A. An Elementary Introduction to Manifolds and Lie Groups

A.1 Introduction

The following is an elementary introduction to some basic concepts in modern differential geometry. The aim is to provide the reader with a clear understanding of the key ideas and to motivate the concepts rather than to promote the rigor and generality pursued in mathematical texts. Specifically, our main objective is to arrive at a comprehensive description of *smooth manifolds*, *Lie groups*, and their most basic properties.

In particular, we shall introduce the following specific concepts:

1. metric space;
2. open and closed subsets of a metric space;
3. equivalence relation;
4. bijection;
5. isometry;
6. continuous function between metric spaces;
7. topological space;
8. continuous function between topological spaces;
9. homeomorphism;
10. connected topological space;
11. compact topological space;
12. topological manifold;
13. differentiable and smooth manifolds;
14. diffeomorphism;
15. complex projective space;
16. smooth curve on a smooth manifold;
17. tangent and cotangent spaces;
18. vector and tensor fields;
19. associative algebra;
20. differential forms and exterior differentiation;
21. push-forward map;
22. pullback map for differential forms;
23. compact manifold;
24. group;
25. group homomorphism and isomorphism;

26. Lie group;
27. action of a Lie group on a manifold;
28. transitive and free actions;
29. Lie algebra of a Lie group;
30. abstract Lie algebra;
31. unitary groups;
32. representation of a Lie group;
33. irreducible representation;
34. representation of a Lie algebra;
35. enveloping algebra;
36. Casimir operator.

The abstract definitions are supplemented with concrete examples.

For a more thorough treatment of this material, we wish to refer the reader to the following introductory textbooks:

- B. Schutz: *Geometrical Methods of Mathematical Physics* (Cambridge University Press, Cambridge, 1980);
- C. Isham: *Modern Differential Geometry for Physicists* (World Scientific, Singapore, 1989);
- M. Nakahara: *Geometry, Topology and Physics* (Adam Hilger, Bristol, 1990);
- C. Nash and S. Sen: *Topology and Geometry for Physicists* (Academic Press, London, 1983);
- Y. Choquet-Bruhat, C. De Witt-Morette, and M. Dillard-Bleick: *Analysis, Manifolds and Physics*, Parts I and II (North-Holland, Amsterdam, 1989);
- R. Geroch: *Mathematical Physics* (The University of Chicago Press, Chicago, 1985);
- V. Guillemin and A. Pollack: *Differential Topology* (Prentice-Hall, Inc., New Jersey, 1974).

A more advanced book on the subject is:

- S. Helgason: *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic Press, New York, 1978).

Before going through the discussion of manifolds and Lie groups, we would like to introduce the reader to some more basic mathematical concepts.

Consider the real line $\mathbb{R}$ as a set of points (numbers). Let $[a, b]$ and (a, b) denote the "closed" and "open" intervals in $\mathbb{R}$, with $a < b$. The points a and b are called the "boundary" or "limit points" of both $[a, b]$ and (a, b). Alternatively, we say that the set $\{a, b\}$ is the "boundary" of $[a, b]$ and (a, b). The distinguishing feature between these two intervals is that $[a, b]$ includes its boundary whereas (a, b) does not. We can use this property to define the notions of "open" and "closed" sets in $\mathbb{R}$, namely we define a *closed set* to be a subset of $\mathbb{R}$ that includes its boundary. Similarly, we call a subset *open*, if it does not include any of its boundary points.[1]

In view of these definitions, we can make several interesting observations. First, the complement of an open set is closed and vice versa. However, this does not contradict the existence of subsets that are neither open nor closed. The half-open (half-closed) intervals such as $[a, b)$ are examples of such subsets. Next, we can easily convince ourselves that the union and the intersection of two open (closed) sets are open (respectively closed). The only trouble seems to be the case where the two open (closed) sets have an empty intersection. This is remedied by postulating that the empty set and the universal set, in this case $\mathbb{R}$, are both open and closed. This assumption is also in agreement with the statement that the complement $\mathbb{R} - O$ of an open (closed) set O is closed (open). Furthermore, we can show that the union of any infinite collection of open sets is also open whereas their intersections may or may not be open. A standard example is the infinite family of open intervals defined by

$$\left\{\left(-\frac{n+1}{n}, \frac{n+1}{n}\right) : n \in \mathbb{Z}^+\right\}.$$

Evidently, the union of all such intervals is the open interval

$$\bigcup_{n=1}^{\infty}\left(-\frac{n+1}{n}, \frac{n+1}{n}\right) = (-2, 2),$$

whereas their intersection,

$$\bigcap_{n=1}^{\infty}\left(-\frac{n+1}{n}, \frac{n+1}{n}\right) = [-1, 1]$$

is a closed interval.

An important fact about point sets such as $\mathbb{R}$ is that each point p can be included in some open interval. For example, for every positive number $\epsilon \in \mathbb{R}^+$, $p \in (p - \epsilon, p + \epsilon)$. The open subset

$$O_\epsilon(p) := (p - \epsilon, p + \epsilon) = \{x \in \mathbb{R} : |x - p| < \epsilon\}$$

is called an "open neighborhood" or simply a "neighborhood" of p.

So far, we have introduced several simple mathematical concepts using the example of the real line as our universal set. The set of real numbers enjoys the property of having a natural notion of distance between its elements. This is indeed the main ingredient that allows us to define limit points, open and

[1] These definitions are clearly based on the definition of a *boundary* or *limit point*. We do not offer the latter at this stage hoping that the reader is guided by his or her intuition. We shall present precise definitions of open and closed sets later in this appendix.

closed subsets, and the neighborhoods of points. The set $\mathbb{R}$ with its usual notion of distance is an example of a large class of mathematical structures called *metric spaces.* These are sets of points with a well-defined concept of distance between their elements. The following definition describes metric spaces in precise terms.

Definition 1: Let X be a set of points and $d : X \times X \to [0, \infty)$ be a function that assigns a non-negative real number to any two elements of X and satisfies the following conditions:

1) For every $p, q \in X$, $d(p, q) = 0$ if and only if $p = q$.
2) For every $p, q \in X$, $d(p, q) = d(q, p)$.
3) For every $p, q, r \in X$, $d(p, q) + d(q, r) \geq d(p, r)$.

Then, the pair (X, d) is said to be a **metric space** and d is called a **metric** or a **distance function**. Alternatively, it is said that X has a **metric structure** specified by d.

Once a set is endowed with a metric structure we can immediately define the *open neighborhoods* of its points. For every $p \in X$ and $\epsilon > 0$, we define the open neighborhoods of p by

$$N_\epsilon(p) := \{x \in X \ : \ d(x, p) < \epsilon\}.$$

These are also called the *open balls* centered at p. Having defined the notion of an open neighborhood, we proceed with the definitions of open and closed subsets of a metric space.

Definition 2: Let (X, d) be a metric space. A subset $O \subseteq X$ is said to be an **open subset** if every point $x \in O$ has at least an open neighborhood $N_\epsilon(x)$ that lies entirely inside O, i.e., $N_\epsilon(x) \subset O$. A subset $C \subseteq X$ is said to be a **closed subset** if its complement, $X - C$, is open.[2]

Using this definition and assuming that the empty set $\emptyset$ and the universal set X are both open and closed, we can show that the finite and infinite unions and finite intersections of open sets are open. We shall see that these properties play an important role in generalizing the concepts of openness and closedness to the structures that lack the notion of distance. This is a general pattern in the methodology of mathematics. The results or theorems that are derived as logical consequences for special structures can be employed as postulates for more general structures. In this case, these more general structures are called *topological spaces.*

In mathematics once one defines a structure such as a metric space, the next task becomes to investigate the problem of the classification of such structures. In order to pursue the classification problem, one first needs to

[2] We can alternatively give a definition of limit points for metric spaces and retain our definitions of open and closed sets for $\mathbb{R}$. Following this approach, a point p is said to be a *limit point* of a subset $Y \subset X$, if every open neighborhood of p intersects both Y and $X - Y$. Then, a subset is called open if it does not include any of its limit points. It is called closed if it includes all its limit points.

have a clear understanding of the notion of "equivalence" of two structures. For example, to be able to compare two metric spaces one must know under what conditions they are "identical" or "equivalent." This raises the question of what "equivalence" means in precise mathematical language. It does not take much effort to realize that an "equivalence" is a sort of "relation" between two objects. It has three intuitively sound and simple properties.

Definition 3: Let $\mathcal{S}$ be a collection of objects and $\sim$ be a relation between any two of its elements. The relation $\sim$ is said to be an **equivalence relation** if it satisfies the following conditions:

1) For every $s \in \mathcal{S}$, $s \sim s$ (reflexivity).
2) For every $s_1, s_2 \in \mathcal{S}$, $s_1 \sim s_2$ implies $s_2 \sim s_1$ (symmetry).
3) For every $s_1, s_2, s_3 \in \mathcal{S}$, if $s_1 \sim s_2$ and $s_2 \sim s_3$, then $s_1 \sim s_3$ (transitivity).

An important property of an equivalence relation is that it divides the universal collection $\mathcal{S}$ into distinct (non-intersecting) subcollections. These are called the *equivalence classes*. As is clear from the nomenclature, each equivalence class consists of objects that are equivalent, i.e., related by the equivalence relation. Whence, each member of an equivalence class can represent the whole class as equally well as any other member.

In physics, one usually uses the word *symmetry* when such a situation occurs. Physicists would say that there is a symmetry between the members of each equivalence class that allows one to represent the whole class using a particular member. Symmetry is a desirable quality because it permits the freedom of choice of the representative for each class.[3] In this sense, the notion of symmetry is associated with the notion of equivalence. In the following, we shall examine some examples of equivalence relations in the mathematical arena.

Let us consider the set of positive integers (natural numbers) $\mathbb{Z}^+$. For any pair $p, n \in \mathbb{Z}^+$, we have

$$p = mn + r,$$

where m and r are integers such that $m \geq 0$ and $0 \leq r < n$. "r" is called the *remainder* (of the division of p by n). We know from arithmetic that m and r are uniquely determined. We can use this result to set up an equivalence relation between any two integers. Let us choose a positive integer n and define any two positive integers p_1 and p_2 to be equivalent if they correspond to the same remainder r upon division by n. In mathematical symbols, we write

$$p_1 \equiv p_2 \pmod{n}.$$

This means that there are $m_1, m_2 \in \mathbb{Z}^+$, such that

$$p_1 = m_1 n + r \quad \text{and} \quad p_2 = m_2 n + r.$$

[3] Often, a clever choice can simplify the study of the particular problem appreciably. A concrete example of this is the gauge symmetry of electromagnetism.

We can check that, indeed, this satisfies the requirements of Def. 3. The equivalence classes of this equivalence relation are labelled by $r \in \{0, 1, \cdots, n-1\}$ and denoted by $\bar{r}$. The set of all the equivalence classes

$$\mathbb{Z}_n := \{\bar{0}, \bar{1}, \cdots, \overline{n-1}\}$$

is called *integers modulo n*. For example, let us choose $n = 2$. Then, there are two equivalence classes $\bar{0}$ and $\bar{1}$. These are known as *even* and *odd* numbers.

Next, let us consider the collection of all finite sets with no further structures on them. Two finite sets are distinguished by the "number" of their elements. In other words, two sets are said to be *equivalent* if they have the same "number" of elements. This is obviously an equivalence relation. It divides the collection of all finite sets into equivalence classes of sets with the same number of elements. The same definition is rather unsatisfactory for infinite sets since the notion of the number of elements is not well defined. A simple generalization of this notion, however, works perfectly well.

Definition 4: Let X and Y be two sets and $f : X \to Y$ be a function (map), i.e., f assigns to each element of X one and only one element of Y. The subset

$$f(X) := \{y \in Y \ : \ y = f(x) \text{ for some } x \in X\} \subseteq Y$$

is called the *image* of X under f. In some cases $f(X)$ is identical to Y but not always. If $f(X) = Y$, then f is called an *onto* or *surjective* function. Let Y_1 be a subset of Y. Then the subset

$$f^{-1}(Y_1) := \{x \in X \ : \ f(x) \in Y_1\} \subseteq X$$

is called the *preimage* or *inverse image* of Y_1 under f. Inverse images of subsets of Y are subsets of X. The inverse image of a subset Y_1 of Y which includes only a single point, i.e., $Y_1 = \{y\}$, is called the inverse image of that point, $y \in Y$. It may happen that the inverse image of a point $y \in Y$ is empty, $y \notin f(X)$, or that it consists of many elements. If the inverse images of all the points of Y have at most a single element, then f is called a *one-to-one* or *injective* function. This simply means that every $y \in f(X)$ is the image of a single point $x \in X$. The condition of one-to-oneness is the necessary and sufficient condition for the existence of the inverse function

$$f^{-1} : f(X) \subseteq Y \longrightarrow X$$

of f. The function f is said to be a **bijection** or a **one-to-one correspondence**, if it is both onto and one-to-one. Two sets X and Y are said to be *bijective* if there exists a bijection f between them.[4]

The relation of being bijective for arbitrary sets is the appropriate generalization of having the same number of elements for finite sets. We can easily

[4] A bijection is also called a bijective function.

show that bijective finite sets have the same number of elements. Moreover, we can check that the relation defined in this way is an equivalence relation, i.e., it satisfies all the necessary requirements of Def. 3. The utility of this equivalence relation is in the classification of all point sets. In fact, in set theory, one does not distinguish between bijective sets.

Let us return to our discussion of metric spaces where the issues of the identification and classification of mathematical structures were raised. Naturally, two "equivalent" metric spaces must be necessarily "equivalent" as sets. Therefore, the notion of equivalence is again linked to the existence of certain functions between metric spaces. Since a metric space has an additional metric structure besides the point set structure, the notion of equivalence of metric spaces is a refinement of that of point sets.

Definition 5: Let (X_1, d_1) and (X_2, d_2) be two metric spaces and $f : X_1 \to X_2$ be a bijection. f is said to be an **isometry** if f preserves the metric structures, d_1 and d_2.[5] In more precise language, for all $p_1, q_1 \in X_1$ and $p_2, q_2 \in X_2$

$$d_1(p_1, q_1) = d_2(f(p_1), f(q_1)) \,.$$

Two metric spaces are said to be **isometric** if there exists an isometry between them.

