

C2 Consider the Euclidean space $\mathbb{E}^n$ with rectangular Cartesian coordinates x^j. Introduce a covariant tensor by

$$\widehat{G}^{\otimes *} = \sum_{i,j=1}^{n} \delta_{ij}\, dx^i \otimes dx^j. \tag{1.244}$$

Then, we have

$$\widehat{G}^{\otimes *}\big(\widehat{X}, \widehat{Y}\big) = \sum_{i,j=1}^{n} \delta_{ij}\, X^i Y^j = \sum_{j=1}^{n} X^j Y^j, \tag{1.245}$$

showing that $\widehat{G}^{\otimes *}$ is symmetric, non-degenerate and positive definite. So, $\widehat{G}^{\otimes *}$ is a Riemannian metric on $\mathbb{E}^n$, making $\mathbb{E}^n$ a Riemannian manifold. On a transformation to a new coordinate chart x'^r, possibly defined only for an open subset $\mathcal{D}$ in $\mathbb{E}^n$, we have

$$dx^i = \sum_{r=1}^{n} \frac{\partial x^i}{\partial x'^r}\, dx'^r. \tag{1.246}$$

In the new coordinate chart the metric tensor on $\mathcal{D}$ will be of the form

$$\widehat{G}^{\otimes *} = \sum_{r,s=1}^{n} g'_{rs}\, dx'^r \otimes dx'^s, \tag{1.247}$$

$$g'_{rs} = \sum_{i,j=1}^{n} \frac{\partial x^i}{\partial x'^r} \frac{\partial x^j}{\partial x'^s}\, \delta_{ij} \tag{1.248}$$

with

$$\widehat{G}^{\otimes *}(\widehat{X}, \widehat{Y}) = \sum_{r,s=1}^{n} g'_{rs}\, X'^r Y'^s, \tag{1.249}$$

where X'^r and Y'^s are respectively the components of $\widehat{X}$ and $\widehat{Y}$ in the new coordinates.

Generally a Riemannian manifold $\mathbb{M}^\ell$ admits only local coordinate charts x'^j which may be non-rectangular, the sphere $\mathbb{M}^2 = \mathcal{S}^2$ in $\mathbb{E}^3$ being an example. In each coordinate chart the metric tensor is still of the form

$$\widehat{G}^{\otimes *} = \sum_{i,j=1}^{\ell} g'_{ij}\, dx'^i \otimes dx'^j. \tag{1.250}$$

On a coordinate transformation to $\bar{x}^j$ the expression changes to

$$\widehat{G}^{\otimes *} = \sum_{i,j=1}^{\ell} \bar{g}_{ij}\, d\bar{x}^i \otimes d\bar{x}^j, \tag{1.251}$$

with

$$\bar{g}_{ij} \;=\; \sum_{m,k=1}^{\ell} \frac{\partial x'^m}{\partial \bar{x}^i} \frac{\partial'^k}{\partial \bar{x}^j}\, g'_{mk}. \qquad (1.252)$$

Comparing with Eq. (1.35) we can see that generally the coefficients g'_{ij} of the metric $\widehat{G}^{\otimes *}$ transform like the components of a covariant tensor. Hence g'_{ij} are often referred to as the components of the metric tensor in coordinate chart x'^j.

C3 A Riemannian metric generates a scalar product on each tangent space. In $I\!\!E^n$ the scalar product between two tangent vectors $\widehat{X}_\alpha$ and $\widehat{Y}_\alpha$ in $\widehat{T}_\alpha I\!\!E^n$ is given by $\widehat{G}^{\otimes *}(\widehat{X}, \widehat{Y})$ evaluated at α, i.e., by evaluating Eq. (1.245) at α. This agrees with our previous discussion in §1.2.1C(1) C4.

C4 A Riemannian manifold admits a *volume element* which enables us to set up integrations on $I\!\!M^\ell$. Let

$$\widehat{G}^{\otimes *} = \sum_{i,j=1}^{\ell} g'_{ij}\, dx'^i \otimes dx'^j \qquad (1.253)$$

be the metric tensor in $I\!\!M^\ell$. Then in coordinate system x'^j the volume $\delta\mu'$ of an infinitesimal rectangle $(x'^j, x'^j + \Delta x'^j)$ is defined to be

$$\delta\mu' = \sqrt{g'}\, \Delta x'^1 \Delta x'^2 \cdots \Delta x'^\ell, \qquad (1.254)$$

where g' is the determinant formed by the elements g'_{ij}. The volume of any bounded region Λ is given by the limit of the sum of these infinitesimal rectangles which make up Λ.[28] Such a volume is written in the form of the following integral:

$$\int_\Lambda d\mu' = \int_\Lambda \sqrt{g'}\, dx'^1 dx'^2 \cdots dx'^\ell. \qquad (1.255)$$

Given any function f on $I\!\!M^\ell$ we also have an integral

$$\int_\Lambda f d\mu = \int_\Lambda f \sqrt{g'}\, dx'^1 dx'^2 \cdots dx'^\ell. \qquad (1.256)$$

Most relevant is the set of functions whose square is integrable in $I\!\!M^\ell$, since such a set of functions will form the basis for the construction of a Hilbert space in quantum mechanics.

[28] Schutz (1980) p. 121.

The integral in Eq. (1.256)) is an invariant independent of any choice of coordinate system. On a transformation to new coordinates $\bar{x}^j$ the integral becomes

$$\int_\Lambda f\, d\bar{\mu} = \int_\Lambda f\sqrt{\bar{g}}\, d\bar{x}^1 d\bar{x}^2 \cdots d\bar{x}^\ell, \tag{1.257}$$

where $\bar{g}$ is the determinant formed by the transformed $\bar{g}_{ij}$ which are related to g'_{ij} by Eq. (1.252). We have

$$\int_\Lambda f\sqrt{g'}\, dx'^1 dx'^2 \cdots dx'^\ell = \int_\Lambda f\sqrt{\bar{g}}\, d\bar{x}^1 d\bar{x}^2 \cdots d\bar{x}^\ell. \tag{1.258}$$

Finally we note that the volume element is identifiable with the ℓ-form $dx'^1 \wedge dx'^2 \wedge \cdots \wedge dx'^\ell$ although we shall not delve into this more abstract concept of volume.[29]

C5 The 4-dimensional space-time in Special Relativity is an example of a pseudo-Riemannian space. We shall not pursue pseudo-Riemannian spaces further in this book.

