

# **The Principle of the Most Elementary Particles**

Cai Xiaotian

Guangdong Province, China

Email: [caixt26@163.com](mailto:caixt26@163.com)

蔡孝天

中国广东省

## Abstract

This theory employs topology and differential geometry as mathematical tools to investigate the properties of particles. It first proposes two fundamental postulates, and based on these, integrates and develops special relativity, quantum mechanics, electromagnetic field theory, general relativity, particle theory, and cosmology, thereby forming a coherent theoretical system that completely describes particles—particularly a theory that elucidates the most fundamental properties and principles of particles. Taking these two postulates as logical starting points, the theory derives a large number of conclusions consistent with experimental facts and free of contradictions, including at least the following qualitative and quantitative results:

1. It explains the reason for the constancy of the speed of light and concludes that it is matter that expands, not empty space. The expansion process described by this theory agrees with astronomical observations, and it concludes that expansion is a geometric phenomenon rather than a mechanical one; hence, there is no need to invoke the hypothesis of dark energy to explain the cause or driving force of expansion.

2. The theory provides simple and elegant answers to formidable and profound problems. For example: Why does the  $z$ -component of electron spin have two possible values? Why does the  $z$ -component of neutrino spin have only one value? What is the nature of electric charge? Why are the charges of the electron and proton opposite? Why are photons and neutrinos electrically neutral? Why are neutrinos so difficult to detect? Why are light waves transverse rather than longitudinal? Why does a gravitational field exist? And so on.

3. It explains the basic principles of quantum mechanics using Lie group theory, providing an intuitive understanding of these principles and, at the same time, proving their validity.

4. It derives the quantitative relationships among the three mixing angles for neutrino oscillations, which are very close to experimental values, and shows that the cause of neutrino oscillations is not that neutrinos possess rest mass, but that the neutrino itself is a transformation group.

5. It explains the mechanisms that generate electromagnetic and gravitational fields, pointing out that they belong to different geometric structures.

6. It calculates the spin magnetic moments of the proton and neutron with very high precision. This result directly negates the existence of quarks and strongly supports the two fundamental postulates of this theory.

7. It explains the short-range mechanism underlying weak and strong interactions.

8. It gives a new definition of rest mass and analytically computes the rest mass ratio of the proton to the electron, achieving rather high accuracy.

9. It explains why particles can transform into one another and the internal mechanism of such transformations.

10. It provides a classification and description of the properties of elementary dark particles, as well as the mechanism by which all particles are born, and explains why there is so little antimatter in the universe.

11. It calculates the ratio of observable matter to dark matter in the universe, which is close to the observed value.

**Keywords:** elementary particles, Lie group, principal bundle, cosmic expansion, electromagnetic field, gravitational field, neutrino oscillation, spin magnetic moment, rest mass, dark particle.

## Preface

The development of physics has now extended its research into the realms of microscopic particles and the depths of the cosmos, achieving significant accomplishments, yet it has also encountered many formidable difficulties. This has prompted attempts to establish a theory of everything capable of explaining all observed physical phenomena, but such efforts have been fraught with challenges. For instance, Einstein sought to formulate an equation from which both his gravitational field equations and Maxwell's equations could be derived, yet after more than thirty years of endeavor, he did not succeed. Since Einstein's passing in 1955, attempts to develop a theory that uniformly describes gravitational, electromagnetic, weak, and strong interactions have likewise not succeeded. Efforts to unify quantum mechanics with general relativity have also not yet borne fruit.

The reasons for these failures are multifaceted. On one hand, there is the constraint of traditional notions, such as the belief that matter is infinitely divisible and that quarks exist within the proton. On the other hand, there is a growing tendency for theoretical research to diverge from physical reality. For example, based on elastic scattering experiments of electrons and protons, the assumption was made that quarks exist inside the proton. When experiments failed to detect free quarks, the hypothesis of permanent confinement was introduced—that quarks are forever bound and can never be observed—rather than questioning the initial assumption that quarks exist inside the proton. Furthermore, a misalignment between research objectives and physical phenomena is another important reason. Einstein, for instance, attempted to formulate a single equation that could uniformly describe gravity and electromagnetism. The prerequisite for such a unified description is that the essential properties of gravity and electromagnetism must be the same. Yet, clearly, their essential properties differ significantly, making it exceedingly difficult to unify them within a single equation. In fact, if gravity and electromagnetism shared the same essential properties, people would naturally or more readily describe them uniformly using one or a few equations based on long-standing empirical evidence.

Therefore, to achieve success, it is necessary to break free from traditional constraints, to adopt a correct research attitude, to ensure theoretical research remains closely aligned with physical reality, and to ensure that goals are commensurate with reality. In this theory, based on the fact that photons, electrons, protons, and neutrinos (and their antiparticles) do not decay while all other particles do, and that the ultimate decay products are precisely these eight stable particles, I put forward the first hypothesis: The non-decaying photons, electrons, protons, and neutrinos (and their antiparticles) are all elementary particles; particles that decay are composed of combinations of these decay products. Then, based on the fact that the spin of photons is 1, while the spin of electrons, protons, and neutrinos is  $1/2$ , I propose the second hypothesis: Photons correspond to an  $SO(3)$  group; electrons, protons, and neutrinos correspond to an  $SU(2)$  group. Are these two hypotheses correct? First, they are grounded in experimental facts; second, on the basis of these two hypotheses, special relativity, Maxwell's equations, quantum mechanics (including quantum field theory), and general relativity can be organically integrated, and the new conclusions derived are all consistent with experimental observations. This compellingly suggests that the two hypotheses are correct. Lie groups are composites of smooth manifolds and groups, possessing rich properties, while the properties of particles are also diverse. Thus, it is fitting to describe particles—with their diverse properties—using Lie groups, which themselves have rich properties. Attempting to encapsulate all the diverse properties of particles within a single equation or a few equations would inevitably lead to trade-offs and plunge research into a quagmire of difficulties. In contrast, by focusing on the spin properties of these eight non-decaying particles and formulating the hypotheses that they correspond to  $SO(3)$  and  $SU(2)$  groups—much like glimpsing the whole from a single part—it becomes possible to comprehensively describe all the properties of particles, naturally unifying existing theories and yielding numerous new conclusions.