Once more, the relation of isometry satisfies the axioms of an equivalence relation and it divides the collection of all metric spaces into distinct classes of isometric metric spaces.

We saw that the notions of open and closed subsets can be easily defined for metric spaces. Let us collect all the open sets of a metric space and then consider this collection without any reference to the metric structure on the universal set. There are important properties of this derived structure which can be defined and analyzed regardless of the details of the metric structure, i.e., the distance function. One of these properties is related to the existence of "continuous" functions.

The definition of a continuous function between two metric spaces is almost identical to the one presented in elementary calculus texts for the continuity of a function of a real variable.

Definition 6: Let (X_1, d_1) and (X_2, d_2) be two metric spaces. A function $f : X_1 \to X_2$ is said to be **continuous** at a point $p_1 \in X_1$ if for every $\epsilon > 0$, there is a $\delta > 0$ such that

$$d_1(p_1, x_1) < \delta \stackrel{\text{implies}}{\Longrightarrow} d_2(f(p_1), f(x_1)) < \epsilon,$$

for all such $x_1 \in X$. Alternatively, f is continuous at p_1, if for every neighborhood $N_\epsilon(f(p_1)) \subset X_2$ of the point $f(p_1)$ there is a neighborhood $N_\delta(p_1)$ of p_1 such that

$$f(N_\delta(p_1)) \subset N_\epsilon(f(p_1)).$$

---

[5] It is not difficult to see that if f preserves the metric structure, so does its inverse.

This definition is a little abstract. The reader is advised to draw the graph of a simple function of a real variable and examine the utility of the above definition in practice.

An important characterization of continuous functions is the following result.

Proposition 1: A function $f : X_1 \to X_2$ is continuous if and only if the inverse image of every open subset of X_2 is open in X_1.

A simple but rather instructive consequence of this result is that the notion of continuity does not directly depend on the particular metric structures of the corresponding metric spaces. For example, let us assume that (X_1, d_1') is another metric structure on X_1 such that both distance functions, d_1 and d_1', define the same collection of open subsets in X_1. This is to say that every open set in (X_1, d_1) is also open in (X_1, d_1') and vice versa. Then, the continuity of a function f will not depend on which metric function we choose on X_1. The same is true for X_2. The concept of continuity, therefore, is only sensitive to the collection of open sets of the two spaces.

Let us consider the following two choices of metric function on $\mathbb{R}^2$:

$$\begin{aligned} d_1(\mathbf{x}, \mathbf{y}) &:= \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)^2} \\ d_1'(\mathbf{x}, \mathbf{y}) &:= |x^1 - y^1| + |x^2 - y^2|. \end{aligned}$$

It is easy to observe that both d_1 and d_1' satisfy the axioms of a metric function. Further, they define the same notion of open subsets and consequently continuity in $\mathbb{R}^2$.

As we argued in the preceding paragraphs, the continuity of a function does not depend on the details of the metric structure. This triggers the question of identifying the minimal structure on a point set that allows for the concept of continuity to be defined. Evidently, such a structure will be more general than a metric structure and less trivial than a plain set structure. This structure is called a *topological structure* or simply a *topology* on a set.

Definition 7: Let X be a set and $\mathcal{T}$ be a family of subsets of X such that

1) The empty set $\emptyset$ and the universal set X belong to $\mathcal{T}$.
2) If O_1 and O_2 belong to $\mathcal{T}$, then so does $O_1 \cap O_2$, i.e., the intersection of a finite number of elements of $\mathcal{T}$ is also an element of $\mathcal{T}$.
3) The union of any finite or infinite number of elements of $\mathcal{T}$ is also an element of $\mathcal{T}$.

Then, the pair $(X, \mathcal{T})$ is said to be a **topological space**. The collection $\mathcal{T}$ is called a **topology** on X. The elements of $\mathcal{T}$ are called the *open subsets* of X.

In other words, a topology on a set is an assignment of the word "open" to a collection of its subsets that possess certain properties of open subsets of say metric spaces. If the set is endowed with a metric structure, then the open subsets defined by the metric satisfy the requirements of a topological space. Thus any metric space has a canonical topological structure which is called the *metric topology*. However, there is no unique topology on a given set (unless the set is empty). In particular, a metric space may be given a non-metric topology.

The notion of *continuity* can be easily defined for functions between topological spaces. This can be done either by first defining the open neighborhoods and then using a similar definition to the one presented for metric spaces, namely Def. 6, or by promoting Proposition 1 to a definition.

Definition 8: Let $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$ be two topological spaces and $f : X_1 \to X_2$ be a function. Then, f is said to be continuous if the inverse image of every open subset of X_2 (every element of $\mathcal{T}_2$) is an open subset of X_1 (an element of $\mathcal{T}_1$).

Having defined topological spaces, we shall next try to define an appropriate concept of "equivalence" for topological spaces. This is done in complete analogy to the cases of point sets and metric spaces. Again the equivalence of topological spaces must reduce to that of total or universal sets. Thus, we need a bijection that preserves the topological structures. Such a function is called a *homeomorphism*.

Definition 9: A function $f : X_1 \to X_2$ between two topological spaces $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$ is said to be a **homeomorphism**, if it is a continuous bijection with a continuous inverse. If there exists a homeomorphism between two topological spaces, they are called **homeomorphic**.

The existence of a homeomorphism between two topological spaces defines an equivalence relation. This equivalence relation divides the collection of all topological spaces into equivalence classes of homeomorphic topological spaces. The members of each class share the same *topological properties*. These are properties that are defined using the notion of open subsets. An intuitively simple example of a topological property is *connectedness*.

Definition 10: A topological space $(X, \mathcal{T})$ is said to be **disconnected** if there are two open subsets O_1 and O_2 such that $X = O_1 \cup O_2$ and $O_1 \cap O_2 = \emptyset$. If a topological space is not disconnected, it is said to be **connected**.

We can show that under a homeomorphism a connected topological space is mapped to another connected topological space. Alternatively, there is no homeomorphism between a connected and a disconnected topological space. Hence, connectedness is a *topological property*.

We saw that in a topological space the unions of open sets are also open. This simple property suggests a practical way of generating all the open subsets as the unions of some "more basic" ones. The collection of these basic open sets is called a *basis* of the topological space. More precisely, a subfamily $\mathcal{B}$ of a topology $\mathcal{T}$ is called a *basis* if all the elements of $\mathcal{T}$, i.e., all the open subsets, are obtained as the unions of the elements of $\mathcal{B}$.

There are other important collections of open subsets of a topological space $(X, \mathcal{T})$, subfamilies of $\mathcal{T}$. For example, let O be a subset and consider a family of open subsets $\mathcal{C} = \{O_\alpha\}$, i.e. a subfamily of $\mathcal{T}$, such that $O \subseteq \cup_\alpha O_\alpha$. Then, $\mathcal{C}$ is called an *open covering* of O. Certainly, a basis of $\mathcal{T}$ is an open covering of every subset of X. The converse is certainly not true. A subset $\mathcal{C}'$ of an open covering $\mathcal{C}$ is naturally called a subcovering.

Definition 11: A topological space $(X, \mathcal{T})$ is said to be **compact** if every (infinite) open covering of X has a finite subcovering.

Compactness is also a topological property, i.e., under a homeomorphism a compact topological space is mapped to another compact topological space. Compactness plays a very substantial role in the study of a special class of topological spaces, called manifolds, and particularly Lie groups. We shall give a more intuitive definition of compactness which is valid for manifolds in Sect. A.2.

A simple example of a topological space is the space $\mathbb{R}^n$ with a metric topology. Usually, we choose the Euclidean metric,

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2},$$

to define the open neighborhoods and hence the open subsets. A simple basis for this topological space is given by the subfamily of all open balls:

$$\mathcal{B} = \{N_r(\mathbf{x}) \ : \ \mathbf{x} \in \mathbb{R}^n \ , \ r \in \mathbb{R}^+\}.$$

Now, let us consider the following family of open subsets:

$$\mathcal{C} = \{N_2(\mathbf{n}) \ : \ \mathbf{n} \in \mathbb{Z}^n \subset \mathbb{R}^n\}.$$

$\mathcal{C}$ is an open covering of $\mathbb{R}^n$. It is however not a basis. One can easily show that none of these coverings has a finite subcovering. Hence $\mathbb{R}^n$ with the Euclidean metric topology is not compact.

There are other topological structures on the set $\mathbb{R}^n$. In fact, there is an infinite number of them. Two rather trivial examples of non-metric topologies on $\mathbb{R}^n$ are

1) $\mathcal{T}_0 := \{\emptyset, \mathbb{R}^n\}$, i.e., the only open subsets are the empty set and the total space.
2) $\mathcal{T}_{\text{discrete}} := \{O \ : \ O \subseteq \mathbb{R}^n\}$, i.e., all the subsets are open.

Similar topologies can be given to any other point set X. They are known as the *trivial topology* and the *discrete topology* on X, respectively.

We shall always assume that $\mathbb{R}^n$ is endowed with the Euclidean metric topology.

Another useful fact about topological spaces is that we can induce a topology on a subset X_1 of the universal set X. This is simply done by defining the open subsets of X_1 to be the intersections of the open subsets of X and X_1. This topology is called the *subspace topology.* This allows us, for example, to speak of a homeomorphism between the subsets of two topological spaces.

As for point sets, we can define the Cartesian product of two topological spaces $(X, \mathcal{T})$ and $(X', \mathcal{T}')$. The result is called the *product topology*. It is naturally defined to be a topology on the Cartesian product $X \times X'$, whose elements (open subsets) are the Cartesian products of the elements of $\mathcal{T}$ and $\mathcal{T}'$ (open subsets of X and X'). An example of a product topology is the metric topology on $\mathbb{R}^{m+n}$. This is the product of the metric topologies on $\mathbb{R}^m$ and $\mathbb{R}^n$. Therefore, the metric topology on $\mathbb{R}$ generates $\mathbb{R}^m$, for all $m \in \mathbb{Z}^+$, as its products.

So far, we have introduced many mathematical concepts that are usually not familiar to non-mathematicians. We shall need these concepts to arrive at a fairly precise definition of a *manifold*. Our list of related topological concepts is however far from being complete. We wish to refer the interested reader to textbooks on topology, such as

- J. G. Hocking and G. S. Young: *Topology* (Dover Publications Inc., New York, 1988);
- G. F. Simmons: *Topology and Modern Analysis* (R. E. Krieger Publishing Company, Malabar, Florida, 1983).

A.2 Differentiable Manifolds

Throughout the development of geometry and topology the space $\mathbb{R}^m$ has served mathematicians as a source of key ideas and properties. Many of these ideas and properties could be generalized to define various abstract mathematical structures which are occasionally used to formulate and solve concrete physical problems. An important class of such structures that are extensively used in theoretical physics is the class of the so-called *manifolds*. Manifolds are certain topological spaces which are obtained by patching together open pieces of $\mathbb{R}^m$. In other words, a manifold is the union of a number of open subsets each of which is homeomorphic to (an open subset) of $\mathbb{R}^m$. These open subsets are called *coordinate patches* or *charts*. On each of these patches one can set up a coordinate system. This is done by mapping the points of the patch into $\mathbb{R}^m$ using the corresponding homeomorphism. In this way, one translates the local properties of a manifold into those of $\mathbb{R}^m$ and employs the knowledge of $\mathbb{R}^m$ to develop calculus, analysis, and even geometry. The main complication arises from the fact that the results obtained in one patch are only valid locally. Thus, it is necessary to check the validity of local results when applied to the whole space. On the other hand, since the choice of the coordinate charts is not unique, the physical results of any computation must be independent of this choice.

Definition 12: Let M be a topological space with a countable basis.[6] M is said to be a **topological manifold** of dimension m, if there exists an

[6] This means that the elements of the basis can be labelled by integers.

open covering $\{O_\alpha\}$ of M such that each O_α is homeomorphic to (some open subset of) $\mathbb{R}^m$, for a fixed m. The homeomorphisms

$$\phi_\alpha : O_\alpha \longrightarrow \phi_\alpha(O_\alpha) \subseteq \mathbb{R}^m,$$

together with the open subsets O_α are called the **charts** of the manifold. A complete collection of charts is called an **atlas**. For every pair of intersecting charts, e.g., $O_\alpha \cap O_\beta \neq \emptyset$, the functions

$$g_{\alpha\beta} := \phi_\alpha \circ \phi_\beta^{-1}\Big|_{O_\alpha \cap O_\beta} : \phi_\beta(O_\alpha \cap O_\beta) \longrightarrow \phi_\alpha(O_\alpha \cap O_\beta)$$

are homeomorphisms between open subsets of $\mathbb{R}^m$. These functions are called the **transition** or **overlap functions**.

The topology of a manifold, i.e., the collection of all its open subsets, can be recovered from the (Euclidean metric) topology of $\mathbb{R}^m$. In fact, all the open subsets of a manifold are unions of images of open subsets of $\mathbb{R}^m$ under the ϕ_α^{-1}'s. Moreover, since all m-dimensional manifolds are locally "identical", the global or topological properties of a manifold depend on the way these open subsets are patched or glued together. This is done by the transition functions. Therefore, the topology of a manifold is determined by its transition functions.

Definition 13: Let $U \subseteq \mathbb{R}^{m_1}$ and $V \subseteq \mathbb{R}^{m_2}$. Then, a function $g : U \to V$ is said to be a C^N, $0 < N \leq \infty$, function if g is N-times differentiable.[7] A C^∞ function is also called a **smooth function**. g is called a **diffeomorphism** if it is a differentiable homeomorphism with a differentiable inverse. Similarly, one defines C^N diffeomorphisms by requiring a diffeomorphism and its inverse to be C^N. If the transition functions $g_{\alpha\beta}$ are C^N diffeomorphisms, then the manifold is called a C^N manifold. In particular, C^1 and C^∞ manifolds are called **differentiable** and **smooth manifolds**, respectively. Throughout this book all the manifolds are assumed to be smooth.

The homeomorphisms ϕ_α and their inverses ϕ_α^{-1} enable us to treat the points of O_α as those of a subset of $\mathbb{R}^m$. Clearly, the space $\mathbb{R}^m$ is itself a manifold. It can be covered by a single chart. If the minimum number of charts that cover a manifold is more than one, then the manifold has a different topology than $\mathbb{R}^m$. A simple example of such a manifold is the two-dimensional sphere, S^2. We shall examine the manifold structure of S^2 in detail.