1.5.3 Hamiltonian manifolds

Definition 1.5.3(1) Hamiltonian structures and Hamiltonian manifolds

- A *Hamiltonian structure* $\Omega_H^{(2)}$ on a $2n$-dimensional manifold is a closed and non-degenerate two-form on the manifold.
- A $2n$-dimensional manifold together with a Hamiltonian structure is called a *Hamiltonian manifold*, to be denoted by $I\!M_H^{2n}$.

Theorem 1.5.3(1) Darboux's theorem on Hamiltonian manifolds[30]

- In a $2n$-dimensional Hamiltonian manifold $I\!M_H^{2n}$ there exist coordinates $(\mathsf{q}^j, \mathsf{p}_j), j = 1, 2, \ldots, n$, in the neighbourhood of every point $m \in I\!M_H^{2n}$ such that the Hamiltonian structure $\Omega_H^{(2)}$ is given by

$$\Omega_H^{(2)} = \sum_{j=1}^n d\mathsf{q}^j \wedge d\mathsf{p}_j. \tag{1.259}$$

These are called *canonical coordinates*.

[29] Nash and Sen (1983). We have avoided the definition of a volume element in $I\!M^\ell$ in terms of an ℓ-form. Instead we introduce the notion of a volume element in line with the usual and intuitive concept.

[30] Abraham and Marsden (1978).

Comments 1.5.3(1) Canonical transformations, functions and vector fields on $I\!M_H^{2n}$

C1 A Hamiltonian structure is also known as a *symplectic structure* and a Hamiltonian manifold as a *symplectic manifold*. The name Hamiltonian arises from the fact that such manifolds form the basis of Hamilton's formulation of classical mechanics. To see how a Hamiltonian structure comes about in the traditional Hamiltonian formulation of classical mechanics let us consider a particle moving in $I\!E^n$. Let the particle's position be given by rectangular Cartesian coordinates x^j in $I\!E^n$ and its motion specified by their conjugate momenta p_j. These momenta p_j form the components of a covariant vector[31] and this is why we use a subscript rather than a superscript. We can combine position and momentum variables to form a $2n$-dimensional space Γ known as the *phase space* which is coordinated by $2n$ variables x^j and p_j. We can construct a two-form

$$\Omega_H^{(2)} = \sum_{j=1}^{n} dx^j \wedge dp_j \tag{1.260}$$

which is seen to be closed and non-degenerate, i.e., $\Omega_H^{(2)}$ is a Hamiltonian structure on Γ, rendering Γ a Hamiltonian manifold. This phase space in classical mechanics is a paradigm of Hamiltonian manifolds, and this also explains why one set of canonical variables is labelled by a superscript and the other set by a subscript in Darbourx's theorem.

C2 In a Hamiltonian manifold we may also call q^j coordinate variables and p_j momentum variables. Moreover the momentum variable p_j is also said to be *canonically conjugate* to q^j. We may consider that the momentum variables formally span a space, called the *momentum space*, in a similar way to that in which the coordinate variables q^j span a *coordinate space*.

C3 Canonical coordinates are not unique, i.e., we can have other canonical coordinates $(\bar{\mathsf{q}}^j, \bar{\mathsf{p}}_j), j = 1, 2, \ldots, n$, in the neighbourhood of any point satisfying Eq. (1.259) of Darboux's theorem. The transformation from one set of canonical coordinates $(\mathsf{q}^j, \mathsf{p}_j)$ to another set $(\bar{\mathsf{q}}^j, \bar{\mathsf{p}}_j)$ is called a *canonical transformation*. A simple example is the transformation

$$\bar{\mathsf{q}}^j = -\mathsf{p}_j, \quad \bar{\mathsf{p}}_j = \mathsf{q}^j. \tag{1.261}$$

This shows that a coordinate variable $\bar{\mathsf{q}}^j$ need not necessarily refer to position and a momentum variable $\bar{\mathsf{p}}_j$ need not refer to motion.

C4 A function f on $I\!M_H^{2n}$ gives rise to a one-form $\Omega^{(1)} = df$. We can relate such a one-form to a vector field. To achieve this relationship we must

[31] See §1.1.3C(1) C3.

first choose a two-form, and then form the interior product between the vector field and the two-form.[32] In a Hamiltonian manifold $I\!M_H^{2n}$ the Hamiltonian structure $\Omega_H^{(2)}$ can serve as the chosen two-form. Then a one-form $\Omega^{(1)} = df$ together with $\Omega_H^{(2)}$ defines a vector field $\widehat{\mathcal{X}}_f$ on $I\!M_H^{2n}$ by

$$\widehat{\mathcal{X}}_f \rfloor\ \Omega_H^{(2)} = df. \tag{1.262}$$

Explicitly we have

$$\widehat{\mathcal{X}}_f = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial \mathsf{p}^j} \frac{\partial}{\partial \mathsf{q}^j} - \frac{\partial f}{\partial \mathsf{q}^j} \frac{\partial}{\partial \mathsf{p}_j} \right). \tag{1.263}$$

To see how such an expression is obtained we first note that following Eq. (1.75) a vector field $\widehat{\mathcal{X}}$ on $I\!M_H^{2n}$ is expressible as

$$\widehat{\mathcal{X}} = \sum_{j=1}^{n} \left(\mathcal{X}^j \frac{\partial}{\partial \mathsf{q}^j} + \mathcal{X}^{n+j} \frac{\partial}{\partial \mathsf{p}_j} \right). \tag{1.264}$$