This theory provides accessible and concise answers to dauntingly profound questions, such as: Why are there two possible  $z$ -axis components for electron spin? Why does the neutrino have only one possible  $z$ -axis spin component? What is the nature of electric charge? And so forth.

This theory concludes that matter expands, rather than that empty space itself expands, and

that this expansion is a geometric phenomenon rather than a mechanical one, thereby eliminating the need to introduce the hypothesis of dark energy to explain the cause of expansion.

The theory's calculations of the spin magnetic moments of the proton and neutron directly challenge the existence of quarks and strongly support the two fundamental hypotheses proposed herein.

The relationships among the three mixing angles for neutrino oscillations derived in this theory are in close agreement with experimental values.

The theory computes the proton-to-electron rest mass ratio with remarkably high precision.

The theory derives the ratio of observable matter to dark matter in the universe, which closely matches observational values.

These quantitative results confirm the validity of the two hypotheses proposed in this theory.

This paper is written in a textbook style, aiming to be accessible and lower the threshold for learning. It is divided into two parts. The first part covers the mathematical foundations, and the second part covers the physical content. The first part consists of Chapters 1 through 12, which present the mathematical knowledge necessary for reading the physical part of the paper, primarily covering topics such as topology, differential geometry, Riemannian geometry, and Lie group theory. In principle, all knowledge from these disciplines could be applied to the research in this theory; however, constrained by space and time, only the relevant definitions, formulas, theorems, and other conclusive content from these mathematical fields are presented. The derivations or proofs of most formulas and theorems are not provided. Readers interested in the proofs of the formulas and theorems listed in the paper may consult the mathematical references provided at the end. Because these modern mathematical theories are highly abstract, I have attempted to illustrate them with examples wherever possible. The physical part consists of Chapters 13 through 21. The content of the physical part is not detailed here, but readers can get a general idea by browsing the table of contents. Again, constrained by space and time, only the most fundamental conclusions are presented. Readers are encouraged to build upon these foundational conclusions to conduct further research and obtain more new results.

Due to the author's limited expertise, comments and criticisms from readers are warmly welcomed.

Cai Xiaotian

June 24, 2024



## Conventions

1. In the mathematical part, the notation follows mathematical conventions: English letters are used as indices, and smooth tangent vector fields are denoted by  $X$ , etc. In the physical part, the notation follows physical conventions: Greek letters are used as indices, and operators or smooth tangent vector fields are denoted by  $\hat{X}$ , etc.

2. The inner product of two smooth tangent vector fields  $X$  and  $Y$  is denoted by  $(X, Y)$ , or by  $\langle X, Y \rangle$ , or by a dot  $\cdot$ ; i.e.,  $X \cdot Y$ .

3. In mathematics texts, results are often categorized as lemmas, theorems, corollaries, propositions, etc. In this paper, such distinctions are not made; all are uniformly referred to as theorems.

4. In this paper, the signature of the metric for the gravitational field is taken to be +2, i.e.,  $(-, +, +, +)$ .

5. This paper mainly adopts the International System of Units (SI). Only in a few instances is the Gaussian unit system used, and in those cases, explicit indication is provided.

# Part One Foundations of Mathematics

## Contents

<b>Chapter 1 Foundations of Topology</b>	12
§1.1 Set	12
1.Power set	12
2.Direct product of sets	12
3.Equivalence relation	12
4.Quotient Set	13
5.Mapping of Sets	13
6.Cardinality of a Set	14
§1.2 Metric Space and Euclidean Space	15
1.Metric Space	15
2.Euclidean Space	15
§1.3 Topology and Topological Space	16
1.Topology	16
2.Topological Space	16
3.Topological Basis	17
§1.4 Cover	17
1.Open Cover	17
2.Finite Subcover	17
3.Compact Space	18
§1.5 Connected Space	18
1.Complement	18
2.Closed Set	18
3.Continuous Mapping	19
4.Connected Space	19
5.Path-connected	19
6Simply connected	20
§1.6 Homeomorphic Mapping	21
§1.7 Hausdorff Space	22
§1.8 Quotient space	22
§1.9 Product of Topological Spaces	23
§1.10 Euler characteristic	24
§1.11 Group	25
1.Group	25
2.Subgroup	26
3.Invariant Subgroup	26
4.Quotient Group	26
<b>Chapter 2 Tensors as Invariants Under Coordinate Transformations</b>	27
§2.1 Coordinate Transformation	27