As the reader must have noticed, the concept of a manifold is quite abstract. A practical way of thinking about manifolds is to think about them as the generalizations of surfaces. In fact, the simplest, and at the same time interesting, examples of a manifold are two-dimensional surfaces such as the sphere and torus. These examples are easily understood, for they can be visualized as sitting inside $\mathbb{R}^3$. In mathematical language, one says that these

[7] Occasionally, the symbol C^0 is used for the set of *continuous* functions.

manifolds are *submanifolds* of $\mathbb{R}^3$. A submanifold is a manifold which is a subset of another manifold. Its charts are the restrictions of those of the "bigger" manifold. If M' is a submanifold of M, then the charts $(O'_\alpha, \phi'_\alpha)$ of M' are given by

$$O'_\alpha := M' \cap O_\alpha, \ \phi'_\alpha := \phi_\alpha|_{O'_\alpha} .$$

A substantial result of differential geometry is the *Whitney Embedding Theorem.* This theorem says that every manifold is a submanifold of $\mathbb{R}^d$ for some $d \in \mathbb{Z}^+$. Thus, we can always think of manifolds as direct generalizations of two-dimensional surfaces to higher dimensions. Nevertheless, we must keep in mind that manifolds are well-defined mathematical objects by themselves. They can be studied independently of their relation to an embedding space.

As we mentioned earlier, the sphere S^2 is an example of a "non-trivial" smooth manifold.[8] It is usually thought of as the submanifold of $\mathbb{R}^3$ defined by the set of unit vectors,

$$\{\mathbf{x} \in \mathbb{R}^3 \ : \ |\mathbf{x}| = 1\} \subset \mathbb{R}^3.$$

This is called the *round sphere.* The round sphere has more structure than a smooth manifold. Specifically, it inherits a "geometric" or "metric structure" from $\mathbb{R}^3$. This brings us to the problem of the classification and the necessity of a definition of "equivalence" for smooth manifolds. We know that a manifold is a topological space. Thus, the concept of equivalence should be defined via certain homeomorphisms between two manifolds so that the equivalent manifolds have the same topological structures. Furthermore, this homeomorphism must reflect the information about the smoothness and the chart structures. Such a function is also called a *diffeomorphism.*

Definition 14: Let M_1 and M_2 be two differentiable (C^N) manifolds whose charts are given by $(O_{\alpha_1}, \phi_{\alpha_1})$ and $(O_{\alpha_2}, \phi_{\alpha_2})$. Then, any function $f : M_1 \to M_2$, can be defined by its restrictions:

$$f_{\alpha_{1,2}} := \phi_{\alpha_2} \circ f \circ \phi_{\alpha_1}^{-1} : \phi_{\alpha_1}(O_{\alpha_1}) \longrightarrow \phi_{\alpha_2}(O_{\alpha_2}).$$

These are functions from $\mathbb{R}^{m_1}$ to $\mathbb{R}^{m_2}$, where m_1 and m_2 are the dimensions of M_1 and M_2, respectively. If all $f_{\alpha_{1,2}}$ are differentiable (C^N) functions then the function f is said to be differentiable (C^N). Similarly, a homeomorphism $f : M_1 \to M_2$ is said to be a (C^N) **diffeomorphism of manifolds**, if the corresponding functions $f_{\alpha_{1,2}}$ are (C^N) diffeomorphisms.[9]

The notion of (C^N) *diffeomorphy* defines an equivalence relation. This relation divides the collection of all (C^N) differentiable manifolds into the equivalence classes of (C^N) diffeomorphic manifolds. The elements of each class are treated as identical manifolds. In this respect, the round sphere is a

---

[8] A trivial manifold means that it is topologically equivalent (homeomorphic) to $\mathbb{R}^m$ for some $m \in \mathbb{Z}^+$. In other words, a manifold is called trivial if it can be covered by a single chart.

[9] Note that we have already defined the notion of a (C^N) diffeomorphism for $\mathbb{R}^m$.

representative of an infinite class of smooth manifolds which are diffeomorphic to one another. For example, the surface of an ellipsoid can be obtained by a smooth deformation of the round sphere and vice versa. Thus, it belongs to the same diffeomorphy class and can represent the sphere S^2 equally well.

Let us examine the manifold structure of S^2. As we mentioned, S^2 is (topologically) different from $\mathbb{R}^2$. In fact, we need at least two charts to cover S^2. A practical choice of coordinates for the points of S^2 is the spherical coordinates (θ, φ). The fact that the global (θ, φ) coordinate system is ill defined at $\theta = 0$ and $\theta = \pi$, is an indication of the necessity for at least two different coordinate charts. A rather standard set of coordinate charts for S^2 is obtained by what is known as the *stereographic projection* of S^2 on $\mathbb{R}^2$. The corresponding open covering consists of two open subsets:

$$O_1 := S^2 - \{\mathcal{S}\} \text{ and } O_2 := S^2 - \{\mathcal{N}\}.$$

Here $\mathcal{S}$ and $\mathcal{N}$ denote the south and the north poles of S^2, respectively. The homeomorphisms

$$\phi_i : O_i \longrightarrow \mathbb{R}^2, \quad i = 1, 2,$$

are defined by the following projective procedure.

Consider the tangent plane P_1 to S^2 at the north pole, $\mathcal{N}$. Let R be an arbitrary point of O_1 and draw a line through R and the south pole, $\mathcal{S}$ (Fig. A.1). This line intersects P_1 at a point p_1. The map ϕ_1 is defined by

$$\phi_1(R) = p_1.$$

By interchanging the roles of the north and the south poles we define ϕ_2 similarly. It is clear that ϕ_1 and ϕ_2 are homeomorphisms of a punctured sphere to $\mathbb{R}^2$.

Using ϕ_1 and ϕ_2, we can treat the points of S^2 as those of P_1 or P_2. On each of these planes, we can set up a coordinate system and use our knowledge

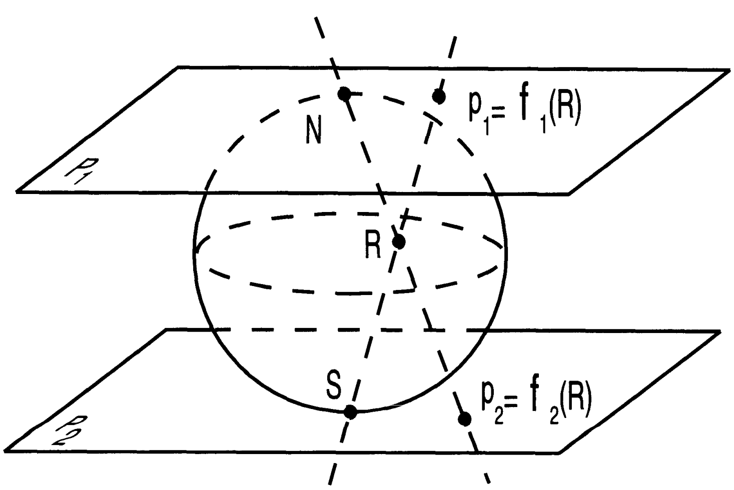

Fig. A.1. Stereographic projection of sphere.

of $\mathbb{R}^2$ to do calculations. Let us choose the usual Cartesian coordinates on P_1 and P_2,

$$\begin{aligned} p_1 \in P_1 : \quad p_1 &\equiv (x_1, y_1) \\ p_2 \in P_2 : \quad p_2 &\equiv (x_2, y_2). \end{aligned} \tag{A.1}$$

We can use the Euclidean geometry of $\mathbb{R}^3$ to obtain an explicit formula for the transition function,

$$g_{21} := \phi_2 \circ \phi_1^{-1}.$$

Let us choose the x_1 and y_1 axes in P_1 to be parallel to the x_2 and y_2 axes in P_2. On the intersection of O_1 and O_2, i.e., for any R different from the north and the south poles, we have

$$\begin{aligned} (x_2, y_2) \equiv p_2 &= \phi_2(R) \\ &= \phi_2\left(\phi_1^{-1}(p_1)\right) \\ &= \left(\phi_2 \circ \phi_1^{-1}\right)(x_1, y_1) \\ &=: g_{21}(x_1, y_1). \end{aligned}$$

Then, a simple calculation shows that

$$g_{21}(x_1, y_1) = \left(\frac{x_1}{x_1^2 + y_1^2}, \frac{-y_1}{x_1^2 + y_1^2} \right). \tag{A.2}$$

Note that on $O_1 \cap O_2$, $x_1^2 + y_1^2 \neq 0$. Thus, the transition function $g_{21} : \mathbb{R}^2 - \{0\} \to \mathbb{R}^2 - \{0\}$ is a well-defined smooth diffeomorphism, and consequently S^2 is a smooth manifold.

The Cartesian coordinates (x_i, y_i), $i = 1, 2$, have the disadvantage that they do not reflect the desirable symmetries of S^2. The natural choice of a coordinate system that inherits these symmetries is the spherical coordinates (θ, φ). We pointed out that this coordinate system cannot be used globally as one uses the spherical coordinates (r, θ, φ) on R^3. A compromise can be reached, however, by adopting two sets of spherical coordinates to represent the points of O_1 and O_2. Denoting these by (θ_1, φ_1) and (θ_2, φ_2), respectively, we find the following simple expression for the transition function:

$$(\theta_2, \varphi_2) = \tilde{g}_{21}(\theta_1, \varphi_1) = (\pi - \theta_1, \varphi_1).$$

This choice has its ambiguities at the poles. Nevertheless, these turn out to be unimportant. The situation is analogous to the use of the polar coordinates (r, φ) on $\mathbb{R}^2$. At $r = 0$, φ is not well defined. However, this does not prevent us from using polar coordinates on $\mathbb{R}^2$.

If we use complex coordinates in (A.1), namely

$$p_i \equiv w_i := x_i + i y_i, \qquad i = 1, 2,$$

then the expression (A.2) for the transition function takes the particularly simple form

$$w_2 = g_{21}(w_1) = \frac{1}{w_1}. \tag{A.3}$$

As a function of a complex variable $g_{21} : \mathbb{C} - \{0\} \to \mathbb{C} - \{0\}$ is a complex analytic (holomorphic) function. This makes S^2 an example of a *complex (holomorphic) manifold.* Briefly, a complex manifold is an even-dimensional manifold that is locally homeomorphic to $\mathbb{C}^m$ and whose transition functions are analytic functions from $\mathbb{C}^m$ to $\mathbb{C}^m$. m is called the complex dimension of the manifold. Its real dimension is $2m$.

S^2 is a member of an important class of smooth (complex) manifolds called *complex projective spaces.*

Definition 15: Consider the space of $(N+1)$-tuple complex numbers, $\mathbb{C}^{N+1}$. As a point set the **complex projective space** $\mathbb{C}P(N)$ is the set of all complex lines in $\mathbb{C}^{N+1}$ that pass through the origin. These lines are also called "rays". Each ray is represented by a non-zero complex vector, $\mathbf{z} \in \mathbb{C}^{N+1}$, via

$$l = [z] := \{\lambda \mathbf{z} \,:\, \lambda \in \mathbb{C} - \{0\}\}. \tag{A.4}$$

In fact, as seen in (A.4), $\mathbf{z}$ can be chosen to have unit length (norm). The length or the norm of $\mathbf{z}$ is defined by

$$\| \mathbf{z} \| := \sqrt{|z^1|^2 + \cdots + |z^{N+1}|^2},$$

where z^i are the complex components of $\mathbf{z}$ and $|z^i|$ are their moduli.[10] The set of all unit vectors in $\mathbb{C}^{N+1}$ defines the unit sphere S^{2N+1},

$$S^{2N+1} := \left\{\mathbf{z} \in \mathbb{C}^{N+1} \,:\, \| \mathbf{z} \| = 1 \right\} \subset \mathbb{R}^{2N+2} = \mathbb{C}^{N+1}.$$

However, the rays are still not in one-to-one correspondence with the points of S^{2N+1}. This is because we can represent l with another unit vector $\mathbf{z}' := w\mathbf{z}$, where $w \in \mathbb{C}$ is of unit modulus. Clearly, $l = [z] = [z']$ and $\mathbf{z}' \in S^{2N+1}$, but $\mathbf{z} \neq \mathbf{z}'$. This suggests that although $\mathbb{C}P(N) \neq S^{2N+1}$, it may be viewed as a set of equivalence classes of points of S^{2N+1}. The desired equivalence relation is given by

$$\mathbf{z} \sim \mathbf{z}' \text{ iff there is } w \in \mathbb{C} \text{ with } \mathbf{z}' = w\mathbf{z}.$$

The equivalence class of $\mathbf{z} \in S^{2N+1}$ is denoted by $[z]$ as defined in (A.4). These equivalence classes are precisely the points of $\mathbb{C}P(N)$.

There is a standard choice of coordinate charts for $\mathbb{C}P(N)$. These are called *homogeneous coordinate* charts. They are naturally induced from $\mathbb{C}^{N+1} - \{0\}$. Let us use the notation z for the coordinates of $\mathbf{z}$ in $\mathbb{C}^{N+1}$, i.e., $z := (z^1, \cdots, z^{N+1})$. Consider the $N+1$ open sets in $\mathbb{C}P(N)$ defined by

$$O_i := \{[z] \in \mathbb{C}P(N) \,:\, z^i \neq 0\}, \qquad i = 1, 2, \cdots N+1.$$

[10] If $z^i = x^i + iy^i$, the modulus of z^i is given by $|z^i| = \sqrt{(x^i)^2 + (y^i)^2}$.