Making use of Eqs. (1.221) and (1.224) we get

$$\begin{aligned} \widehat{\mathcal{X}} \rfloor\ \Omega_H^{(2)} &= \widehat{\mathcal{X}} \rfloor\ \sum_{j=1}^{n} \left(d\mathsf{q}^j \wedge d\mathsf{p}_j \right) & (1.265) \\ &= \sum_{j=1}^{n} \left\{ \left(\widehat{\mathcal{X}} \rfloor\ d\mathsf{q}^j \right) \wedge d\mathsf{p}_j - d\mathsf{q}^j \wedge \left(\widehat{\mathcal{X}} \rfloor\ d\mathsf{p}_j \right) \right\} & (1.266) \\ &= \sum_{j=1}^{n} \left(-\mathcal{X}^{n+j} d\mathsf{q}^j + \mathcal{X}^j d\mathsf{p}_j \right), & (1.267) \end{aligned}$$

on account of the fact that

$$\widehat{\mathcal{X}} \rfloor\ d\mathsf{q}^j = \mathcal{X}^j, \quad \widehat{\mathcal{X}} \rfloor\ d\mathsf{p}_j = \mathcal{X}^{n+j}. \tag{1.268}$$

Since

$$df = \sum_{j}^{n} \left(\frac{\partial f}{\partial \mathsf{q}^j} d\mathsf{q}^j + \frac{\partial f}{\partial \mathsf{p}_j} d\mathsf{p}_j \right), \tag{1.269}$$

we obtain from Eqs. (1.262) and (1.267) the following results:

$$\mathcal{X}_f^{n+j} = -\frac{\partial f}{\partial \mathsf{q}^j}, \quad \mathcal{X}_f^j = \frac{\partial f}{\partial \mathsf{p}_j}. \tag{1.270}$$

[32] See §1.4.4E(1) E3.

We conclude that a function f generates a vector field $\widehat{\mathcal{X}}_f$ through the Hamiltonian structure $\Omega_H^{(2)}$. This result leads to a number of concepts which are indispensable in formulating Hamiltonian mechanics geometrically. These are set out in Definition 1.5.3(2).

Definition 1.5.3(2) Hamiltonian vector fields, Poisson bracket

- In a Hamiltonian manifold $\mathbb{M}_H^{2n}$ a vector field is called the *Hamiltonian vector field* generated by a function f on $\mathbb{M}_H^{2n}$ if Eq. (1.262) is satisfied. Such a Hamiltonian vector field is denoted by $\widehat{\mathcal{X}}_f$. Explicitly, we have

$$f \quad \rightarrow \quad \widehat{\mathcal{X}}_f = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial \mathsf{p}_j} \frac{\partial}{\partial \mathsf{q}^j} - \frac{\partial f}{\partial \mathsf{q}^j} \frac{\partial}{\partial \mathsf{p}_j} \right). \tag{1.271}$$

- The Poisson bracket $\{f, g\}$ between two functions f and g on $\mathbb{M}_H^{2n}$ is defined to be the function

$$\{f, g\} = -\widehat{\mathcal{X}}_f(g). \tag{1.272}$$

Comments 1.5.3(2) Coordinate expressions and Lie algebras

C1 The coordinate expression for the Poisson bracket $\{f, g\}$ is

$$\{f, g\} = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial \mathsf{q}^j} \frac{\partial g}{\partial \mathsf{p}_j} - \frac{\partial f}{\partial \mathsf{p}_j} \frac{\partial g}{\partial \mathsf{q}^j} \right). \tag{1.273}$$

This agrees with the usual definition of Poisson bracket in classical mechanics. Care should be taken if the coordinates employed are not global. Canonical variables $\mathsf{q}^i, \mathsf{p}_j$ satisfy the following characteristic Poisson bracket relations:

$$\{\mathsf{q}^i, \mathsf{p}_j\} = \delta^i_j. \tag{1.274}$$

C2 One can verify that the Lie bracket between two Hamiltonian vector fields $\widehat{\mathcal{X}}_{f_1}$ and $\widehat{\mathcal{X}}_{f_2}$ is proportional to the Hamiltonian vector field $\widehat{\mathcal{X}}_f$ generated by the function f which is the Poisson bracket of f_1 and f_2, i.e.,[33]

$$[\widehat{\mathcal{X}}_{f_1}, \widehat{\mathcal{X}}_{f_2}] = -\widehat{\mathcal{X}}_f, \quad \text{where} \quad f = \{f_1, f_2\}. \tag{1.275}$$

C3 The set of all (smooth) functions on $\mathbb{M}_H^{2n}$ constitutes a *Lie algebra* with Lie algebra multiplication defined by the Poisson bracket. The set of Hamiltonian vector fields under Lie bracket also forms a Lie algebra.

[33] Abraham and Marsden (1978) p. 194.

1.6 Classical Dynamical Systems

1.6.1 Classical systems of finite order

A *classical system of order* $\ell < \infty$ is traditionally defined by three properties:[34]

1. States of the system are specifiable by ℓ parameters $\gamma = \{\gamma^1, \gamma^2, \ldots, \gamma^\ell\}$. Each parameter γ^j assumes a continuous set of real values. To start with we shall assume that each γ^j ranges over the entire $\mathbb{R}$.

2. Physical observables are represented by real-valued functions $A(\gamma)$ of the above parameters.

3. The time evolution of the system is represented by the time-dependence of the parameters which satisfy certain first-order differential equations of the form
$$\frac{d\gamma^j}{dt} = \mathcal{X}^j(\gamma, t) \tag{1.276}$$
for some smooth functions $\mathcal{X}^j(\gamma, t)$ of γ and t. These are referred to as *equations of motion*. Given any state γ_0 at time $t = 0$ the state at a later time is determined by the solution $\gamma(t)$ of these equations satisfying the initial conditions $\gamma_0 = \gamma(0)$, i.e., $\gamma_0^j = \gamma^j(0)$ for all j.