1. Transformation of coordinate differentials.....	27
2. Einstein Summation Convention.....	27
3. Jacobian Matrix.....	27
§2.2 Definition of Tensors.....	28
1. Zero-Order Tensor.....	28
2. Contravariant Tensor.....	28
3. Covariant Tensor.....	30
4. Mixed Tensor.....	31
§2.3 Symmetry and Antisymmetry of Tensors.....	32
§2.4 Algebraic Operations on Tensors.....	32
1. Equality of Tensors.....	32
2. Zero Tensor.....	32
3. Addition and Subtraction of Tensors.....	32
4. Product of Tensors.....	33
5. Contraction of Tensors.....	33
6. Inner Product of Tensors.....	34
§2.5 Flexible Use of Tensors.....	34
<b>Chapter 3 Differentiable Manifolds.....</b>	<b>36</b>
§3.1 Functions in Euclidean Space.....	36
1. $k$ -times Differentiable Functions.....	36
2. Smooth Functions.....	36
3. Analytic Functions.....	36
§3.2 Manifolds.....	37
1. Topological Manifold.....	37
2. Introducing Coordinates on a Manifold.....	37
3. Smooth Manifold.....	39
§3.3 Real Projective Space.....	40
§3.4 Product Manifolds.....	43
<b>Chapter 4 Tangent Vectors and Cotangent Vectors.....</b>	<b>45</b>
§4.1 Smooth Functions on Manifolds.....	45
§4.2 Tangent Vectors and Tangent Spaces.....	45
1. Parametric Curves on a Manifold.....	45
2. Tangent Vector.....	46
3. Tangent Space.....	46
4. Tangent Vector of a Curve.....	47
§4.3 Cotangent Vectors and Cotangent Spaces.....	49
1. Cotangent Vector.....	49
2. Basis of the Cotangent Space.....	50
3. Tangent Bundle and Cotangent Bundl.....	51
§4.4 Transformation of Vectors.....	51
<b>Chapter 5 Mappings and Submanifolds.....</b>	<b>53</b>
§5.1 Mappings.....	53
1. Smooth Mappings.....	53
2. Cotangent Map.....	54

3. Tangent Map.....	55
§5.2 Submanifolds.....	58
1. Immersed Submanifolds and Embedded Submanifolds.....	58
2. Open Submanifold.....	59
3. Closed Submanifold.....	59
4. Regular Submanifold.....	60
§5.3 Tangent Space of a Product Manifold.....	61
1. Direct Sum of Linear Spaces.....	61
2. Tangent Space of a Product Manifold.....	62
<b>Chapter 6 Smooth Tangent Vector Fields.....</b>	<b>63</b>
§6.1 Tangent Vector Fields and Smooth Tangent Vector Fields.....	63
1. Tangent Vector Field.....	63
2. Smooth Tangent Vector Field.....	63
3. Parallelizability.....	64
§6.2 Integral Curves.....	64
§6.3 Poisson Bracket.....	66
1. Poisson Bracket.....	66
2. Algebraic Properties of the Poisson Bracket.....	67
3. Geometric Interpretation of the Poisson Bracket.....	67
§6.4 Mapping of Smooth Tangent Vector Fields.....	69
§6.5 Coordinate Basis.....	70
§6.6 One-Parameter Differentiable Transformation Groups.....	71
§6.7 Lie Derivative.....	74
§6.8 Singular Points.....	77
§6.9 Phase Plane.....	78
1. Phase Plane.....	78
2. Singular Points of the System.....	78
3. Types of Singular Points.....	79
<b>Chapter 7 Tensors as Linear Operators.....</b>	<b>85</b>
§7.1 Vector Spaces.....	85
§7.2 Dual Space.....	86
§7.3 Linear Mappings.....	86
§7.4 Tensors.....	87
§7.5 Tensor Product.....	88
§7.6 Tensor Fields.....	89
<b>Chapter 8 Differential Forms.....</b>	<b>92</b>
§8.1 Permutations and Permutation Groups.....	92
§8.2 Symmetric and Antisymmetric Covariant Tensors.....	93
§8.3 Symmetrization and Antisymmetrization Operators.....	94
§8.4 Exterior Product.....	95
§8.5 Exterior Differentiation.....	96
1. Differential Forms.....	96
2. Wedge Product of Differential Forms.....	98
3. Exterior Differentiation of Differential Forms.....	98

4.Pullback of Differential Forms.....	100
5.Closed Differential Forms and Exact Differential Forms.....	102
§8.6 Interior Product.....	102
<b>Chapter 9 Orientation and Integration on Smooth Manifolds.....</b>	<b>104</b>
§9.1 Conditions for Orientability.....	104
1.Orientation of a Vector Space.....	104
2.Conditions for Orientability.....	104
§9.2 Methods of Orientation.....	106
1.Orientation via Coordinate Systems.....	106
2.Orientation by an $m$ -Form.....	106
3.Orientation of Hypersurfaces.....	108
§9.3 Types of Orientations.....	109
§9.4 Preservation and Reversal of Orientation.....	109
§9.5 Orientation of Product Manifolds.....	110
§9.6 Integration of Differential Forms on Manifolds.....	112
§9.7 Degree of a Map and Brouwer Degree.....	114
<b>Chapter 10 Affine Connections.....</b>	<b>120</b>
§10.1 Affine Connections.....	120
§10.2 Geodesics.....	124
§10.3 Curvature and Torsion.....	125
1.Curvature Tensor.....	125
2.Torsion Tensor.....	127
<b>Chapter 11 Riemannian Manifolds.....</b>	<b>129</b>
§11.1 Metric Tensor.....	129
1.Metric Tensor and Riemannian Metric.....	129
2.Levi-Civita Connection.....	131
3.A Simplified Method for Calculating Christoffel Symbols.....	132
4.Koszul Formula.....	132
5.Exponential Map.....	132
§11.2 Curvature.....	133
1.Curvature Tensor.....	133
2.Sectional Curvature.....	134
3.Ricci Curvature.....	136
4.Scalar Curvature.....	137
5.Einstein Tensor.....	137
6.Einstein Manifolds.....	138
7.Locally Euclidean Spaces.....	139
§11.3 Gauss–Bonnet Theorem.....	139
§11.4 Isometries and Conformal Transformations.....	140
1.Isometries.....	140
2.Conformal Transformations.....	140
§11.5 Riemannian Symmetric Spaces.....	141
§11.6 Divergence of Tangent Vector Fields.....	141
§11.7 Completeness and the Bonnet–Myers Theorem.....	143