Clearly every point in $\mathbb{C}P(N)$ is included in at least one of these open subsets. Thus, $\{O_i\}_{i=1,\cdots N+1}$ forms an open covering of $\mathbb{C}P(N)$. The homogeneous coordinate charts are given by (O_i, ϕ_i), where for every $[z] \in O_i$

$$\phi_i([z]) = \left(\frac{z^1}{z^i}, \cdots, \frac{z^{i-1}}{z^i}, 1, \frac{z^{i+1}}{z^i}, \cdots, \frac{z^{N+1}}{z^i}\right),$$

or more correctly

$$\phi_i([z]) := \left(\frac{z^1}{z^i}, \cdots, \frac{z^{i-1}}{z^i}, \frac{z^{i+1}}{z^i}, \cdots, \frac{z^{N+1}}{z^i}\right) \in \mathbb{C}^N.$$

We shall next obtain the transition functions. In order to do this, we compare the coordinates of a point associated with two different coordinate charts. Let $[z] \in \mathbb{C}P(N)$ such that $z^i \neq 0$ and $z^j \neq 0$ for some $i \neq j$. Then $[z] \in O_i \cap O_j$, and we have

$$\phi_j([z]) = g_{ji}\left(\phi_i([z])\right).$$

Since z^i and z^j are both non-zero, we have

$$\begin{aligned} &\left(\frac{z^1}{z^i}, \cdots, \frac{z^{i-1}}{z^i}, 1, \frac{z^{i+1}}{z^i}, \cdots, \frac{z^{N+1}}{z^i}\right) \equiv [z] \\ &\equiv \left(\frac{z^1}{z^j}, \cdots, \frac{z^{j-1}}{z^j}, 1, \frac{z^{j+1}}{z^j}, \cdots, \frac{z^{N+1}}{z^j}\right). \end{aligned}$$

The action of the transition function g_{ji} on an arbitrary element $w = (w^1, \cdots, w^N) = \phi_i([z]) \in \mathbb{C}^N$ is described by the following steps:

1) Identify $(w^1, \cdots, w^N)$ with $(w^1, \cdots, w^{i-1}, 1, w^i, \cdots, w^N)$.

2) Multiply this $(N+1)$-tuple by $\frac{z^i}{z^j}$, to obtain

$$\left(\frac{w^1 z^i}{z^j}, \cdots, \frac{w^{i-1} z^i}{z^j}, \frac{z^i}{z^j}, \frac{w^i z^i}{z^j}, \cdots, \frac{w^N z^i}{z^j}\right). \tag{A.5}$$

3) Since $w = \phi_i([z])$,

$$w^k = \begin{cases} \frac{z^k}{z^i} & \text{for } k = 1, \cdots, i-1 \\ \frac{z^{k+1}}{z^i} & \text{for } \quad k = i, \cdots, N. \end{cases}$$

This implies that one of the components displayed in (A.5) is identically 1. In fact, if we assume that $i < j$, $w^{j-1} = \frac{z^j}{z^i}$, that is, $\frac{w^{j-1} z^i}{z^j} = 1$. This allows us to undo the first step by dropping this 1 from (A.5) and obtain

$$\left(\frac{w^1 z^i}{z^j}, \cdots, \frac{w^{i-1} z^i}{z^j}, \frac{z^i}{z^j}, \frac{w^i z^i}{z^j}, \cdots, \frac{w^{j-2} z^i}{z^j}, \frac{w^j z^i}{z^j}, \cdots, \frac{w^N z^i}{z^j}\right) := \tilde{w}.$$

This is an N-tuple of complex numbers that we denote by $\tilde{w}$. For $i > j$, we obtain another N-tuple similarly.

4) The transition function g_{ji} is given by

$$g_{ji}(w) := \tilde{w}.$$

We could have associated each step of this list with a function and define the transition functions as the composition of these functions. Clearly, the transition functions are (complex) analytic.

Let us look at the case of $N = 1$. There are two coordinate charts:

$$O_1 = \{[(z^1, z^2)] \ : \ z^1 \neq 0\} \ \text{ and } \ O_2 = \{[(z^1, z^2)] \ : \ z^2 \neq 0\}.$$

On $O_1 \cap O_2$, both z^1 and z^2 are non-zero and

$$\phi_1\left([(z^1, z^2)]\right) = \frac{z^2}{z^1} \quad \text{and} \quad \phi_2\left([(z^1, z^2)]\right) = \frac{z^1}{z^2}.$$

Following the above procedure, we denote $\frac{z^2}{z^1}$ by w. Then,

$$g_{21}(w) = \frac{1}{w}.$$

This is exactly the expression given in (A.3). Thus, in view of the fact that the structure of a manifold is determined by its transition functions, we have the following identity

$$\mathbb{C}P(1) \simeq S^2.$$

The symbol $\simeq$ stands for the word "diffeomorphic."

We have given many examples of smooth manifolds of dimension two or higher. As for one-dimensional manifolds, there are two possibilities. These are the real line $\mathbb{R}$ and the circle S^1. In fact, a more correct statement is that these are the only *connected* one-dimensional manifolds that do not have a boundary. The adjective "connected" refers to the same property as defined for topological spaces: Connectedness is a topological property.

Another basic concept that generalizes to the discussion of manifolds is that of taking products of manifolds. The resulting objects, which are themselves manifolds, are called *product manifolds.* As a topological space, a product manifold, $M = M_1 \times M_2$, has the product topology. The coordinate charts of M are obtained as the products of the coordinate charts of M_1 and M_2. In particular, if M_1 and M_2 are of dimension m_1 and m_2, then M is (m_1+m_2)-dimensional. Some of the higher-dimensional manifolds are product manifolds. A typical example of this is the torus, T^2, that is the product of two circles,

$$T^2 \simeq S^1 \times S^1.$$

More trivial examples of product manifolds are

$$\mathbb{R}^m \simeq \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{m\text{-times}}.$$

We have also higher dimensional tori,

$$T^m := \underbrace{S^1 \times \cdots \times S^1}_{m\text{-times}}.$$

The examples we have considered are all connected manifolds. The disconnected manifolds are usually not quite as interesting. They are essentially a collection of two or more copies of connected manifolds. An example of a one-dimensional disconnected manifold is $\{-1, 1\} \times S^1$. Here, we view the finite set $\{-1, 1\}$ as a zero-dimensional manifold. In fact, every finite set can be seen as a submanifold[11] of $\mathbb{R}$. In particular, $\{-1, 1\}$ is also called the *zero-dimensional circle*, S^0.

We have discussed the importance of the notion of a diffeomorphism in some detail. The action of a diffeomorphism can be represented by its effect on the transition functions. In fact, a diffeomorphism takes one set of smooth transition functions into another. What remains unchanged is a little more than the topological structure of the manifold. There are cases in which we can find a function between two manifolds which is a homeomorphism but not a diffeomorphism. In this case, we say that the two manifolds have identical topological structures but different *differential (smooth) structures.*[12]

The differentiability or smoothness requirement enables us to introduce the notion of a *tangent space.* This is an essential step towards defining analysis and geometry on a smooth manifold.

Tangent spaces of a manifold are direct generalizations of the tangent spaces of a two-dimensional surface in $\mathbb{R}^3$. They can be defined without any reference to the embedding of the manifold into $\mathbb{R}^d$. However, we would like to give a "geometric" definition of the tangent space at a point of a manifold. This approach makes use of such an embedding. We shall first give the definition of a *curve* on a manifold.

Definition 16: Let M be an m-dimensional smooth manifold. Then, any smooth function $C : [0, T] \to M$ is called a smooth **curve** in M.[13]

A smooth curve C on M can be viewed as a smooth curve in a Euclidean space $\mathbb{R}^d$ by embedding M into $\mathbb{R}^d$, where d is a sufficiently large positive integer.

Definition 17: Let M be a smooth manifold and $C : [0, T] \to M$ be a smooth curve on M with $C(0) = p \in M$. Choose a Euclidean space $\mathbb{R}^d$ such that M is a submanifold of $\mathbb{R}^d$. The vector $v := \frac{d}{dt}C(t)\big|_{t=0} \in \mathbb{R}^d$ is called the *tangent vector* to $C \subset \mathbb{R}^d$ at $t = 0$. It is also called the **tangent vector** to $C \subset M$ at $p \in M$. The set of all tangent vectors to all curves in M that

[11] The topology given to a finite set is the subset topology which is, in this case, the same as the discrete topology. This means that every subset of a finite set is postulated to be open.

[12] For a precise definition of a differential structure of a manifold see [186].

[13] Consider $[0, T]$ as a submanifold of $\mathbb{R}$. Then, C is a function between two smooth manifolds. The notion of smoothness for such a function is defined earlier.

originate at $p \in M$ forms an m-dimensional vector space ($\cong \mathbb{R}^m$).[14] This vector space is called the **tangent space** of M at p and denoted by T_pM. The dual (vector) space of T_pM – the space of all linear real-valued functions on T_pM – is called the **cotangent space** of M at p. It is denoted by T_pM^*.

The elements of a cotangent space are called *cotangent vectors* or simply *covectors*. Another terminology is to call the tangent and cotangent vectors, *contravariant* and *covariant* vectors, respectively.

Let us consider the set of all the tangent spaces of a manifold. This set inherits the structure of a smooth manifold from M. It has the special property that the tangent vectors belonging to the tangent space of each point behave as the elements of a finite-dimensional vector space. In other words, this manifold consists of an infinite collection of vector spaces which are labeled by the points of the original manifold M. This is an example of a *vector bundle*. Specifically, it is called the *tangent bundle* of M and denoted by TM. Similarly, the set of all cotangent vectors is called the *cotangent bundle* and denoted by TM^*. A review of vector bundles is provided in Chap. 5. We suffice to say that a vector bundle is locally homeomorphic to the Cartesian product of an open subset of a manifold and a vector space. For example, if $\{(O_\alpha, \phi_\alpha)\}_\alpha$ is an atlas of M, then the subsets of TM consisting of the tangent vectors at the points of O_α are homeomorphic to $O_\alpha \times \mathbb{R}^m$. This allows us to represent the points of TM by pairs of the form: (p, v_p), where $p \in M$ and $v_p \in T_pM$.

Definition 18: Let $V : M \to TM$ be a (smooth) function such that for every $p \in M$, $V(p) = (p, v(p))$ for some $v(p) \in T_pM$. Then, V is called a (smooth) **vector field** on M. Because V is determined by its values, $v(p)$, we usually identify $V(p)$ and $v(p)$ and write $V(p) \in T_pM$. Similarly, a smooth function $\Omega : M \to T_pM^*$ with $\Omega(p) = (p, \omega(p))$ and $\omega(p) \in T_pM^*$, is called a **differential one-form** or simply a **one-form**. Again, we identify $\Omega(p)$ with $\omega(p)$.

Vector fields and one-forms are also called **contravariant** and **covariant** vector fields, respectively. They have many applications in theoretical physics. This is because we can write explicit local formulas for their components and use them to do calculations. Let us choose a local coordinate chart, (O_α, ϕ_α), that includes $p \in M$. The open subset O_α is represented by its image $\phi_\alpha(O_\alpha) \subset \mathbb{R}^m$. We have the following identification:

$$p \equiv \phi_\alpha(p) \equiv (x^1, \cdots, x^m) \in \mathbb{R}^m, \quad \forall p \in O_\alpha.$$

The same applies for the points of a curve $C : [0, T] \to M$. The portion of C that belongs to O_α is mapped to a curve in $\phi_\alpha(O_\alpha)$:

$$C(t) \equiv \big(x^1(t), \cdots, x^m(t)\big) =: x(t).$$

Let us now consider an arbitrary tangent vector $v_p = \frac{d}{dt}C(t)\big|_{t=0}$. The tangent curve C can be assumed to lie entirely inside O_α. v_p is given by

[14] We shall often use $\cong$ to denote an *isomorphism*.

Let us denote the components of v_p in the two coordinate charts by v_p^i and $v_p^{'i}$. Then, the chain rule

$$\frac{dx^{'i}}{dt} = \frac{\partial x^{'i}}{\partial x^j}\frac{dx^j}{dt}, \qquad i = 1, \cdots, m, \tag{A.7}$$

dictates the following transformation rule:

$$v_p^{'i} = \frac{\partial x^{'i}}{\partial x^j} v_p^j. \tag{A.8}$$

All the repeated indices are understood to be summed over their ranges. This is known as the *Einstein summation convention.*

Relation (A.8) can be taken as the definition of a (contravariant) vector. According to this definition a (contravariant) vector on an m-dimensional manifold is an object with m local components which satisfy the coordinate transformation rule given by (A.8). A simple way of remembering this transformation rule is to use the symbols $\frac{\partial}{\partial x^i}$ for the local coordinate basis vectors, i.e.,

$$v_p = v_p^i \frac{\partial}{\partial x^i}. \tag{A.9}$$

Then, recognizing that v_p does not depend on the choice of coordinates,

$$v_p^i \frac{\partial}{\partial x^i} = v_p = v_p^{'i} \frac{\partial}{\partial x^{'i}},$$

and enforcing the chain rule,

$$\frac{\partial}{\partial x^i} = \frac{\partial x^{'j}}{\partial x^i}\frac{\partial}{\partial x^{'j}},$$

we recover (A.8).

The same line of reasoning applies for (contravariant) vector fields. After all, a vector field is a vector-valued function. The only additional fact is that for a vector field, the point p in (A.9) becomes an independent variable. It is represented by its coordinates $(x^1, \cdots, x^m)$ or collectively by x. Therefore, a (contravariant) vector field V on $O_\alpha \subseteq M$ is expressed by

$$V = V^i(x)\frac{\partial}{\partial x^i}, \tag{A.10}$$

where $V^i(x)$ satisfy the coordinate transformation rule listed as (A.8).

The transformation properties of covectors (covariant vectors) is obtained similarly. Let $\omega_p \in T_pM^*$ be a covector. Then, for any $v_p \in T_pM$, $\omega_p(v_p)$ is a real scalar. We know from elementary linear algebra that T_pM^* as an m-dimensional vector space is isomorphic to $\mathbb{R}^m$. Thus, in the local chart (O_α, ϕ_α), we can write ω_p in terms of its components $(\omega_p)_i$, $i = 1, \cdots, m$. The scalar obtained by the action of ω_p on v_p can then be written componentwise,

$$\omega_p(v_p) = (\omega_p)_j (v_p)^j \tag{A.11}$$

However, as we pointed out earlier, a scalar is independent of the choice of coordinate charts. Hence, if we choose another coordinate chart, say, (O_β, ϕ_β) we must obtain the same result, namely

$$(\omega'_p)_j \, (v'_p)^j = \omega_p(v_p) = (\omega_p)_j \, (v_p)^j .$$

This equality together with (A.8) yield the transformation rule for the components of a covector (covariant vector):

$$(\omega'_p)_i = \frac{\partial x^j}{\partial x'^i} \, (\omega_p)_j . \tag{A.12}$$

A convenient notation for the local basis covectors is dx^i. In this notation, any covector is locally written as

$$\omega_p = (\omega_p)_i \, dx^i .$$

The independence of covectors from the choice of coordinate charts and the chain rule

$$dx'^i = \frac{\partial x'^i}{\partial x^j} \, dx^j ,$$

reproduces (A.12) immediately.