The collection of all states can be conveniently identified with the set Γ of ℓ-tuples $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$. To formalize this we call
$$\Gamma = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} \tag{1.277}$$
the *phase space*, or the *state space*, of the system. While it may be possible formally to impose a vector space structure together with a norm to convert Γ to a Euclidean space, the resulting structure, e.g., the norm of an element of Γ, may not have any particular physical meaning generally. So, when referring to Γ as a Euclidean space we should bear this in mind; it is often sufficient to treat the phase space as an ℓ-dimensional manifold without reference to any metric structure. We shall formally regard each set of parameter values as a *point* in Γ and the parameters as *coordinates* in Γ. If the range of variables γ^j is restricted the phase space becomes a subspace of $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$.

In geometric terms a state corresponds to a point γ in the phase space Γ, and an observable corresponds to a function A on Γ. The dynamics of the system can then be described by a vector field
$$\widehat{\mathcal{X}} = \sum_{j=1}^{\ell} \mathcal{X}^j \frac{\partial}{\partial \gamma^j} \tag{1.278}$$

[34] Percival and Richards (1982) §3.1, Perko (1991).

in Γ in that the motion of the system can be identified with the integral curves of $\widehat{\mathcal{X}}$ on the phase space. This vector field is called the *velocity field* of the system. Each point $\gamma_0 \in \Gamma$ can serve as an initial state and the starting point of an integral curve σ_{γ_0} of $\widehat{\mathcal{X}}$ with $\gamma(t) = \sigma_{\gamma_0}(t)$ representing the state at time t. Integral curves are obtained from the solutions of Eq. (1.276).

These dynamics may also be recast into the form of the flow of vector field $\widehat{\mathcal{X}}$, i.e., a one-parameter group of transformations of Γ

$$T_t : \Gamma \mapsto \Gamma, \quad t \in I\!R \qquad \text{or} \qquad T : I\!R \times \Gamma \to \Gamma \tag{1.279}$$

with differentiable and group properties as stated in Definition 1.3.3(2). In other words the dynamics are expressible as a one-parameter group of transformations of the phase space onto itself. The situation is simpler if the vector field is not explicitly time-dependent, i.e., $\mathcal{X}^j = \mathcal{X}^j(\gamma)$.

Definition 1.6.1(1) Stationary states and phase flows

- The map T in Eq. (1.279) is called the *phase flow* of the dynamics.
- A state $\gamma(0)$ is said to be *stationary* if $\gamma(t) = \gamma(0)$ for all t, or equivalently if $A(\gamma(t)) = A(\gamma(0))$ for all observables A.

Clearly a state $\gamma(0)$ is stationary if and only if the velocity field $\widehat{\mathcal{X}}$ vanishes at $\gamma(0)$, i.e., $\mathcal{X}^j(\gamma(0)) = 0$ for every j and for all times. In other words stationary states correspond to critical points of the velocity field. Stationary states can be further classified into various types, e.g., stable and unstable.[35]

1.6.2 First-order systems

The state of a *first-order classical system* is specified by a single real parameter $\gamma = \gamma^1$ lying in a certain range $\mathcal{R} \subset I\!R$. The phase space Γ of such a system is given by $\Gamma = \mathcal{R}$ with equation of motion[36]

$$\frac{d\gamma}{dt} = \mathcal{X}(\gamma), \quad \gamma \in \Gamma. \tag{1.280}$$

The corresponding velocity field is

$$\widehat{\mathcal{X}} = \mathcal{X}\frac{d}{d\gamma}, \qquad \mathcal{X} = \frac{d\gamma}{dt}. \tag{1.281}$$

[35] Percival and Richards (1982) §3.3 to §3.5. Stationary states are often referred to as *fixed points* of the vector field $\widehat{\mathcal{X}}$.

[36] For many physical systems $\mathcal{X}$ can also contain the time variable t explicitly.

Examples 1.6.2(1) Electrical circuits as first-order systems

E1 Consider an idealized classical resistanceless circuit containing an inductor with an inductance L in a time-dependent external magnetic field. The applied magnetic flux Φ_{ex} threading through the inductor varies as the external magnetic field changes, and a current I is generated according to Faraday's and Lenz's laws:[37]

$$L\,\frac{dI}{dt} = -\frac{d\Phi_{ex}}{dt}. \tag{1.282}$$

Integrating from 0 to t we get

$$LI(t) - LI(0) = -\Big(\Phi_{ex}(t) - \Phi_{ex}(0)\Big), \tag{1.283}$$

or

$$LI(t) + \Phi_{ex}(t) = \Phi_T, \quad \Phi_T = LI(0) + \Phi_{ex}(0). \tag{1.284}$$

Physically we can interpret Φ_T as the total flux threading through the inductor, being the sum of the external flux and the induced flux due to the current. The fact that Φ_T is a constant means that the induced current $I(t)$ must vary with $\Phi_{ex}(t)$ in such a way as to keep the total flux Φ_T constant. The state of this system is specifiable by a single parameter, the current I. The energy of the system is known is to be given by

$$E = \frac{1}{2}\,LI^2 = \frac{1}{2L}\,(\Phi_T - \Phi_{ex})^2. \tag{1.285}$$

On the phase space $\Gamma = \{I \in I\!R\}$ the vector field corresponding to the dynamics of the circuit is

$$\widehat{\mathcal{X}} = \mathcal{X}\,\frac{d}{dI}, \quad \mathcal{X} = -\frac{1}{L}\,\frac{d\Phi_{ex}}{dt}. \tag{1.286}$$

We shall return to this system again in §4.5.1.

1.6.3 Second-order Hamiltonian systems

The phase space for second-order systems is parameterized by two coordinates, i.e., $\Gamma = \{(\gamma^1, \gamma^2)\}$. We are particularly interested in second-order Hamiltonian systems which are defined in geometric terms in Definition 1.6.3(1).

[37] Bleaney and Bleaney (1957) p. 151.