§11.8	Submanifolds in Euclidean Space.....	145
§11.9	Induced Metrics and Space Forms.....	145
1.	Induced Metric.....	145
2.	Induced Metric of $S^3$ in $R^4$ .....	146
3.	Metrics on Open Balls.....	147
4.	Space Forms.....	147
§11.10	Products of Riemannian Manifolds.....	148
§11.11	Spherical Trigonometry.....	150
1.	Cosine Theorem for Sides of a Spherical Triangle.....	150
2.	General Cosine Theorem.....	151
3.	Formulas for a Spherical Right Triangle.....	151
§11.12	Toponogov Comparison Theorem.....	151
§11.13	Hodge Star Operator.....	153
§11.14	Jacobi Fields.....	157
<b>Chapter 12</b>	<b>Lie Groups</b> .....	160
§12.1	Topological Groups.....	160
§12.2	Lie Groups and Lie Algebras.....	160
1.	Lie Groups.....	160
2.	Right and Left Translations.....	160
3.	Lie Algebras.....	161
4.	The Tangent Bundle of a Lie Group is Trivial.....	164
§12.3	Local Lie Groups.....	164
§12.4	Exponential Matrix.....	167
§12.5	Homomorphisms and Subgroups of Lie Groups.....	167
1.	Homomorphisms and Isomorphisms of Lie Groups.....	167
2.	Homomorphisms and Isomorphisms of Lie Algebras.....	168
3.	Lie Subgroups.....	168
4.	One-Parameter Subgroups and the Exponential Map.....	169
5.	Sophus Lie's Fundamental Theorems.....	170
§12.6	Lie Transformation Groups.....	170
1.	General Lie Transformation Groups.....	170
2.	Homogeneous Spaces.....	172
3.	Orbits.....	172
4.	Inner Automorphisms.....	172
5.	Adjoint Representation.....	173
§12.7	Orientation of Lie Groups.....	174
§12.8	Bi-invariant Metric Tensors.....	175
§12.9	$SO(3)$ Group.....	176
1.	$SO(3)$ Group as the Special Orthogonal Group.....	176
2.	$SO(3)$ Group as the Real Rotation Group.....	176
3.	Euler Angle Parameterization.....	177
§12.10	$SU(2)$ Group.....	179
1.	General Form of the $SU(2)$ Group.....	179

2	$SU(2)$ Group is Homeomorphic to the Three-Dimensional Unit Sphere.....	180
3	A Real Parameterization of the $SU(2)$ Group.....	180
§12.11	Homomorphism between the $SO(3)$ Group and the $SU(2)$ Group.....	181
§12.12	Direct Product of Lie Groups.....	185
1.	Direct Product Group.....	185
2.	Direct Product of Lie Groups.....	186
3.	Direct Sum of Lie Algebras.....	186
4.	$SO(4, \mathbb{R})$ Group.....	187
5.	Connection of a Direct-Product Lie Group.....	187
6.	Curvature Tensor of a Direct-Product Lie Group.....	188
7.	Ricci Tensor of the Direct-Product Lie Group.....	189
8.	Scalar Curvature of a Direct-Product Lie Group.....	189
9.	Internal Direct Product of Groups.....	190
§12.13	Principal Bundles.....	190
1.	Differentiable Fiber Bundles.....	190
2.	Vector Bundles and Tangent Bundles.....	195
3.	Principal Bundles.....	195
4.	Connection and Curvature of the Base Manifold.....	196
5.	Connection and Curvature of a Principal Bundle.....	197

# Chapter 1 Foundations of Topology

## §1.1 Set

### 1.Power set

**Theorem 1.1.1** If a set  $S = \{a_1, a_2, \dots, a_m\}$  is a finite set with  $m$  elements, then there are  $2^m$  subsets in total.

**Example 1.1.1** Set  $S = \{1, 2\}$  consists of 2 elements, so the total number of subsets of set  $S$  is  $2^2 = 4$ . They are  $\{1\}$ ,  $\{2\}$ ,  $\{1, 2\}$  and the empty set  $\emptyset$ .

Based on this theorem, we denote the set of all subsets of set  $S$  as  $2^S$ , which is called the power set of set  $S$ .

### 2.Direct product of sets

**Definition 1.1.1** Let  $A$  and  $B$  be two sets. The set of all ordered pairs

$$(a, b), \quad a \in A, \quad b \in B$$

is called the **direct product** of  $A$  and  $B$ , denoted by  $A \times B$ ,

$$A \times B = \{(a, b) | a \in A, b \in B\}.$$

The direct product of sets is also called the **Cartesian product** of sets. The direct product  $A \times B$  of sets  $A$  and  $B$  is also a set. The elements of this set are ordered pairs  $(a, b)$ , where the first element  $a$  belongs to set  $A$  and the second element  $b$  belongs to set  $B$ .