We would like to remark that at this stage the use of (the operators) $\frac{\partial}{\partial x^i}$ for the basis vectors and similarly the use of dx^i for the basis covectors are for practical purposes. In this notation, the duality of the basis vectors and covectors takes the following form

$$dx^i (\frac{\partial}{\partial x^j}) = \delta^i_j ,$$

where δ^j_i are the components of the Kronecker delta function,

$$\delta^i_j := \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j . \end{cases}$$

Letting p be an independent variable, we obtain the local expression for a covariant vector field or a one-form,

$$\omega(x) = \omega_i(x) \, dx^i .$$

The transformation rule for the components of a one-form is given by

$$\omega'_i(x') = \frac{\partial x^j}{\partial x'^i} \, \omega_j(x). \tag{A.13}$$

Again, we can take (A.13) as a definition of a covariant vector field.

Covariant and contravariant vector fields are simple examples of *tensor* fields. In general, a tensor is an elements of the tensor products of a bunch of vector spaces. A tensor at a point $p \in M$ is defined similarly as an element of the tensor product space

$$\underbrace{T_pM^* \otimes \cdots \otimes T_pM^*}_{r\text{-times}} \otimes \underbrace{T_pM \otimes \cdots \otimes T_pM}_{s\text{-times}} . \tag{A.14}$$

An element of this space is called an r-times covariant and s-times contravariant tensor at $p \in M$. r and s are also called the *covariant* and *contravariant ranks* of the tensor, respectively. Alternatively, we can define a tensor as a multilinear real-valued function of several vectors and covectors. Consider a function

$$T: \underbrace{T_pM \times \cdots \times T_pM}_{r\text{-times}} \times \underbrace{T_pM^* \times \cdots \times T_pM^*}_{s\text{-times}} \longrightarrow \mathbb{R},$$

such that T is linear in all its entries. Then, T is said to be a *tensor* of covariant and contravariant ranks r and s, respectively. The vector space depicted by (A.14) is the set of all such multilinear functions. This set forms a vector space under the operations of pointwise addition and multiplication by real numbers.

Tensor fields are tensor valued functions on a manifold. They play an important role in describing various physical quantities. Probably, the best example of the use of tensor fields in physics is in the theory of electromagnetism where the electromagnetic field strength[15] F is a tensor field [117, 146]. Tensor (fields) are also defined according to their coordinate transformation properties. For instance, the electromagnetic field tensor F is a covariant tensor of rank 2, because its components, $F_{\mu\nu}$, satisfy the following transformation rule

$$F'_{\mu\nu}(x') = \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x^\rho}{\partial x'^\nu} F_{\sigma\rho}(x). \tag{A.15}$$

The use of the Greek indices in (A.15) is to indicate that they refer to the Minkowski spacetime coordinates. Let us use the notation dx^μ for the basic covectors. Then, we have

$$F(x) = F_{\mu\nu}(x)\, dx^\mu \otimes dx^\nu .$$

In general, a (mixed) tensor of rank (r,s) is expressed locally according to

$$T(x) = T^{j_1 \cdots j_s}_{i_1 \cdots i_r}\, dx^{i_1} \otimes \cdots \otimes dx^{i_r} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_s}} .$$

The transformation properties of the components are easily obtained by requiring that $T(x)$ is coordinate independent.

[15] In the text we used the symbol $F^{(\mathrm{el})}$ to denote the electromagnetic field strength tensor to avoid any ambiguity. Here we drop the label $^{(\mathrm{el})}$ for simplicity.

An important class of tensor fields is the *totally antisymmetric covariant tensor fields*. These are more commonly called **differential forms**. As the name indicates, differential forms are covariant tensor fields whose components are antisymmetric in all their indices. The electromagnetic field tensor F is an example of a differential form of rank 2, a two-form. We shall occasionally drop the adjective "differential" for simplicity.

The space of all tensors (tensor fields) form a vector space. This is, by definition, the tensor product of copies of the tangent and cotangent spaces. The operation of *tensor product* makes this vector space into an (associative) *algebra*. This means that not only we can add tensors and multiply them by numbers, but we can multiply tensors by other tensors as well. To obtain the tensor product of two tensors, we multiply the components and take the tensor product of the basic tensors. For example, suppose

$$T = T_i^j\, dx^i \otimes \frac{\partial}{\partial x^j}, \quad S = S_k dx^k.$$

Then, we have

$$T \otimes S := T_i^j S_k\, dx^i \otimes \frac{\partial}{\partial x^j} \otimes dx^k.$$

In the next section we shall encounter other examples of an associative algebra. Therefore we next present the definition of an associative algebra.

Definition 19: Let $(\mathcal{A}, +, \cdot)$ be a vector space and $\otimes : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ be a binary operation satisfying the following conditions:

1) $(a_1 + a_2) \otimes a_3 = a_1 \otimes a_3 + a_2 \otimes a_3$;
2) $a_3 \otimes (a_1 + a_2) = a_3 \otimes a_1 + a_3 \otimes a_2$;
3) $c \cdot (a_1 \otimes a_2) = (c \cdot a_1) \otimes a_2 = a_1 \otimes (c \cdot a_2)$;
4) $a_1 \otimes (a_2 \otimes a_3) = (a_1 \otimes a_2) \otimes a_3$,

where a_1, a_2, and a_3 are arbitrary elements of $\mathcal{A}$ and $c \in \mathbb{R}$ ($c \in \mathbb{C}$ if $\mathcal{A}$ is a complex vector space). Then $(\mathcal{A}, +, \cdot, \otimes)$ is said to be an **associative algebra**. A subset $\{\mathcal{J}_i\}$ of $\mathcal{A}$ is said to generate $\mathcal{A}$ if every element of $\mathcal{A}$ can be written as a linear combination of products of $\mathcal{J}_i$. The elements $\mathcal{J}_i$ are called the **generators** of the algebra.

Similarly to the space of tensor fields, the space of all differential forms is a vector space. The tensor product, however, does not respect the requirement of antisymmetry. By this, we mean that the tensor product of two forms is in general a tensor whose components are not antisymmetric in all its indices. Thus, we need an alternative algebra operation for differential forms that preserves the antisymmetry. This is called the "antisymmetric tensor product" or the *wedge product*, $\wedge$. For example, we have

$$dx^\mu \wedge dx^\nu := dx^\mu \otimes dx^\nu - dx^\nu \otimes dx^\mu = -dx^\nu \wedge dx^\mu.$$

The electromagnetic field tensor can be written as

$$F = \frac{1}{2} F_{\mu\nu}\, dx^\mu \wedge dx^\nu. \tag{A.16}$$

The vector space of all differential forms together with the wedge product form an algebra known as the *exterior algebra* or the *Grassmann algebra.* A detailed introduction to differential forms can be found in H. Flanders' book: *Differential Forms with Applications to the Physical Sciences* (Dover, New York, 1989).

Nowadays, differential forms are used extensively in theoretical physics. They provide a remarkable tool for keeping the calculations short. In particular, they are useful in keeping track of the transformation properties of physical quantities. We can also develop a sort of calculus on differential forms by defining the so-called "exterior differentiation". This is an operation that takes a p-form to a $(p+1)$-form. The components of the resultant differential form are linear combinations of the first derivatives of the components of the original p-form.

Definition 20: Let Ω^p denote the space of all p-forms on a smooth manifold M. Then the **exterior derivative** is the map

$$d : \Omega^p \longrightarrow \Omega^{p+1},$$

defined locally by

$$d\omega := \left(\frac{\partial}{\partial x^j}\omega_{i_1\cdots i_p}\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p},$$

where

$$\omega = \omega_{i_1\cdots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p} \in \Omega^p.$$

An important observation is the following result.

Proposition 2: $d^2 := d \circ d = 0$.

A simple application of differential forms and exterior differentiation is in the theory of electromagnetism. It is not difficult to check that the components of the *vector potential* A_μ [117, 146], transform like the components of a one-form:

$$A := A_\mu dx^\mu, \qquad \mu = 0, 1, 2, 3.$$

Let us calculate

$$\begin{aligned} dA &= \left(\frac{\partial}{\partial x^\mu} A_\nu\right) dx^\mu \wedge dx^\nu \\ &= \frac{1}{2}\left[\frac{\partial}{\partial x^\mu} A_\nu - \frac{\partial}{\partial x^\nu} A_\mu\right] dx^\mu \wedge dx^\nu. \end{aligned}$$

We immediately recognize the content of the last bracket as the components of the electromagnetic field strength tensor,

$$F_{\mu\nu} = \frac{\partial}{\partial x^\mu} A_\nu - \frac{\partial}{\partial x^\nu} A_\mu.$$

Thus, we have the simple identity

$$F = dA.$$

An application of Proposition 2 is

$$dF = d^2 A = 0.$$

This simple equation is known as the "homogeneous Maxwell's equations" (in component form). These (two) equations together with the "inhomogeneous Maxwell's equations" describe all electromagnetic phenomena.

So far, we have defined different types of vector and tensor fields on smooth manifolds. Let us see how we can use a function between two manifolds to induce these fields from one manifold to another. Consider a smooth function $f : M_1 \to M_2$ between two smooth manifolds. Let v_p be a tangent vector at $p \in M_1$. v_p can be "pushed forward" to define a tangent vector at $f(p) \in M_2$.

Definition 21: Let M_1 and M_2 be smooth manifolds, $p \in M_1$, and $f : M_1 \to M_2$ be a smooth function. Then f induces a *linear* map

$$f_* : T_p M_1 \longrightarrow T_{f(p)} M_2$$

called the **push-forward** or the **differential map**. To describe this map let us consider an arbitrary tangent vector $v_p \in T_p M$ and choose a curve $C_1 : [0, T] \to M_1$ such that $v_p = \frac{dC_1}{dt}\big|_{t=o}$. The image of C_1 under f is a smooth curve

$$C_2 := f \circ C_1 : [0, T] \longrightarrow M_2$$

in M_2. The push-forward map is then defined by

$$f_*(v_p) := \left.\frac{dC_2}{dt}\right|_{t=0}. \tag{A.17}$$

A useful exercise is to show that this definition is independent of the choice of the curve C_1.

The same definition applies for a vector field by taking p to be an independent variable. However, we should point out that the push-forward map is defined locally. This means that, in general, the push-forward map can be used to induce a vector field only on an open neighborhood in M_2. There are cases in which a global vector field on the whole of M_1 cannot be pushed forward to define a smooth global vector field on $f(M_1) \subseteq M_2$.

Similarly, we can induce cotangent vectors and differential forms using a smooth map between two manifolds. However, this time it is the cotangent vectors and differential forms of M_2 that are "pulled back" on M_1.

Definition 22: Let M_1, M_2, and f be as in Def. 21. Then f defines a linear map

$$f^* : T_{f(p)} M_2^* \longrightarrow T_p M_1^*,$$

called the **pullback** map. To describe the pullback map we consider an arbitrary cotangent vector $\omega_{f(p)}$ at $f(p) \in M_2$. As an element of $T_pM_1^*$, the **pullback** of $\omega_{f(p)}$ is a linear map acting on T_pM_1. Thus, it can be defined through its action on arbitrary elements u_p of T_pM_1. We have

$$f^*(\omega_{f(p)})[u_p] := \omega_{f(p)}[f_*(u_p)], \tag{A.18}$$

where $f_*(u_p)$ is the push-forward of u_p.

The pullback map is also generalized to differential forms and general covariant tensor fields. Interestingly, it is a global mapping. That is, every global differential form on M_2 can be pulled back to define a global differential form on $f^{-1}(M_2) \subseteq M_1$. We make extensive use of the pullback operation in our discussion of universal connections and their realization in the phenomenon of geometric phase in Chaps. 6 and 7.

We will end this section by giving the definition of *compactness* for manifolds. As we mentioned in our discussion of general topological spaces, compactness is a topological property. It is defined for arbitrary topological spaces. However, the general definition is rather abstract. If a topological space is a subspace of $\mathbb{R}^d$ for some $d \in \mathbb{Z}^+$, as is a manifold, we can use a result of real analysis to arrive at a more intuitive definition.

Definition 23: Let M be an arbitrary (topological) manifold. Embed M into some $\mathbb{R}^d$. M is said to be a **compact manifold**, if it is a *closed* and *bounded* subset of $\mathbb{R}^d$. As usual, a closed subset is a subset whose complement is open. Moreover, a subset M of $\mathbb{R}^d$ is called bounded if there is an open ball

$$B_r^d = \{\mathbf{x} \in \mathbb{R}^d \ : \ |\mathbf{x}| < r\}$$

in $\mathbb{R}^d$, such that $M \subset B_r^d$.

Clearly, $\mathbb{R}^m$ or an open ball in $\mathbb{R}^m$ is not compact, whereas spheres, tori, and (finite-dimensional) projective spaces are all compact manifolds.

A.3 Lie Groups

In the preceding section, we introduced manifolds and gave several examples of smooth manifolds. In this section, we shall discuss another important class of smooth manifolds called *Lie groups.* The mathematical theory of Lie groups is one of the most well-established and substantial achievements of modern mathematics. It is also of great practical and theoretical use in physics. Our aim will be, therefore, not to attempt to present a complete review of Lie groups. Instead, we shall try to point out the basic concepts and emphasize a few of the most important and useful facts about Lie groups. More detailed discussions of the theory of Lie groups and its applications in physics can be found in

– M. Hamermash: *Group Theory and Its Application to Physical Problems* (Dover, New York, 1989);

- J. P. Elliot and P. G. Dawer: *Symmetry in Physics*, Vol. 1 and 2 (Oxford University Press, New York, 1990);
- R. Gilmore: *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974);
- H. Georgi: *Lie Algebras in Particle Physics* (Addison-Wesley, New York, 1982); and
- S. Sternberg: *Group Theory and Physics* (Cambridge University Press, Cambridge, 1994).