Definition 1.6.3(1) Second-order Hamiltonian systems

- A second-order system is called a *Hamiltonian system* if the phase space Γ is a two-dimensional Hamiltonian manifold $\mathbb{M}_H^2$ with a Hamiltonian structure $\Omega_H^{(2)}$ and a chosen real-valued function H_g on $\mathbb{M}_H^2$, to be called the *Hamiltonian generator* of the system, such that the dynamics is generated by the Hamiltonian vector field $\widehat{\mathcal{X}}_{H_g}$ of H_g.

Comments 1.6.3(1) Hamiltonians, Hamiltonian generators, phase flow and Liouville's theorem

C1 The dynamics of a Hamiltonian system is generated by the Hamiltonian vector field $\widehat{\mathcal{X}}_{H_g}$ acting as the velocity field. As a result the Hamiltonian generator H_g is often referred to as the *generator of time translations*.[38]

C2 The energy of a Hamiltonian system is often referred to as the *Hamiltonian of the system*, denoted by H. We must emphasize that in principle and in general H is not the same as H_g, despite the fact that in many practical cases the energy of the system turns out to be equal to the Hamiltonian generator.[39] We shall return to this important point later in quantum theory.

C3 According to Darboux's theorem there are canonical coordinates (q, p) in the neighbourhood of every point in Γ such that $\Omega_H^{(2)} = d\mathsf{q} \wedge d\mathsf{p}$. The Hamiltonian vector field $\widehat{\mathcal{X}}_{H_g}$ is given in accordance with Eq. (1.271), by

$$\widehat{\mathcal{X}}_{H_g} = \frac{\partial H_g}{\partial \mathsf{p}} \frac{\partial}{\partial \mathsf{q}} - \frac{\partial H_g}{\partial \mathsf{q}} \frac{\partial}{\partial \mathsf{p}}. \tag{1.287}$$

The dynamics is given by the integral curves of $\widehat{\mathcal{X}}_{H_g}$, i.e., solutions of the following equations:

$$\frac{d\mathsf{q}}{dt} = \frac{\partial H_g}{\partial \mathsf{p}}, \quad \frac{d\mathsf{p}}{dt} = -\frac{\partial H_g}{\partial \mathsf{q}}. \tag{1.288}$$

These are the well-known *Hamilton's equations of motion* in classical mechanics.[40]

C4 Stationary states occur at the critical points of the Hamiltonian vector field, i.e., where

$$\frac{\partial H_g}{\partial \mathsf{p}} = 0 \quad \text{and} \quad \frac{\partial H_g}{\partial \mathsf{q}} = 0. \tag{1.289}$$

---

[38] Goldstein (1950) p. 259.

[39] The fact that H_g may not be the energy of the system is clearly pointed out by Goldstein (1950) p. 221. For an example in which $H_g \neq H$ see Leech (1965) p. 48.

[40] Bishop and Goldberg (1968) Chapter 6.

C5 The nature of the phase flow T in Hamiltonian dynamics is characterized by *Liouville's Theorem* which states that a phase flow in Hamiltonian dynamics preserves the *phase volume*.[41] For a second-order system an elementary phase volume is just $d\mathsf{q}d\mathsf{p}$. At time $t = \tau$ later the point (q, p) is mapped to the point $(\mathsf{q}_\tau, \mathsf{p}_\tau) = T_\tau(\mathsf{q}, \mathsf{p})$. Liouville's Theorem implies

$$d\mathsf{q}_\tau d\mathsf{p}_\tau = d\mathsf{q}d\mathsf{p}. \tag{1.290}$$

Examples 1.6.3(1) Hamiltonian generators, Hamiltonians and Hamilton's equations

E1 Consider a particle of mass m in one-dimensional motion in $\mathbb{E}$ under a potential V which is a differentiable function on $\mathbb{E}$. In classical mechanics the state of the particle is fixed by its position x and canonically conjugate momentum p. The dynamics are governed by Hamilton's Eqs. (1.288) with a Hamiltonian generator H_g given by

$$H_g = \frac{p^2}{2m} + V. \tag{1.291}$$

This is a typical example of a second-order Hamiltonian system. The phase space is

$$\Gamma = \{(\gamma^1 = x, \gamma^2 = p)\} \quad \text{with} \quad \Omega_H^{(2)} = dx \wedge dp. \tag{1.292}$$

A point $\gamma = (x, p)$ in Γ corresponds to a state of the particle. The Hamiltonian vector field for the dynamics is

$$\begin{aligned} \widehat{\mathcal{X}}_{H_g} &= \frac{\partial H_g}{\partial p}\frac{\partial}{\partial q} - \frac{\partial H_g}{\partial x}\frac{\partial}{\partial p} \\ &= \frac{p}{m}\frac{\partial}{\partial x} - \frac{\partial V}{\partial x}\frac{\partial}{\partial p}. \end{aligned} \tag{1.293}$$

The integral curves are then solutions of Hamilton's equations

$$\frac{dx}{dt} = \frac{p}{m}, \quad \frac{dp}{dt} = -\frac{\partial V}{\partial x}. \tag{1.294}$$

For such a system the Hamiltonian generator H_g is equal to the Hamiltonian (energy) H of the system. For a free particle the Hamiltonian vector field reduces to

$$\widehat{\mathcal{X}}_{H_g} = \frac{p}{m}\frac{\partial}{\partial x}. \tag{1.295}$$

[41] Arnold (1978) pp. 68-70, Abraham and Marsden (1978) Proposition 3.3.4. For a simple proof see Percival and Richards (1982) §4.9, §6.2 who use the term *phase area* rather than phase volume.

E2 A canonical momentum conjugate to a coordinate variable is not unique. In the case of a particle in motion in $\mathbb{E}$ the variable

$$p' = p + f(x) \tag{1.296}$$

is also canonically conjugate to x. This is seen in the Poisson brackets

$$\{x, p'\} = \{x, p\} = 1, \tag{1.297}$$

which agree with the Poisson bracket relations given by Eq. (1.274) of canonical variables. One can also confirm the canonical nature of x and p' by going back to Theorem 1.5.3(1) to check that $dx \wedge dp'$ and $dx \wedge dp$ give rise to the same Hamiltonian structure. Indeed (x, p') may even be regarded as a canonical transformation of (x, p).