**Example 1.1.2** Suppose set  $A = \{1, 2\}$ ,  $B = \{a, b\}$ , then

$$A \times B = \{(1, a), (1, b), (2, a), (2, b)\}.$$

The direct product set  $A \times B$  of two elements  $(a, b)$  and  $(c, d)$  being equal implies

$$(a, b) = (c, d) \Leftrightarrow a = c, b = d.$$

The direct product of  $n$  sets  $A_1, A_2, \dots, A_n$  can be denoted as

$$A_1 \times A_2 \times \dots \times A_n \quad \text{or} \quad \prod_{i=1}^n A_i.$$

Its elements consist of  $n$  ordered elements:  $(a_1, a_2, \dots, a_n)$ ,  $\forall a_i \in A_i$ , which can be written as

$$A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n) | a_1 \in A_1, a_2 \in A_2, \dots, a_n \in A_n\}.$$

### 3.Equivalence relation

**Definition 1.1.2:** Let  $R$  be a binary relation denoted as “ $\sim$ ” on a set  $A$ . If  $R$  satisfies the following conditions, then  $R$  is called an equivalence relation defined on the set  $A$ .

- 1) Reflexivity if  $a \in A$ ,  $aRa$ , which is denoted as  $a \sim a$ ;
- 2) Symmetry if  $a, b \in A$  and  $aRb, bRa$ , namely if  $a \sim b, b \sim a$ ;
- 3) Transitivity if  $a, b, c \in A$  and  $aRb$  and  $bRc$ ,  $aRc$ , namely if  $a \sim b$  and  $b \sim c, a \sim c$ .

**Example 1.1.3** There is a set  $A = \{\text{carrot, coal, red paper, black skirt, red hat, ink}\}$ . In this set, an equivalence relation  $R$  is defined as the "same color" relation. According to this equivalence relation, a carrot is equivalent to a carrot, denoted as: carrot  $\sim$  carrot. A carrot is equivalent to red paper, denoted as: carrot  $\sim$  red paper. Red paper is equivalent to the red hat, denoted as: red paper  $\sim$  red hat. By transitivity, the carrot is equivalent to the red hat, denoted as: carrot  $\sim$  red hat. Similarly, coal is equivalent to the black skirt, denoted as coal  $\sim$  black skirt.



Carrot, red paper, and red hat belong to the same equivalence class; coal, black skirt, and ink belong to another equivalence class.

#### 4. Quotient Set

**Definition 1.1.3** Let  $\sim$  be an equivalence relation defined on a set  $A$ . The family of all distinct equivalence classes of  $A$  under  $\sim$  is called the quotient set of  $A$  with respect to the equivalence relation  $\sim$ , denoted as  $A/\sim$ .

**Example 1.1.4** In Example 1.1.3, we group the carrot, red paper, and red hat into one class, denoted as  $[a_1] = \{\text{carrot, red paper, red hat}\}$ , and treat  $[a_1]$  as an element of a certain set  $A/\sim$ . Group coal, black skirt, and ink into one class, denoted as  $[a_2]$ , and treat  $[a_2]$  as another element of set  $A/\sim$ . The set  $A/\sim$  contains two elements,  $[a_1]$  and  $[a_2]$ , denoted as

$$A/\sim = \{[a_1], [a_2]\}.$$

The set  $A/\sim$  is called the quotient set of set  $A$ . The quotient set  $A/\sim$  is a new set generated by classifying set  $A$  according to the equivalence relation  $\sim$ .

#### 5. Mapping of Sets

**Definition 1.1.4** Let  $A$  and  $B$  be two non-empty sets. If there exists a rule of correspondence  $f$  such that for every element  $x$  in set  $A$ , there is a unique element  $y$  in set  $B$  corresponding to it, then  $f$  is called a **mapping** from  $A$  to  $B$ , denoted as  $f: A \rightarrow B$ . The relationship between corresponding elements is written as  $y = f(x)$ . Set  $A$  is called the **domain** of  $f$ , and  $f(A) = \{f(x) | x \in A\}$  is the **range** of  $f$ .

**Definition 1.1.5** For a mapping  $f: A \rightarrow B$ , if for every element  $y$  in set  $B$ , there exists at least one element  $x$  in  $A$  such that  $f(x) = y$ , then  $f$  is called a **surjective (onto) mapping**.

When the mapping  $f$  is surjective, the range of  $f$  is the entire set  $B$ ,  $B = f(A) = \{f(x) | x \in A\}$ .

If for any two distinct elements  $x_1$  and  $x_2$  in set  $A$ , we have  $f(x_1) \neq f(x_2)$ , then  $f$  is called an **injective (one-to-one) mapping**.

If  $f$  is both surjective and injective, then  $f$  is called a **bijective mapping** (or a **one-to-one correspondence**).

A bijection establishes a one-to-one correspondence between the elements of the two sets  $A$  and  $B$ .

**Theorem 1.1.2** If  $A$  and  $B$  are finite sets, both containing  $m$  elements, then for the mapping  $f: A \rightarrow B$ , the following relationships hold:

$$f \text{ is injective} \Leftrightarrow f \text{ is bijective} \Leftrightarrow f \text{ is surjective.}$$

**Theorem 1.1.3** For the mapping  $f: A \rightarrow B$ , if  $A_1 \subset A$  and  $A_2 \subset A$ , then

**Theorem 1.1.4** For the mapping  $f: A \rightarrow B$ , if  $A_1 \subset A$  and  $A_2 \subset A$ , then

$$f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2).$$

If  $f$  is injective, then

$$f(A_1 \cap A_2) = f(A_1) \cap f(A_2).$$

**Definition 1.1.6** If  $f: A \rightarrow B$  is a bijection, then for every element  $b \in B$  in set  $B$ , there exists a unique element  $a \in A$  in set  $A$  such that  $f(a) = b$ . Let  $f^{-1}(b) = a$ , then the mapping  $f^{-1}: B \rightarrow A$  is called the **inverse mapping** of  $f$ .