Some of the more advanced textbooks on Lie groups and their representations are

- S. Helgason: *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic Press, New York, 1978);
- T. Bröcker and T. tom Dieck: *Representations of Compact Lie Groups* (Springer-Verlag, New York, 1985);
- W. Fulton and J. Harris: *Representation Theory* (Springer-Verlag, New York, 1991).

Definition 24: Let G be a set of points, and

$$\bullet : G \times G \longrightarrow G$$

be a binary operation. Then the pair $(G, \bullet)$ is said to be a **group** if the following conditions are fulfilled.

1) $\bullet$ is associative:

$$(g_1 \bullet g_2) \bullet g_3 = g_1 \bullet (g_2 \bullet g_3), \quad \forall g_1, g_2, g_3 \in G;$$

2) there is an identity element $e \in G$ such that

$$e \bullet g = g \bullet e = g, \quad \forall g \in G;$$

3) every element $g \in G$ has an inverse, g^{-1} such that

$$g \bullet g^{-1} = g^{-1} \bullet g = e.$$

A subset H of a group G which has the structure of a group with the same group multiplication is called a *subgroup* of G. A necessary and sufficient condition for a subset H to be a subgroup of G is that for every h_1 and h_2 belonging to H, the element $h_1 \bullet h_2^{-1}$ also belongs to H.

Given a subgroup H of a group G, one can define a canonical equivalence relation on G. This equivalence relation is defined according to the requirement:

$$\forall g_1, g_2 \in G, \ g_1 \sim g_2, \quad \text{if and only if} \ \ g_1 \bullet H = g_2 \bullet H.$$

The last equality means that there exist $h_1, h_2 \in H$ with $g_1 \bullet h_1 = g_2 \bullet h_2$.

It is a simple exercise to show that this relation is indeed an equivalence relation. The equivalence class including an element $g \in G$ is denoted by gH

and called a *left coset*. The set of all equivalence classes is called a *left coset space* or a *quotient space* of G and denoted by G/H. Similarly one defines the right cosets by defining a similar equivalence which involves the requirement: $H \bullet g_1 = H \bullet g_2$ for the equivalent elements g_1 and g_2.

Groups have proven to be extremely important mathematical constructions. They have been used in other areas of mathematics as well as different branches of natural sciences. Probably, the main reason for the utility of groups in natural sciences is the existence of "symmetry" in nature.

In our discussion of equivalence classes, we pointed out that once there is an equivalence relation, the universal set (class) of all points (objects) is divided into subsets of equivalence classes. The elements of each equivalence class could be treated equally. Hence, we associated the word "symmetry" with this situation. Groups, as we shall see, "quantify" symmetries. In this sense, "different" groups correspond to different types of symmetries. In the following, we shall first try to explain the meaning of the last couple of sentences. We will start with a familiar example.

In some cases, we have two different equivalence relations on a single universal set. In fact, we have encountered some examples of this. Let us consider the collection of all topological spaces. The elements are topological spaces $(X, \mathcal{T})$ that are distributed among the distinct equivalence classes of homeomorphic topological spaces. Also, we have the collection of all the universal sets X of the above collection. This is the collection of all point sets. We have the equivalence relation defined by the existence of bijections. Obviously, topological equivalence is a stronger condition than set theoretic equivalence. This is simply due to the fact that every homeomorphism is a bijection, but not every bijection is a homeomorphism. Therefore, each equivalence class of bijective sets is further subdivided into equivalence classes of homeomorphic topological spaces. So as we see in this example, we have different types of "symmetries."

Let us concentrate on topological symmetry and see how it is related to groups. Consider a point set X and let us study the set of possible topological structures on X. Any two topological structures on X, equivalent or not, are related to one another through a bijection $f : X \to X$. Hence we study the set of all such bijections which we denote by S_X.

Consider a mapping X by two consecutive bijections,

$$X \xrightarrow{f_1} X \xrightarrow{f_2} X.$$

This is done by composing the bijections f_1 and f_2. We have

$$X \xrightarrow{f_2 \circ f_1} X.$$

It is easy to check that the set S_X with the operation of composition forms a group. The operation of composition of functions is associative. The identity function

$$id : X \ni p \longrightarrow p \in X$$

is clearly a bijection. It provides the identity element of the group of all bijections. This is called the *permutation group* of the set X and denoted by S_X. Finally, the inverses exist and they are also bijections by definition.

Now consider a particular topological structure $\mathcal{T}$ on X. Under the action of a bijection f and its inverse f^{-1}, the open subsets of X are either mapped onto the same collection $\mathcal{T}$ of open subsets or some of them do not. In the first case, the bijection f is by definition a homeomorphism mapping the topological space $(X, \mathcal{T})$ to itself. Let us denote the set of all such homeomorphism by $\mathcal{H}(X, \mathcal{T})$. It is not difficult to show that $\mathcal{H}(X, \mathcal{T})$ forms a subgroup of the permutation group S_X. $\mathcal{H}(X, \mathcal{T})$ is called the *homeomorphism group* of $(X, \mathcal{T})$. Furthermore, the set of all topological spaces which are not homeomorphic to $(X, \mathcal{T})$ corresponds to the quotient set $S_X/\mathcal{H}(X, \mathcal{T})$.

Similarly to the permutation group of a set and the homeomorphism group of a topological space, we can introduce the *diffeomorphism group* $\mathrm{Diff}(M)$, of a differentiable manifold M. This is the set of all diffeomorphisms of M together with the operation of composition.

The examples of groups that we have mentioned are indeed quite complicated mathematical structures. There are much simpler examples of groups with a finite number of elements. These are called *finite groups.* The group which has a single element is called the *trivial group.* The simplest non-trivial group is the group $\mathbb{Z}_2$. It has two elements. There are two defining "representations" of $\mathbb{Z}_2$. These are known as the additive and the multiplicative representations. The additive representation is obtained by viewing $\mathbb{Z}_2$ as the set of all integers modulo 2. The group operation is the simple addition of integers modulo 2, i.e., $\bar{a} + \bar{b} := \overline{a + b}$. The additive representation is, therefore, given by

$$\mathbb{Z}_2 = (\{\bar{0}, \bar{1}\}, +) .$$

The multiplicative representation is

$$\mathbb{Z}_2 = (\{-1, 1\}, \times) ,$$

where "$\times$" means the ordinary multiplication of numbers. Other simple examples of finite groups are the groups $\mathbb{Z}_n$, integers modulo n with addition, and the permutation groups S_n. The latter consists of all the bijections from a finite set I_n of n elements onto I_n, i.e., $S_n = S_{I_n}$. Clearly, S_n has $n!$ elements.

We can divide all groups into two large classes based on whether the group operation is *commutative* or not. A group $(G, \bullet)$ is called a *commutative* or an *Abelian group*, if for every g_1 and g_2 in G,

$$g_1 \bullet g_2 = g_2 \bullet g_1 .$$

Evidently, $\mathbb{Z}_n$ are Abelian groups, whereas S_n $(n > 2)$ are not Abelian. An instructive exercise would be to show that S_n are *non-Abelian* for $n > 2$.[16]

[16] Hint: Show that S_3 is non-Abelian and use the fact that S_3 is a *subgroup* of S_n, for $n > 2$.

As for every class of mathematical structures, we will need to define a concept of *equivalence* for groups. This is also defined by the existence of certain functions between two groups that preserve the *group structure.* Such a function is called a *group isomorphism.*

Let us consider two groups G and H. We will use the same notation for both group operations, "$\bullet$", but denote the elements of G and H by g and h, respectively.

Definition 25: A function $f : G \to H$ is said to be a **group homomorphism** if for every $g_1, g_2 \in G$

$$f(g_1 \bullet g_2) = f(g_1) \bullet f(g_2).$$

A bijective homomorphism is called a **group isomorphism**.

We can easily check that the existence of isomorphisms defines an equivalence relation on the class of all groups. The elements of the isomorphism (equivalence) classes are not distinguished in group theory. Probably, the simplest example of a group isomorphism is

$$\mathbb{Z}_2 \cong S_2.$$

We use the notation "$\cong$" for the word "isomorphic". The classification of all finite groups (up to isomorphy) is one of the most difficult tasks in mathematics.

A natural concept in group theory is the concept of a *product group.* The reader should be able to define this concept by himself or herself.

A standard elementary textbook on group theory is *Topics in Algebra* by I. N. Herstein (Blaisdell Publishing Company, New York, 1964). A more advanced textbook is *Algebra* by T. W. Hungerford (Springer, New York, 1987).

We briefly mentioned that groups are associated with symmetries. A typical and rather instructive example of this is the group of the geometric operations on a plane that leave an equilateral triangle unchanged. These are three (0, 120, and 240 degrees) rotations about the center, and three reflections about the symmetry axes. These operations form a group which is also called *the symmetry group of a triangle.* This group is isomorphic to S_3. It is a non-Abelian group that contains the subset of the three rotations as an Abelian subgroup. The next example of a geometric symmetry group is the group of symmetries of a square. This includes four rotations and four reflections. It is not isomorphic to any of the groups that we have discussed so far, neither is it a product group of some "smaller" groups. We can proceed to introduce the symmetry groups of other geometric objects such as equilateral polygons. Evidently, the number of elements of the symmetry group – this is called the *order* of the finite group – depends on the number of sides of the polygon. The limiting object is the *round circle.* Its symmetry group has an infinite number of elements. These are the rotations about the center by arbitrary angles $\varphi \in [0, 2\pi)$ and reflections about any axis through the center. The latter can also be parameterized by an angle, namely the angle that the axis of reflection makes with, say, the x-axis in $\mathbb{R}^2$. Alternatively, one can perform an arbitrary reflection by a combination of two rotations and a fixed reflection, say about the x-axis. In order to do this, first one performs a suitable rotation so that the axis of reflection is rotated to the x-axis. This is followed by a reflection about the x-axis and finally a rotation by the same angle but in the opposite sense. This shows that the symmetry group of the circle is not parameterized by two independent continuous variables.

Let us consider the subgroup of the symmetry group of the circle consisting of rotations. This is an Abelian group. Its elements are parameterized by the angles or the points of another round circle. Let g be such a rotation, and let us view the original circle S^1 as the set of vectors of unit length in $\mathbb{R}^2$,

$$S^1 := \left\{ \mathbf{v} \in \mathbb{R}^2 \ : \ \mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}, \ x^2 + y^2 = 1 \right\}.$$

Then, g is represented by a matrix that "multiplies" the points of S^1 from the left:

$$\begin{pmatrix} x \\ y \end{pmatrix} \xrightarrow{g} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

In this way, we have an explicit representation of an infinite group, namely

$$SO(2) := \left\{ \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \ : \ \varphi \in [0, 2\phi) \right\}.$$

The notation $SO(2)$ means the set of all *special* (determinant $= 1$), orthogonal (inverse $=$ transpose), two-dimensional real matrices. In $\mathbb{R}^3$, it corresponds to the rotations about a fixed axis.

We can also use the identity $\mathbb{R}^2 = \mathbb{C}$, to view S^1 as the set of complex numbers with unit modulus,

$$\{z \in \mathbb{C} \ : \ |z| = 1\}.$$

If $|z| = 1$, then z is a phase,

$$z = e^{i\theta}, \qquad \theta \in [0, 2\pi).$$

In this representation, the rotation by an angle φ is performed by multiplication by the phase $e^{i\varphi}$,

$$e^{i\theta} \xrightarrow{g} e^{i\theta'} = e^{i\varphi}\, e^{i\theta} = e^{i(\theta+\varphi)}.$$

The group of rotations is then parameterized by the set of all phases,

$$U(1) := \left\{ e^{i\varphi} \ : \ \varphi \in [0, 2\pi) \right\},$$

where the group multiplication is the ordinary multiplication of complex numbers. The letter "U" stands for "unitary", since every phase is indeed a unitary (inverse $=$ Hermitian conjugate) one-by-one matrix.

The two groups $SO(2)$ and $U(1)$ are in fact the same, i.e., they are isomorphic. It is quite easy to see that the group elements, in this case, describe the points of a smooth manifold, namely the circle, S^1. $U(1)$ is a typical example of a (compact, connected) *Lie group.*

Another example of a compact (but disconnected) Lie group is the full symmetry group of the round circle. This group consists of the (subgroup of) rotations and arbitrary reflections. We can construct this group algebraically. For this purpose we consider the round circle S^1 as the set of all unit vectors in $\mathbb{R}^2$ originating at $0 \in \mathbb{R}^2$, and find all the linear transformations which map S^1 onto itself, i.e., preserve the magnitude of the vectors. It is not difficult to see that these transformations correspond to 2×2 orthogonal matrices. The set of all such matrices is denoted by $O(2)$. We know from elementary linear algebra that the determinant of any orthogonal matrix is either 1 or -1. As we showed above the rotations correspond to the special orthogonal matrices which have unit determinant. Thus the reflections are identified by the orthogonal matrices of determinant -1. Alternatively, we can view S^1 as the set of complex numbers of unit norm (modulus), and try to find the symmetry group of S^1 as the group of linear transformations of the complex numbers $\mathbb{C}$ which preserve the norm. In this picture, the relation between reflections and the operation of complex conjugation is most interesting.

Having examined some simple examples of Lie groups, we pursue our review by presenting a precise definition of a Lie group.

Definition 26: Let $(G, \bullet)$ be a group that has, in addition, the structure of a smooth manifold. Then, G is said to be a **Lie group** if the functions defined by the group multiplication, $\bullet : G \times G \to G$,

$$\bullet(g_1, g_2) := g_1 \bullet g_2,$$

and inversion, $i : G \to G$

$$i(g) := g^{-1},$$

are smooth functions.

Other examples of Lie groups are the spaces $\mathbb{R}^m$, where the group multiplication is simply the addition of vectors. We also have the multiplicative group $(\mathbb{R} - \{0\}, .)$ which is a disconnected Lie group. Similarly, the multiplicative group $(\mathbb{C} - \{0\}, .)$ has the structure of a Lie group with the group space being the punctured complex plane. Other, more interesting groups are the so-called *classical groups.* These are different types of matrix groups. We will discuss the *unitary groups* $U(n)$ in some detail. The other classical groups are discussed in most of the textbooks on Lie group theory.