E3 Another example is a classical resistanceless LC circuit consisting of a capacitor with capacitance C and an inductor with inductance L. When a current I flows there will be a charge Q on the capacitor linked to the current by $I = dQ/dt$ and a magnetic flux $\Phi = LI$ through the inductor. A voltage $V = Q/C$ will develop across the capacity. The circuit equation is

$$L\,\frac{dI}{dt} + \frac{Q}{C} = 0, \tag{1.298}$$

or

$$\frac{d^2Q}{dt^2} = -\omega^2 Q, \quad \omega^2 = 1/LC. \tag{1.299}$$

We can also express the circuit equation in terms of Φ, i.e., we have

$$\frac{d^2\Phi}{dt^2} = -\omega^2\,\Phi. \tag{1.300}$$

This is a Hamiltonian system. We have a phase space

$$\Gamma = \{(\mathsf{q} = \Phi, \mathsf{p} = Q)\} \quad \text{with} \quad \Omega_H^{(2)} = d\Phi \wedge dQ. \tag{1.301}$$

Here Φ and Q are treated as a pair of canonical variables. One can easily verified that an appropriate Hamiltonian generator for this system is

$$H_g = \frac{Q^2}{2C} + \frac{\Phi^2}{2L}. \tag{1.302}$$

This circuit system is essentially the same as a harmonic oscillator in mechanics. The only stationary state is at $(\Phi = 0, Q = 0)$.

E4 Consider a particle confined to move in a circle $\mathcal{C}$ which is coordinated by a variable $\theta \in [0, 2\pi]$ as discussed in §1.5.1E(1) E2 and E5. The phase space

Γ is parameterized by θ and its conjugate momentum[42] $P_\theta(\mathcal{C}) \in (-\infty, \infty)$. Here Γ can be visualized as a 2-dimensional cylinder, i.e., $\Gamma = \mathcal{C} \times I\!\!R$, showing the non-Euclidean nature of the Hamiltonian manifold.

E5 The definition of Hamiltonian systems can be extended to higher order, i.e., to order $2n$ with a $2n$-dimensional manifold $I\!\!M_H^{2n}$ and canonical coordinates $(\mathsf{q}^j, \mathsf{p}_j)$, $j = 1, 2, \ldots, n$.

1.6.4 Momentum observables, vector fields and operators

Many of the geometric methods developed in recent years are very powerful and are directly applicable to the formulation of many physical theories. These methods are capable of producing a unified treatment of seemingly diverse subjects such as differential equations, differential geometry and dynamical systems such as classical mechanical systems. Moreover, they lead to a deeper understanding of classical mechanics and its transition to quantum mechanics. This is why we have devoted so much space to their discussion. As illustrations we shall consider a class of observables, to be referred to as *momentum observables*, which play a fundamental role in both classical mechanics and quantum mechanics. This section is devoted to reviewing these observables within classical mechanics as preparation to their quantization which will be discussed in Chapter 3.

Consider a particle moving in $I\!\!E^n$ with rectangular Cartesian coordinates x^j. Let us call the space $I\!\!E^n$ in which the particle moves the *physical coordinate space*, or **physical space** or **coordinate space** for short. Let the particle's momenta conjugate to x^j be denoted by p_j. Then the phase space Γ is a Hamiltonian manifold coordinated by x^j, p_j with a Hamiltonian structure $\Omega_H^2 = \sum_j dx^j \wedge dp_j$. An observable is by definition a function A on the $2n$-dimensional phase space. We start by examining functions linear in p_j, i.e., functions of the form

$$P = \sum_{k=1}^{n} \xi^k(x) p_k, \tag{1.303}$$

where $\xi^k(x)$ are functions of coordinates x^j only. According to Eq. (1.271) such a momentum observable generates a Hamiltonian vector field on Γ, i.e.,

$$\widehat{\mathcal{X}}_P = \sum_{j=1}^{n} \left(\xi^j(x) \frac{\partial}{\partial x^j} - \sum_k \left(\frac{\partial \xi^k}{\partial x^j} p_k \right) \frac{\partial}{\partial p_j} \right). \tag{1.304}$$

[42] See §1.6.4E(1) E5.

The fact that ξ^j are functions of coordinates only enables us to project this vector field down from the phase space Γ to the coordinate space $I\!E^n$, i.e., we can generate a vector field $\widehat{X}_P$ on $I\!E^n$ given by

$$\widehat{X}_P = \sum_{j=1}^{n} \xi^j(x) \frac{\partial}{\partial x^j}. \tag{1.305}$$

We shall call $\widehat{X}_P$ the *Hamiltonian vector field* of P on $I\!E^n$.

Definition 1.6.4(1) Complete and incomplete momenta

- For a Hamiltonian system an observable of the form

$$P = \sum_{k=1}^{n} \xi^k(x) p_k, \tag{1.306}$$

  where ξ^k are functions of coordinates x^j only, is called a *momentum observable based on the physical space* $I\!E^n$ or simply a momentum.

- A momentum $P = \sum_{k=1}^{n} \xi^k(x) p_k$ is said to be
  - *complete* if its Hamiltonian vector $\widehat{X}_P$ given by Eq. (1.305) on the coordinate space $I\!E^n$ is complete;
  - *incomplete* if $\widehat{X}_P$ is incomplete in $I\!E^n$.