**Theorem 1.1.5** If the mapping  $f: A \rightarrow B$  is a bijection, then its inverse mapping  $f^{-1}: B \rightarrow A$  is also a bijection.

**Definition 1.1.7** If  $A_1 \subset A$ , then the set  $f(A_1) = \{f(x) | x \in A_1\}$  is called the **image** of  $A_1$ . If  $B_1 \subset B$ , then the set  $f^{-1}(B_1) = \{x \in A | f(x) \in B_1\}$  is called the **preimage** (or **inverse image**) of  $B_1$ .

If the mapping  $f$  is injective, then for each element  $y$  in set  $B$ , the preimage  $f^{-1}(y)$  of  $y$  contains at most one element.

**Definition 1.1.8** Given mappings  $f: A \rightarrow B$  and  $g: B \rightarrow C$ , the expression  $g(f(x))$  defines the **composite mapping**  $g \circ f: A \rightarrow C$ .

**Theorem 1.1.6** For the mappings  $f: A \rightarrow B$ ,  $g: B \rightarrow C$ , and the composite mapping  $g \circ f: A \rightarrow C$ , the following relations hold:

- 1) If  $f$  and  $g$  are injective, then  $g \circ f$  is also injective;
- 2) If  $f$  and  $g$  are surjective, then  $g \circ f$  is also surjective;
- 3) If  $f$  and  $g$  are bijective, then  $g \circ f$  is also bijective.

**Definition 1.1.9** The mapping  $f$ , which keeps every element of set  $S$  unchanged, i.e., for any  $a \in S$ , we have  $f(a) = a$ , is called the **identity mapping** or the identity map, denoted as  $id_S: S \rightarrow S$ . It is often simply written as  $id$ .

**Definition 1.1.10** Let  $A \subset S$ . If there exists a mapping  $i: A \rightarrow S$  such that for  $\forall x \in A$ ,  $i(x) = x$ , then the mapping  $i$  is called the **inclusion mapping**.

**Example 1.1.5** The set of integers is  $S = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$  and the set of natural numbers is  $A = \{0, 1, 2, 3, \dots\}$ . Define the mapping  $i(x) = x$ ,  $\forall x \in A$ , then the mapping  $i: A \rightarrow S$  is an inclusion mapping.

## 6. Cardinality of a Set

**Definition 1.1.11** Let  $A$  and  $B$  be two sets. If there exists a bijection from  $A$  to  $B$ , then sets  $A$  and  $B$  are said to be **equipotent**, denoted as  $A \sim B$ .

**Theorem 1.1.7** The equipotence of sets satisfies:

- 1) Reflexivity  $A \sim A$ ;
- 2) Symmetry if  $A \sim B$ ,  $B \sim A$ ;
- 3) Transitivity if  $A \sim B$  and  $B \sim C$ ,  $A \sim C$ .

**Definition 1.1.12** If sets  $A$  and  $B$  are equipotent, then they are said to have the same cardinality (or power), denoted as  $\overline{A} = \overline{B}$ .

**Definition 1.1.13** A set containing a finite number of elements is called a **finite set**, and its cardinality is defined as the number of its elements. A set that is equipotent to the set of natural numbers is called a **countable set**, and its cardinality is defined as  $\aleph_0$  ( $\aleph$  is read as "aleph," and  $\aleph_0$  is read as "aleph-null"). A set containing infinitely many elements that is not equipotent to the set of natural numbers is called an **uncountable set**.

Both the set of natural numbers and the set of real numbers contain infinitely many elements, but the elements of the natural number set are countable or enumerable, while the elements of the real number set are uncountable or non-enumerable; they are different in this regard.

**Theorem 1.1.8** If sets  $A_1, A_2, \dots, A_k$  are all countable sets, then the direct product set

$$A_1 \times A_2 \times \dots \times A_k$$

is also countable.

**Theorem 1.1.9** The closed interval  $[0, 1]$  is an uncountable set.

**Definition 1.1.14** A set that is equipotent to the closed interval  $[0, 1]$  is called a **continuum**, and its cardinality is denoted as  $\aleph$ .

**Theorem 1.1.10** 1)  $[0, 1] \sim (0, 1)$ ;

2)  $R \sim (0, 1)$ ;

3) Any interval containing infinitely many elements has cardinality  $\aleph$ .

**Theorem 1.1.11** The direct product of a finite number of continua is still a continuum.

A straight line is a set equipotent to the open interval  $(0, 1)$  and is a continuum. According to Theorem 1.1.11, the direct product of two straight lines forms a plane, so a plane is also a continuum, and its cardinality is equal to that of a straight line. The direct product of a straight line and a plane is a three-dimensional Euclidean space, which is also a continuum, and its cardinality is likewise equal to that of a straight line. Similarly, an  $m$ -dimensional Euclidean space  $R^m$  is

also a continuum, and its cardinality is equal to that of a straight line. All are equal to  $\aleph$ .