Lie groups enter into physical problems as the transformation groups of physical systems. Often, a physical system is defined on a smooth manifold M. Depending on the specific geometry of the problem there may exist quantities that are invariant under certain transformations of the manifold. These quantities are called *conserved quantities.* In fact, the set of all manageable transformations is the diffeomorphism group of the manifold. Here one can view the manifold as the submanifold of some Euclidean space $\mathbb{R}^d$ and interpret a diffeomorphism as a smooth and smoothly reversible deformation of the manifold in the embedding Euclidean space.

In field theories, physical quantities are usually represented by tensor fields $T = T(x)$.[17] In general, under a diffeomorphism a point $x \in M$ is mapped to another point $x' \in M$. As regards the (tensor) fields T, there are two possibilities. Either one keeps using the original fields evaluated at the new point, i.e., under the diffeomorphism $T(x) \to T(x')$, or one also transforms the tensor fields according to the diffeomorphism $T(x) \to T'(x')$. The latter is performed by pushing-forward or pulling-back the field T (or a combination of these) using the diffeomorphism and its inverse. Adopting this latter point of view, one may interpret the effect of a diffeomorphism as a reparameterization of the physical quantities. Thus the corresponding symmetries are related to the reparameterization invariance of the physical quantities. These symmetries are known as the internal symmetries. In a sense they are indispensable qualities of every sensible field theory. The *diffeomorphism invariance* of Einstein's general theory of relativity and the gauge symmetries of other field theories are examples of internal symmetries. The former point of view, in contrast, corresponds to the external symmetries which are related to the specifics of a physical system. They may or may not be present.

Usually, the (tensor) fields used to represent physical quantities are given by their local components. The use of local components is often necessary to perform computations. These local components, however, depend on the choice of a local coordinate chart and a local basis of the tangent and cotangent spaces. The choice of coordinates is completely subjective. Hence, the physical quantities must be independent of such a choice. A local coordinate transformation on a manifold is identical with a coordinate transformation on a copy of $\mathbb{R}^m$, where m is the dimension of the manifold. Coordinate transformations on $\mathbb{R}^m$ also form a group. The elements of this group may be used to parameterize the local coordinate transformations of the manifold. The physical quantities are however invariant under the "action" of this group. In fact, a coordinate transformation does not move the points of the manifold. It is merely a relabeling of the points of the corresponding local chart and thus corresponds to the identity element of the diffeomorphism group. Consequently we do not associate the coordinate invariance (covariance) of the physical quantities with a symmetry of a physical system.[18]

Consider a free particle moving on a plane. The manifold M is $\mathbb{R}^2$. The dynamics of the free particle must certainly be independent of the choice of the coordinate axes. Let us fix a Cartesian coordinate system. Then, any

---

[17] This includes the scalar, covariant, and contravariant vector fields.

[18] The local representation of a diffeomorphism, i.e., when it is represented in local coordinates, resembles a local coordinate transformation. This is the basis of the terminology according to which diffeomorphism symmetry is also called symmetry under general coordinate transformations.

other such system is obtained from the first one through (a combination of) rotations, translations, and interchange of the coordinate axes. The set of all such transformations in $\mathbb{R}^2$ form a group called the *Euclidean group*, $E(2)$. Note that the arbitrariness of the choice of a coordinate system is different from say an $E(2)$ symmetry of a problem.

Let us consider the longitudinal symmetry of the electrostatic properties of a long homogeneous charged metal bar of fixed cross-sectional geometry. The symmetry group is clearly $\mathbb{R}$. The presence of this symmetry allows us to reduce the three-dimensional problem to a two-dimensional one. The full solution is then obtained via the symmetry argument that the results must be independent of the longitudinal coordinate. If further we suppose that the metal bar is cylindrical, the symmetry group is even larger. It is the product group $\mathbb{R} \times O(2)$. Since the symmetry group is two-dimensional the problem is further reducible to a one-dimensional one. The final result will only depend on the distance of the observer from the metal bar.

We have seen how symmetries are related to groups. Specially, we examined the finite and infinite (Lie) groups of transformations. In general, the space which undergoes a transformation is an arbitrary smooth manifold. A transformation of a manifold by a group element is called the *action of the group element on the manifold.* As any transformation, the action of a group element is defined through a function acting on the manifold. More generally we have the following definition.

Definition 27: Let $(G, \bullet)$ and M be a Lie group and a smooth manifold, respectively. A smooth function $f : G \times M \to M$ is said to be a **left action** of G on M, if

1) for all $p \in M$, $f(e,p) = p$, where e is the identity element of G;
2) for all $g_1, g_2 \in G$ and $p \in M$,

$$f\left(g_1, f(g_2,p)\right) = f(g_1 \bullet g_2,\, p).$$

Usually, one abuses the notation and writes "$g \bullet p$" or even "$g\,p$" for $f(g,p)$. In this notation, the requirements of Def. 27 become $e\,p = p$ and $g_1(g_2p) = (g_1g_2)p$. The function f is called a **right action** if instead of the second requirement, we have

$$f\left(g_1, f(g_2,p)\right) = f(g_2 \bullet g_1,\, p).$$

Similarly, one denotes a right action by

$$f(g,p) \equiv p \bullet g \quad \text{or} \quad f(g,p) \equiv p\,g.$$

Every Lie group G has a natural left action on itself. This is simply given by group multiplication from the left:

$$f(g_1, g_2) = \bullet(g_1, g_2) = g_1 \bullet g_2 \equiv g_1g_2.$$

Similarly, one can define the right action of G on itself. For each group element $g \in G$, the left and the right actions of G on itself define two *canonical* smooth functions $L_g : G \to G$ and $R_g : G \to G$, respectively. They are given by

$$L_g(h) := g \bullet h \tag{A.19}$$

$$R_g(h) := h \bullet g. \tag{A.20}$$

These functions and their push-forward and pullback maps have many applications in the theories of Lie groups and fiber bundles.

Definition 28: An action f is called **transitive** if for every two points p_1 and p_2 of M, there is some $g \in G$ such that $p_2 = f(g, p_1)$. f is called a **free action**, if for every $p \in M$ and $g \in G - \{e\}$, $f(g,p) \neq p$.

For example, let us consider the left action of $SO(2)$, rotations about the z-axis, on the unit (round) sphere S^2 inside $\mathbb{R}^3$. This is neither transitive nor free. To show this, first we need two points on S^2 that are not linked via a rotation about the z-axis. This is easily done by choosing two points with different values of z-coordinate in $\mathbb{R}^3$. This means that the action is not transitive. Next, consider the poles. They are left unchanged by all such rotations. Thus, the action is not free either.

An important example of a Lie group is the full rotation group in $\mathbb{R}^3$. An arbitrary rotation in $\mathbb{R}^3$ is specified by three numbers; these are known as the *Euler angles* [147]. An equivalent specification of an arbitrary element of this group is obtained by choosing a unit vector centered at the origin and a point of the unit sphere centered at the tip of this unit vector. One can show that this group is isomorphic to $SO(3)$. Geometrically, it is the manifold obtained by identifying the opposite points of the three sphere S^3 (with respect to the center). In fact, the whole sphere S^3 is another interesting example of a Lie group, namely the group $SU(2)$. $SU(2)$ is defined as the subgroup of the unitary group $U(2)$ that consists of the matrices of unit determinant.

We can study the action of $SO(3)$ on $\mathbb{R}^3$ by identifying the points of $\mathbb{R}^3$ by column vectors and multiplying them by elements of $SO(3)$, i.e., orthogonal 3×3 matrices with unit determinant. We can also restrict this action onto the submanifold S^2 of $\mathbb{R}^3$. Since every rotation preserves the magnitude of a vector, the action of $SO(3)$ maps S^2 to itself. In fact, we can easily see that this action is transitive but not free. An example of an action of a Lie group G that is both transitive and free is the left (right) action of G on itself.

Probably the most important concept in the theory of Lie groups is the concept of the *Lie algebra* of a Lie group. To present a definition of the Lie algebra of a Lie group we need to recall some basic facts about vector fields on a general manifold.

In our discussion of vector fields on smooth manifolds, we introduced a practical local expression for arbitrary vector fields. We labeled this formula as (A.10). In this expression, we denoted every vector field by a differential operator. In fact, it turns out that there is a one-to-one correspondence between the set of all (contravariant) vector fields and the set of all the differential operators of the form given by (A.10). Under this correspondence every vector field is viewed as a differential operator acting on the space of all scalar fields. An interesting property of operators is that they can be composed. Let us choose two vector fields $V = V^i(x)\frac{\partial}{\partial x^i}$ and $W = W^j(x)\frac{\partial}{\partial x^j}$ and consider their *commutator*:

$$
\begin{aligned}
[V,W] &:= V \circ W - W \circ V \\
&= V^i \frac{\partial}{\partial x^i}\left(W^j \frac{\partial}{\partial x^j}\right) - W^j \frac{\partial}{\partial x^j}\left(V^i \frac{\partial}{\partial x^j}\right) \\
&= \underbrace{\left(V^i \frac{\partial W^k}{\partial x^i} - W^j \frac{\partial V^k}{\partial x^j}\right)}_{U^k} \frac{\partial}{\partial x^k}.
\end{aligned}
\tag{A.21}
$$

We can readily check that the components U^k of $[V,W]$ transform like the components of a (contravariant) vector field, i.e., according to (A.8). Let us denote the set of all vector fields of a smooth manifold M by $\mathcal{X}(M)$. The operation defined by (A.21) is a binary operation,

$$[\cdot,\cdot] \,:\, \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M),$$

that promotes $\mathcal{X}(M)$ into a *non-associative algebra.*[19] This operation is called the *Lie bracket* of two vector fields.

Definition 29: Let G be a Lie group and $\mathcal{X}(G)$ be the algebra of vector fields on G. A vector field $X \in \mathcal{X}(G)$ is said to be a **left-invariant vector field**, if for every $g, h \in G$, it satisfies

$$L_{g*}\left(X(h)\right) = X(g \bullet h),$$

where $X(h)$ and $X(g \bullet h)$ are the values of the vector field at the argument points and L_{g*} is the push-forward map induced by the left action of G on itself (A.19). It can be shown that the Lie bracket of two left-invariant vector fields is also left-invariant. Thus, the set of all left-invariant vector fields form a subalgebra of $\mathcal{X}(G)$. The algebra operation is obviously the Lie bracket. This algebra is called the **Lie algebra** of G. It is denoted by $\mathcal{G}$ or LG.

A "geometrical" interpretation of the Lie algebra $\mathcal{G}$ is offered by the following simple result.

Proposition 3: As vector spaces $\mathcal{G}$ and T_eG are isomorphic.

Although this result may seem rather mysterious, it is shown quite straightforwardly. The key point is to recognize that every left-invariant vector field X can be constructed from its value at the identity, $X(e) \in T_eG$. This is done via the left action map L_g of (A.19). We have, for any $g \in G$,

[19] A non-associative algebra satisfies all the requirements of an associative algebra except the condition of associativity, i.e., condition 4 of Def. 19.

$$X(g) = L_{g*}\left(X(e)\right). \tag{A.22}$$

Using this equation, we can push-forward a basis of T_eG to define a basis for $\mathcal{G}$. This is sufficient to prove the isomorphy of these two vector spaces. Since $\dim(T_eG) = \dim(G)$, $\mathcal{G}$ is finite dimensional if and only if G is. Another important implication of (A.22) is that every basis of T_eG induces a set of global basis vector fields in TG. In the language of vector bundles this means that the tangent bundle of every Lie group is a trivial bundle,[20] i.e., it is a product manifold,

$$TG = G \times \mathbb{R}^m$$

where $m = \dim(G)$.

Furthermore, since the Lie algebra $\mathcal{G}$ is isomorphic to T_eG, we can write down the defining (commutation) relations in terms of the tangent vectors at the identity. This in turn indicates that if we compute the Lie bracket of the elements of the Lie algebra via (A.17), the coefficients of the right-hand side of (A.21) will be constant. For example, let us choose a basis $\{J_i\}$ of $\mathcal{G}$. Then, we have

$$[J_i, J_j] = c_{ij}^k\, J_k. \tag{A.23}$$

The basis elements J_i of the Lie algebra are also called the *generators* of the Lie group. The coefficients c_{ij}^k are called the *structure constants*. They determine the Lie algebra. This means that a complete set of commutation relations such as (A.23) specifies the Lie algebra without any reference to the structure of the Lie group. In fact, there is a way to find a Lie group associated to a given Lie algebra. The association is however not unique, i.e., we can find several Lie groups for a given Lie algebra.

In order to define a Lie algebra independently of a Lie group, we shall first try to determine the important properties of the Lie algebra of a Lie group. These are simply the antisymmetry of the Lie bracket:

$$[V, W] = -[W, V],$$

and the so-called *Jacobi identity*:

$$[[U, V], W] + [[W, U], V] + [[V, W], U] = 0,$$

where U, V, W are arbitrary Lie algebra elements. We can easily verify these identities for the Lie algebra of a Lie group using (A.21). For an *abstract Lie algebra* however, they serve as the defining postulates or axioms.

Definition 30: Let $(\mathcal{G}, +, \cdot, [\cdot, \cdot])$ be a non-associative algebra with the algebra multiplication denoted by

$$[\cdot, \cdot] : \mathcal{G} \times \mathcal{G} \longrightarrow \mathcal{G}.$$

[20] A manifold whose tangent bundle is trivial is called a *parallelizable manifold*, so every Lie group is parallelizable.

Then, $\mathcal{G}$ is called an **abstract Lie algebra** if the algebra operation is antisymmetric and it satisfies the Jacobi identity.