Comments 1.6.4(1) Generators of transformation groups

C1 The simplest momenta are canonical variables p_j conjugate to coordinates x^j. It is possible to effect a canonical transformation from (x^j, p_j) to another canonical variables $(\bar{x}^j, \bar{p}_j)$ such that $\bar{x}^j$ do not relate solely to position in the coordinate space. We could have, according to Eq. (1.261),

$$\bar{x}^j = -p_j, \quad \bar{p}_j = x^j. \tag{1.307}$$

The canonical momentum $\bar{p}_j$ conjugate to canonical coordinate $\bar{x}^j$ is no longer a momentum observable based on the physical space $I\!E^n$. This example serves to demonstrate the distinction between a momentum observable based on the coordinate space and a canonical momentum

When there is no risk of confusion we shall simply call a momentum based on the coordinate space a momentum.

C2 A momentum generates a Hamiltonian vector field $\widehat{\mathcal{X}}_P$ on Γ and a separate Hamiltonian vector field $\widehat{X}_P$ on $I\!E^n$.

C3 A *complete momentum* is often referred to as the *generator of a one-parameter group of transformations* of the coordinate space $\mathbb{E}^n$ in the form of translations along the integral curves of its Hamiltonian vector field $\widehat{X}_P$ in the coordinate space $\mathbb{E}^n$. An incomplete momentum does not generate a one-parameter group of transformations. The distinction between complete and incomplete momenta will have a dramatic effect in their quantization. In the examples presented below we shall consider the motion of a particle in $\mathbb{E}^3$ with the usual notation (x, y, z) for the rectangular Cartesian coordinates with (p_x, p_y, p_z) for their respective conjugate momenta. We also present an example of a particle constrained to move in a circle $\mathcal{C}$; we shall return to their quantization in §6.7, §6.8 and §6.9 in Chapter 6.

Examples 1.6.4(1) Linear and angular momenta

E1 Linear momenta in $\mathbb{E}^3$ The momentum p_x conjugate to x is referred to as a *linear momentum*. Its Hamiltonian vector field on $\mathbb{E}^3$ is

$$\widehat{X}_{p_x} = \frac{\partial}{\partial x}. \tag{1.308}$$

This vector field is clearly complete. As pointed out in §1.3.3C(2) C2 such a vector field also generates a one-parameter group of transformations of $\mathbb{E}^3$ in the form of translations along the coordinate variable x by

$$x \mapsto x(\tau) = T_\tau x = \tau + x. \tag{1.309}$$

We shall call this a *group of translations* along x.

E2 Angular momenta in $\mathbb{E}^3$ The momentum $L_z = yp_x - xp_y$ gives rise to a Hamiltonian vector field

$$\widehat{X}_{L_z} = y\partial/\partial x - x\partial/\partial y \tag{1.310}$$

on $\mathbb{E}^3$ which generates rotations about z-axis. This vector field is complete, as shown in §1.3.2E(1) E4, and its integral curves are circles lying on planes parallel to the x-y plane and centred at z-axis. We identify L_z with the angular momentum component along z-axis. This momentum is complete and may be interpreted as the generator for rotations about z-axis. Both L_z and $\widehat{X}_{L_z}$ are zero along z-axis.

E3 Momentum p_θ in $\mathbb{E}^3$ In §1.2.3C(1) C3 we introduce the canonical momentum p_θ conjugate to the angle variable $\theta \in (0, \pi)$ in spherical coordinates. From Eq. (1.31) we can write down its Hamiltonian vector field on $\mathbb{E}^3$ as

$$\widehat{X}_{p_\theta} = \frac{xz}{\sqrt{x^2+y^2}}\frac{\partial}{\partial x} + \frac{yz}{\sqrt{x^2+y^2}}\frac{\partial}{\partial y} - \sqrt{x^2+y^2}\,\frac{\partial}{\partial z}. \tag{1.311}$$

In spherical coordinates this expression becomes

$$X_{p_\theta} = \frac{\partial}{\partial \theta}. \tag{1.312}$$

We can work out the integral curve starting from a point $(r_0, \theta_0, \varphi_0)$ explicitly. It is a solution of

$$\frac{dr}{d\tau} = 0, \quad \frac{d\theta}{d\tau} = 1, \quad \frac{d\varphi}{d\tau} = 0, \tag{1.313}$$

namely

$$r = r_0, \quad \theta = \tau + \theta_0, \quad \varphi = \varphi_0. \tag{1.314}$$

Since τ is confined to the range $(-\theta_0, \pi - \theta_0)$ to keep θ in the range $(0, \pi)$ the vector field, and hence the momentum, are both incomplete.

E4 Radial momentum in $I\!\!E^3$ The momentum p_r in §1.2.3C(1) C3 is called the *radial momentum.* Its Hamiltonian vector field $\widehat{X}_{p_r}$ on $I\!\!E^3$ is

$$\widehat{X}_{p_r} = \frac{x}{r}\frac{\partial}{\partial x} + \frac{y}{r}\frac{\partial}{\partial y} + \frac{z}{r}\frac{\partial}{\partial z}, \quad r > 0. \tag{1.315}$$

In the usual spherical coordinates (r, θ, φ) this expression becomes

$$\widehat{X}_{p_r} = \frac{\partial}{\partial r}, \quad r > 0. \tag{1.316}$$

The integral curve starting from a point $(r_0, \theta_0, \varphi_0)$ is given by

$$r = \tau + r_0, \quad \theta = \theta_0, \quad \varphi = \varphi_0. \tag{1.317}$$

Since τ is confined to the range $(-r_0, \infty)$ to keep r in the range $(0, \infty)$ the vector field is incomplete. To illustrate the link between the group nature of translations and completeness stated in §1.3.3C(2) C2 we can see here that the constraint on the values of r to the range $(0, \infty)$ also implies that translations along coordinate r do not form a group, e.g., there is no inverse translation since such a translation will render some r negative. Similarly we can see that the incompleteness of p_θ is related to the restricted range of the conjugate coordinate θ which makes a translation along θ undefinable for every value of θ.

In passing we should point out that p_r and $\widehat{X}_{p_r}$ are not defined along the z-axis where $x = y = 0$.