A familiar example of an abstract Lie algebra is the algebra of operators in one-dimensional quantum mechanics. The basic elements are the coordinate operator $\hat{x}$, the momentum operator $\hat{p}$, and the identity operator $\hat{1}$. The algebra operation is the commutator of the operators. This operation satisfies all the requirements of Def. 30. The corresponding Lie algebra is called the Weyl–Heisenberg algebra. It is defined by the following commutation relations:

$$\begin{aligned}
[\hat{x},\hat{p}] &= i\hat{1} \\
[\hat{x},\hat{x}] &= [\hat{p},\hat{p}] = \left[\hat{1},\hat{1}\right] = 0 \\
\left[\hat{x},\hat{1}\right] &= \left[\hat{p},\hat{1}\right] = 0.
\end{aligned}$$

Another simple example is the Lie algebra of all $(N \times N)$-matrices with the algebra operation being the commutator of two matrices. This Lie algebra is denoted usually by $\mathcal{G}\ell(N,\mathbb{R})$ or $\mathcal{G}\ell(N,\mathbb{C})$ depending on whether the entries of the matrices are real or complex.[21]

Lie algebras of Lie groups are used extensively in almost all aspects of Lie group theory. A particularly important tool that makes the applications of the Lie algebra possible is the so-called *exponential map.* This is a smooth map from the Lie algebra to the Lie group. In fact, if the Lie group is compact and connected then the exponential map is onto. That is, every group element can be obtained as the *exponential* of some element of the Lie algebra (a tangent vector at the identity). This is extremely important because in practice it is much easier to study a Lie algebra than a Lie group.

Let us consider a smooth manifold M and suppose that $V = V^i(x)\frac{\partial}{\partial x^i}$ is a vector field on M. At each point $x \in M$, we can obtain the value of V as the tangent vector to some curve in M. Let us choose $x_0 \in M$ and denote $V(x_0)$ by v_0. Then, we can find a curve C that starts at x_0 and is tangent to the vector field V at all its points,

$$V(C(t)) = \frac{dC(t)}{dt}, \quad \forall t \in [0,T].$$

The curve C is called an *integral curve* of the vector field V. Every integral curve is uniquely defined up to the starting point x_0. This is a consequence of the existence and uniqueness theorem for ordinary differential equations. For, an integral curve is the solution of the following first-order differential equation:

$$\begin{aligned}
\frac{d}{dt}x^i(t) &= V^i(x(t)) \\
x^i(0) &= x_0^i,
\end{aligned} \tag{A.24}$$

[21] Note that the set of such matrices also form an associative algebra under matrix multiplication.

where $x(t) = (x^1(t), \cdots, x^m(t))$ are the coordinates of the points $C(t)$ of C, and x_0^i are the components of the initial point x_0. The integral curve C is obtained by integrating (A.24).

We can use the results of the previous paragraph to define a smooth function on M that maps an arbitrary point x_0 to the end point of the corresponding integral curve. If we set $T = 1$, i.e., $t \in [0, 1]$, this function maps $x_0 = C(0)$ to $C(1)$, where C is the solution of (A.24). There are manifolds with pathological problems that render this construction inapplicable. Lie groups and, as a matter of fact, all the manifolds we will encounter in this book do not have such problems and the above construction is valid. We denote the "end point" function by

$$\exp_M : \mathcal{X}(M) \times M \longrightarrow M.$$

It is defined by

$$\exp_M(V, x_0) := C(1). \tag{A.25}$$

In words, the function $\exp_M$ yields the end point of the integral curve defined by the vector field $V \in \mathcal{X}(M)$ and the initial condition $x_0 \in M$.

Let us return to our discussion of Lie groups and their Lie algebras. Every element of the Lie algebra of a Lie group is a left-invariant vector field X. We can obtain a function from the Lie algebra into the Lie group by restricting (A.25) to the left-invariant vector fields on G and choosing x_0 to be the identity element of the Lie group, $e \in G$. The resulting function is called the *exponential map*, $\exp : \mathcal{G} \to G$:

$$\exp(X) := \exp_G(X, e).$$

An examination of matrix groups justifies the name "exponential map". For example, the Lie algebra of $U(1)$ is the purely imaginary numbers, $u(1) = i\mathbb{R}$. Then for any $i\varphi \in i\mathbb{R}$, we have

$$\exp(i\varphi) = e^{i\varphi}$$

The same is true for all other matrix groups. An element of the Lie algebra of every matrix group is itself a matrix of the same dimension. The exponential map for these groups reduces to

$$\exp(X) = e^X := \sum_{k=0}^{\infty} \frac{1}{k!} X^k.$$

We shall next review an important class of matrix groups, called the unitary groups, $U(N)$.

Definition 31: Let $GL(N, \mathbb{C})$ denote the set of all invertible $(N \times N)$ complex matrices. $GL(N, \mathbb{C})$ form a Lie group under the operation of matrix multiplication. It is called the *(complex) general linear group*. The Lie algebra

$\mathcal{G}\ell(N,\mathbb{C})$ of $GL(N,\mathbb{C})$ consists of all $(N\times N)$ complex matrices. The **unitary group** $U(N)$ is a compact Lie subgroup of $GL(N,\mathbb{C})$ defined by

$$U(N) := \left\{U \in GL(N,\mathbb{C}) \; : \; U^{-1} = U^{\dagger} \right\},$$

where "†" stands for Hermitian conjugation. The unitary matrices of unit determinant form a Lie subgroup of $U(N)$ called the **special unitary group**:

$$SU(N) := \{U \in U(N) \; : \; \det(U) = 1 \}.$$

The Lie algebras of $U(N)$ and $SU(N)$ are denoted usually by $u(N)$ and $su(N)$, respectively. They are Lie subalgebras of $\mathcal{G}\ell(N,\mathbb{C})$ given by

$$\begin{aligned} u(N) &:= \left\{X \in \mathcal{G}\ell(N,\mathbb{C}) \; : \; X^{\dagger} = -X \right\} \\ su(N) &:= \left\{X \in \mathcal{G}\ell(N,\mathbb{C}) \; : \; X^{\dagger} = -X, \; \mathrm{tr}(X) = 0 \right\}, \end{aligned}$$

where "tr" stands for "trace". Therefore, elements of $u(N)$ are *anti-Hermitian matrices.*

Both $U(N)$ and $SU(N)$ are compact connected Lie groups and their exponential maps are onto. Hence, every unitary and special unitary matrix is obtained as the exponential of some element of the Lie algebra. Specifically, we have

$$\begin{aligned} U(N) &= \left\{\exp(X) \; : \; X \in \mathcal{G}\ell(N,\mathbb{C}), \; X^{\dagger} = -X \right\} \qquad \text{(A.26)} \\ SU(N) &= \left\{\exp(X) \; : \; X \in \mathcal{G}\ell(N,\mathbb{C}), \; X^{\dagger} = -X, \; \mathrm{tr}(X) = 0 \right\}. \end{aligned}$$

In physics, we are accustomed to work with Hermitian matrices rather than the anti-Hermitian matrices. This is the reason for the extra factor of "$i = \sqrt{-1}$" in physicists' definition of the exponential map. Equations (A.26) are often written in the following form:

$$\begin{aligned} U(N) &= \left\{\exp(iX) \; : \; X \in \mathcal{G}\ell(N,\mathbb{C}), \; X^{\dagger} = X \right\} \\ SU(N) &= \left\{\exp(iX) \; : \; X \in \mathcal{G}\ell(N,\mathbb{C}), \; X^{\dagger} = X, \; \mathrm{tr}(X) = 0 \right\}. \end{aligned}$$

$U(N)$ and $SU(N)$ can also be viewed as transformation groups acting on $\mathbb{C}^N$. The situation is similar to the action of $O(2)$ on $\mathbb{R}^2$. The group elements are (represented by) the $N\times N$ complex matrices that multiply the complex column vectors on the left. Let U be a unitary matrix and $\mathbf{z} \in \mathbb{C}^N$. Then, we have

$$\| U\mathbf{z} \|^2 = \mathbf{z}^{\dagger} U^{\dagger} U \mathbf{z} = \mathbf{z}^{\dagger}\mathbf{z} = \| \mathbf{z} \|^2 .$$

Hence, the unitary transformations preserve the (Euclidean) norm on $\mathbb{C}^N$.

Clearly, $U(N)$ is a subgroup of $U(N+1)$. This can be shown by representing elements of $U(N)$ by $(N+1)\times(N+1)$ matrices of the form

$$\begin{pmatrix} * & * & \cdots & * & 0 \\ * & * & \cdots & * & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ * & * & \cdots & * & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \qquad \text{(A.27)}$$

where "$*$" are complex numbers forming an $N \times N$ unitary matrix in the upper left block of (A.27). This is a trivial example of an $(N+1)$-dimensional complex "representation" of $U(N)$.

Definition 32: Let G be a group and V be a complex (real) vector space. The space of all invertible linear transformations of V is denoted by $GL(V)$. A function

$$\rho : G \longrightarrow GL(V),$$

is called a **complex (real) representation** of G, if for every $v \in V$ it satisfies:

1) $\rho(e)(v) = v$;
2) $\rho(g_1 \bullet g_2)(v) = \rho(g_1)\left(\rho(g_2)(v)\right) = \left[\rho(g_1) \circ \rho(g_2)\right](v)$,

where "e" is the identity element of G and "$\circ$" is the composition of linear transformations (multiplication of matrices).

In practice, we usually choose V to be either $\mathbb{C}^N$ or $\mathbb{R}^N$ (for the finite-dimensional representations). In this case, a representation maps the group elements to non-singular matrices. In particular, this allows us to study an abstract group in terms of a subgroup of $GL(N, \mathbb{C})$ or $GL(N, \mathbb{R})$. A representation is said to be a *faithful representation* if it is a one-to-one function.

Next let us consider a representation (ρ, V) of G. It is possible that all the elements of the $\rho(G)$ are represented by block diagonal matrices of the same form, with each block corresponding to a proper vector subspace of V. If this happens, then the representation is said to be a reducible representation. The precise definition is as follows:

Definition 33: Let (ρ, V) be a representation of a group G. Then a vector subspace V' of V is said to be an *invariant* subspace if for every $g \in G$ and every $v' \in V'$, $\rho(g)[v'] \in V'$, alternatively $\rho(G)[V'] \subseteq V'$. The representation (ρ, V) is said to be an **irreducible representation** of G if the only invariant subspaces of V are the trivial vector subspace $\{0\}$ and V itself. A representation which is not irreducible is called a **reducible** representation. It can be decomposed into irreducible representations.

In physics, we are often interested in the so-called *unitary irreducible representations* of Lie groups. A finite-dimensional unitary representation takes the group elements to unitary matrices. It turns out that if the group is non-compact there are no finite-dimensional unitary representations. However, in many cases the groups of interest are compact and we can represent them as subgroups of some unitary group. In fact, a result of group theory, namely the *Peter–Weyl theorem*, says that every compact connected Lie group is isomorphic to a subgroup of some unitary group.

Earlier in this section we emphasized the importance of the notion of the Lie algebra of a Lie group and mentioned that it is much easier to study the Lie algebras. This extends also to the subject of the representations of the Lie groups, in particular, the irreducible unitary representations.

Given an arbitrary representation (ρ, V) of a Lie group G, one can use the identification of the Lie algebra $\mathcal{G}$ of G with the tangent space at the identity T_eG, to obtain a representation of the Lie algebra.

Definition 34: Let $\mathcal{G}$ be an arbitrary abstract Lie algebra, V a vector space, and $\mathcal{G}\ell(V)$ be the vector space of all the linear transformations of V. Then a linear function,

$$\lambda : \mathcal{G} \longrightarrow \mathcal{G}\ell(V),$$

is said to be a **representation of the Lie algebra** $\mathcal{G}$ if for every $X, Y \in \mathcal{G}$,

$$\lambda([X,Y]) = \lambda(X) \circ \lambda(Y) - \lambda(Y) \circ \lambda(X), \tag{A.28}$$

where "[,]" and "$\circ$" denote the Lie algebra operation of $\mathcal{G}$ and the composition of linear functions of V, respectively. In particular, if (ρ, V) is a representation of a Lie group G and $\mathcal{G}$ is the Lie algebra of G, then the push-forward map:

$$\rho_* : T_eG = \mathcal{G} \longrightarrow \mathcal{G}\ell(V) = T_1GL(V),$$

defines a representation of $\mathcal{G}$. Here "1" stands for the identity operator on V.

Once a representation (λ, V) of a Lie algebra $\mathcal{G}$ is chosen, the elements of $\mathcal{G}$ and in particular its basis elements J_i, may be identified with some linear operators acting on V. These however may be composed. In fact the space $\mathcal{G}\ell(V)$ together with the operation of multiplication by numbers, addition of linear transformations, and their composition forms an associative algebra. This algebra which is also related with the Lie algebra structure of $\mathcal{G}\ell(V)$, with the Lie bracket defined by the right-hand side of (A.28), is an example of an *enveloping algebra.*

Definition 35: Let $(\mathcal{G}, [\ ,\ ])$ be an abstract Lie algebra with a basis $\{J_i\}$, $(\mathcal{A}, ., +, \otimes)$ be an associative algebra generated by $\{\mathcal{J}_j\}$ and $\tilde{\mathcal{G}}$ be the vector subspace of $\mathcal{A}$ spanned by $\{\mathcal{J}_j\}$. Then $(\mathcal{A}, ., +, \otimes)$ is said to be the **enveloping algebra** of $(\mathcal{G}, [\ ,\ ])$ and denoted by $\mathcal{E}(\mathcal{G})$, if there exists a vector space isomorphism

$$f : \mathcal{G} \to \tilde{\mathcal{G}},$$

and for all $X, Y \in \mathcal{G}$ the following condition is satisfied:

$$f([X,Y]) := f(X) \otimes f(Y) - f(Y) \otimes f(X).$$

More simply, one says that the enveloping algebra $\mathcal{E}(\mathcal{G})$ of a Lie algebra $\mathcal{G}$ is an associative algebra generated by a basis of $\mathcal{G}$.

Definition 36: An element of the enveloping algebra of a Lie algebra is called a **Casimir operator**, if it commutes with all the generators $\mathcal{J}_i$.

If $\mathcal{G}$ is the Lie algebra of a Lie group G which acts as a transformation group of a physical system, then the Casimir operators represent the invariant quantities. In quantum mechanics a symmetry is generated by a linear operator which commutes with the Hamiltonian. Hence, if G is the symmetry group, then the Hamiltonian must be a (representation of a) Casimir operator.

We end this appendix by emphasizing that the theory of group representations has played a substantial role in the development of quantum physics. In fact, many of the pioneering works in this subject were conducted by physicists such as Eugene Wigner.