E5 Momentum $P_\theta(\mathcal{C})$ conjugate to θ in $\mathcal{C}$ For a particle constrained to move in a circle $\mathcal{C}$ of radius a in $I\!\!E^2$ and centered at the coordinate origin its position can be specified by the polar angle θ in §1.5.1E(1) E2. Let us denote the momentum conjugate to θ by $P_\theta(\mathcal{C})$. In accordance with Eq. (1.271) its associated Hamiltonian vector field $\widehat{X}_{P_\theta(\mathcal{C})}$ on $\mathcal{C}$ is given by Eq. (1.241). This vector field is complete. Hence $P_\theta(\mathcal{C})$ is a complete momentum for a particle constrained to move in a circle $\mathcal{C}$. Momentum $P_\theta(\mathcal{C})$ is identifiable with the traditional angular momentum. Note that this is quite different from p_θ in $I\!E^3$ defined by Eq. (1.31) and discussed in E3 above.

1.6.5 Concluding remarks

In this chapter we have presented an intrinsic approach to establish a number of geometric quantities on Euclidean spaces and manifolds. The starting point is the set of smooth functions which serve as the basic coordinate independent constructs on a manifold on which to build up other geometric quantities. Derivations are then introduced as linear operators satisfying the product rule of differentiation when acting on the set of functions. Naturally derivations are identifiable with differential operators. All these quantities are defined in a coordinate independent way. Fundamentally contravariant vectors are seen to be differential operators. In other words differential operators provide an intrinsic definition of contravariant vectors. Covariant vectors can then be identified with differentials of functions. Smooth assignments of vectors to every point in the manifold lead to the establishment of vector fields.

We formulate classical mechanics in a geometric language.[43] The basic geometric space is the phase space Γ as a Hamiltonian manifold. In the phase space every function f generates a vector field $\widehat{\mathcal{X}}_f$ known as its Hamiltonian vector field on the phase space. The dynamics can be described in terms of the Hamiltonian vector field $\widehat{\mathcal{X}}_{H_g}$ of a function H_g on Γ, known as the Hamiltonian generator. An important class of observables is the momentum observables P. A momentum P can generate a vector field $\widehat{X}_P$ on the coordinate space, referred to as its Hamiltonian vector field on the coordinate space. This property leads us to the following conclusions:

- We can interpret a momentum as a generator for spatial translations.
- We can classify momenta into complete ones and incomplete ones.
- Bearing in mind that vector fields are identifiable with differential operators we see that momentum observables are intrinsically linked to differential operators. This indicates a link between classical mechanics and quantum mechanics where observables are described directly in

[43] We have limited the discussion to what we need later for quantization purposes, and refrained from going deeper into the mathematical and geometric foundations of classical mechanics, e.g., we have not mentioned momentum mapping on sympletic geometry (Abraham and Marsden (1978)) and Poisson manifolds (Choquet-Bruhat and de Witt-Morette (1989)).

terms of operators, i.e., a classical momentum P may be quantized in terms of its Hamiltonian vector field $\widehat{X}_P$ in the coordinate space. This we shall do later in Chapter 3. This method of quantization, referred to as geometric quantization, will be seen to ease the conceptual transition from classical mechanics to quantum mechanics.

References

1. Abraham, R. and Marsden, J. E. (1978). *Foundations of Mechanics*, Benjamin/Cummings, Reading, Mass.
2. Arnold, V. I. (1978). *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York.
3. Benn, I. M. and Tucker, R. W. (1987). *An Introduction to Spinors and Geometry*, Adam Hilger, Bristol.
4. Bishop, R. L. and Goldberg, S. I. (1968). *Tensor Analysis on Manifolds*, Macmillan, New York.
5. Bleaney, B. I. and Bleaney, B. (1957). *Electricity and Magnetism*, Clarendon Press, Oxford.
6. Brickell, F. and Clark, R. S. (1970). *Differentiable Manifolds*, Van Nostrand Reinhold, London.
7. Choquet-Bruhat, Y., de Witt-Morette, C., with Dillard-Bleick, M. (1982). *Analysis, Manfolds and Physics Part I: Basics*, North-Holland, Amsterdam.
8. Choquet-Bruhat, Y., de Witt-Morette, C. (1989). *Analysis, Manfolds and Physics Part II: 92 Applications*, North-Holland, Amsterdam.
9. Darling, R. W. R. (1994). *Differential Forms and Connections*, Cambridge University Press, Cambridge.
10. Frankel, T. (1997). *The Geometry of Physics*, Cambridge University Press, Cambridge.
11. Goldstein, H. (1950). *Classical Mechanics*, Addison-Wesley, Reading, Mass.
12. Isham, C. J. (1989). *Modern Differential Geometry for Physicists*, World Scientific, Singapore.
13. Leech, J. W. (1965). *Classical Mechanics*, Methuen & Co Ltd, London.
14. Lipschutz, S. (1965). *General Topoloogy*, Schaum's Outline Series, McGraw-Hill, New York.
15. Loomis, L. H. and Sternberg, S. (1968). *Advanced Calculus*, Addison-Wesley, Reading, Mass.

16. Martin, D. (1991). *Manifold Theory*, Ellis Harwood, New York.
17. Matsushima, Y. (1972). *Differentiable Manifolds*, Marcel Dekker, New York.
18. Nash, C. and Sen, S. (1983). *Topology and Geometry for Physicists*, Academic Press, London.
19. Percival, I. and Richards, D. (1982). *Introduction to Dynamics*, Cambridge University Press, Cambridge.
20. Perko, L. (1991). *Differential Equations and Dynamical Systems*, Springer-Verlag, New York.
21. Schutz, B. (1980). *Geometrical Methods of Mathematical Physics*, Cambridge University Press, Cambridge.
22. Simmons, G. F. (1963). *Introduction to Topology and Modern Analysis*, McGraw-Hill International Edition, Tokyo.
23. Sutherland, W. A. (1995). *Introduction to Metric and Topological Spaces*, Clarendon Press, Oxford.
24. Synge, J. L. and Schild, A. (1966). *Tensor Calculus*, University of Toronto Press, Toronto.