## §1.2 Metric Space and Euclidean Space

### 1. Metric Space

**Definition 1.2.1** Let  $S$  be a set. If for any  $x, y, z \in S$  a mapping  $d : S \times S \rightarrow R$  satisfies the following conditions, then the function  $d$  is called a **metric** on the set  $S$ :

- 1) Positive Definiteness  $d(x, y) \geq 0$ , and  $d(x, y) = 0$  if and only if  $x = y$ ;
- 2) Symmetry  $d(x, y) = d(y, x)$ ;
- 3) Triangle Inequality  $d(x, z) \leq d(x, y) + d(y, z)$ .

The set  $S$  together with the mapping  $d$  is called a **metric space**, denoted as  $(S, d)$ .

**Example 1.2.1** A distance function can be defined on the closed interval  $[0, 1]$ . Let  $a, b \in [0, 1]$ , and define

$$d(a, b) = \sqrt{(a - b)^2}.$$

It is also possible to define

$$d(a, b) = (a - b)^2.$$

It is also possible to define

$$d(a, b) = |a - b|.$$

It can be verified that all definitions satisfy the conditions mentioned above. Therefore, either can serve as the distance function for the closed interval  $[0, 1]$ , making  $[0, 1]$  a metric space. This demonstrates that different distance functions can be defined for the same metric space.

### 2. Euclidean Space

We denote by  $R$  the **field of real numbers**, i.e., the set of all real numbers.

The set of all ordered  $m$ -tuples of real numbers is denoted as  $R^m$ , i.e.,

$$R^m = \{x = (x^1, \dots, x^m) \mid x^i \in R, 1 \leq i \leq m\}.$$

The real number  $x^i$  is called the  $i$ -th coordinate of the point  $x \in R^m$ .

For any  $x, y \in R^m$ ,  $a \in R$ , define

addition:  $(x + y)^i = x^i + y^i$ ;

multiplication by a real number:  $(ax)^i = ax^i$ .

This makes  $R^m$  an  $m$ -dimensional vector space over the real number field  $R$ .

In  $R^m$ , the following metric (distance function) is defined:

$$d(x, y) = \sqrt{\sum_{i=1}^m (x^i - y^i)^2}, \quad (1.2.1)$$

where  $x, y \in R^m$ . Then  $R^m$  becomes a metric space.

**Definition 1.2.2** The  $m$ -dimensional vector space  $R^m$ , equipped with the distance function defined by Equation (1.2.1), is called the  $m$ -dimensional **Euclidean space**.

The three-dimensional Euclidean space  $R^3$  is the familiar space we know. When  $m=1$ , it is the one-dimensional space, which is simply the set of real numbers. When  $m=2$ , it is the two-dimensional plane. A point in  $R^m$  can be represented by  $m$  real numbers  $(x^1, \dots, x^m)$ , so  $R^m$  can be viewed as a set consisting of infinitely many ordered  $m$ -tuples of real numbers  $(x^1, \dots, x^m)$ .  $R^m$  is flat: in  $R^m$ , two lines can be parallel to each other; the sum of the interior angles of a triangle equals  $\pi$  ( $180^\circ$ ); and so on.

## §1.3 Topology and Topological Space

### 1. Topology

**Definition 1.3.1** Let  $S$  be a nonempty set, and let  $\tau$  be a collection of subsets of  $S$  that satisfies the following conditions:

- 1)  $\emptyset \in \tau$ ,  $S \in \tau$ , (where  $\emptyset$  denotes the empty set);
- 2) The intersection of any finite number of elements in  $\tau$  remains in  $\tau$ , i.e., if  $U_1, U_2, \dots, U_m \in \tau$ , then  $\bigcap_{i=1}^m U_i \in \tau$ ;
- 3) The union of any number of elements in  $\tau$  remains in  $\tau$ , i.e., if  $U_1, U_2, \dots, U_m \in \tau$ , then  $\bigcup_i U_i \in \tau$ .

Then  $\tau$  is called a **topology** on  $S$ .

### 2. Topological Space

**Definition 1.3.2** If a set  $S$  is equipped with a topology  $\tau$ , then  $S$  is called a **topological space**, denoted as  $(S, \tau)$ . The elements of  $\tau$  are called the **open sets** of the topological space  $S$ .

For simplicity, we often omit  $\tau$  and simply refer to  $S$  as a topological space.

**Example 1.3.1** Consider a set  $S = \{1, 2\}$ . The collection  $\tau = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$  consisting of all subsets of  $S$  satisfies the three conditions in Definition 1.3.1, as verified below:

- 1)  $\emptyset \in \tau$ ,  $S \in \tau$ ;
- 2)  $\{1\} \cap \{2\} = \emptyset \in \tau$ ,  $\{1\} \cap \{1, 2\} = \{1\} \in \tau$ ,  $\{1\} \cap \emptyset = \emptyset \in \tau$ ,  
 $\{2\} \cap \{1, 2\} = \{2\} \in \tau$ ,  $\{2\} \cap \emptyset = \emptyset \in \tau$ ,  $\{1, 2\} \cap \emptyset = \emptyset \in \tau$ ,  $\{1\} \cap \{2\} \cap \{1, 2\} = \emptyset \in \tau$ ,  
 $\{1\} \cap \{2\} \cap \emptyset = \emptyset \in \tau$ ,  $\{2\} \cap \{1, 2\} \cap \emptyset = \emptyset \in \tau$ ,  $\{1\} \cap \{2\} \cap \{1, 2\} \cap \emptyset = \emptyset \in \tau$ ;
- 3)  $\{1\} \cup \{2\} = \{1, 2\} \in \tau$ ,  $\{1\} \cup \{1, 2\} = \{1, 2\} \in \tau$ ,  $\{1\} \cup \emptyset = \{1\} \in \tau$ ,  
 $\{2\} \cup \{1, 2\} = \{1, 2\} \in \tau$ ,  $\{2\} \cup \emptyset = \{2\} \in \tau$ ,  $\{1, 2\} \cup \emptyset = \{1, 2\} \in \tau$ ,  
 $\{1\} \cup \{2\} \cup \{1, 2\} = \{1, 2\} \in \tau$ ,  $\{1\} \cup \{2\} \cup \emptyset = \{1, 2\} \in \tau$ ,  
 $\{2\} \cup \{1, 2\} \cup \emptyset = \{1, 2\} \in \tau$ ,  $\{1\} \cup \{2\} \cup \{1, 2\} \cup \emptyset = \{1, 2\} \in \tau$ .

Therefore, the collection  $\tau$  can serve as a topology on the set  $S$ .

In general, the collection  $\tau$  consisting of all subsets of  $S$  can be taken as a topology on  $S$ , known as the **discrete topology**.

**Example 1.3.2** For any nonempty set  $S$ , the collection  $\tau = \{\emptyset, S\}$  can serve as a topology on  $S$ . Readers may verify this themselves. This topology is called the **indiscrete** (or **trivial**) **topology**.

Thus, a nonempty set  $S$  can have multiple topologies.

**Example 1.3.3** In the  $m$ -dimensional Euclidean space  $R^m$ , an **open ball** centered at any point  $x_0 \in R^m$  with an arbitrary radius  $r > 0$ ,

$$B(x_0, r) = \{x \in R^m \mid d(x - x_0) < r, x_0 \in R^m\}$$

is a subset of  $R^m$ . An open ball  $B(x_0, r)$  is an open set. When  $m=1$  (i.e., in  $R$ ), an open ball is an open interval. When  $m=2$  (i.e., in the plane  $R^2$ ), an open ball is an open disk. The collection

$$\tau = \{\emptyset, \text{subsets can be expressed as a union of open balls in } R^m\}$$

can be taken as a topology on  $R^m$ , making  $R^m$  a topological space.

A **closed solid ball** is defined as

$$B^m(x_0, r) = \{x \in R^m \mid d(x - x_0) \leq r, x_0 \in R^m\}.$$

**Definition 1.3.3** Let  $(S, \tau)$  be a topological space, and let  $X \subset S$ . Construct the collection

$$\tau' = \{X \cap U_i \mid U_i \in \tau\},$$

then the topological space  $(X, \tau')$  is called a **topological subspace** of  $(S, \tau)$ , and the topology  $\tau'$  is called the **induced topology** (or **subspace topology**) inherited from  $\tau$ .

**Example 1.3.4** In Example 1.3.1, consider the topological space  $(S, \tau)$  where the set  $S = \{1, 2\}$  and the topology is  $\tau = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$ . Let  $X = \{1\}$ . Then

$$X \cap U_1 = \{1\} \cap \{1\} = \{1\}, \quad X \cap U_2 = \{1\} \cap \{2\} = \emptyset,$$

$$X \cap U_3 = \{1\} \cap \{1, 2\} = \{1\}, \quad X \cap U_4 = \{1\} \cap \emptyset = \emptyset.$$

Therefore,  $\tau' = \{\{1\}, \emptyset\}$ . If we denote  $\tau' = \{\{1\}, \emptyset\}$  as the induced topology on  $X$ , then the topological space  $(X, \tau')$  is a topological subspace of  $(S, \tau)$ .

### 3. Topological Basis

**Definition 1.3.4** Let  $(S, \tau)$  be a topological space, and let  $\mathcal{E}$  be a family of subsets of  $\tau$ . If every open set  $U \in \tau$  can be expressed as a union of some sets from  $\mathcal{E}$ , then  $\mathcal{E}$  is called a **basis for the topology**  $\tau$ .

**Example 1.3.5** In Example 1.3.1, the topological space  $(S, \tau)$  has the topology  $\tau = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$ . Consider the subfamily  $\mathcal{E} = \{\{1\}, \{2\}, \emptyset\}$  of  $\tau$ . Every open set  $U \in \tau$  can be expressed as a union of some sets in  $\mathcal{E}$ :

$$\{1\} = \{1\} \cup \emptyset, \quad \{2\} = \{2\} \cup \emptyset, \quad \{1, 2\} = \{1\} \cup \{2\}, \quad \emptyset = \emptyset \cup \emptyset.$$

Therefore,  $\mathcal{E}$  is a basis for the topology  $\tau$ .

**Second Countability Axiom:** A topological space has a countable basis.

**Definition 1.3.5** If a topological space  $(S, \tau)$  has a countable basis, then  $S$  is said to **satisfy the second countability axiom**.

**Example 1.3.6** In Example 1.3.5, the topological basis  $\mathcal{E} = \{\{1\}, \{2\}, \emptyset\}$  is countable, so the topological space  $(S, \tau)$  in that example satisfies the second countability axiom.

## §1.4 Cover

### 1. Open Cover

**Definition 1.4.1** Let  $A$  and  $S$  be two sets, and let  $\{A_j\}$  be a collection of subsets of  $A$ . If  $S \subset \bigcup_j A_j$ , then  $\{A_j\}$  is called a **cover** of  $S$ , or  $\{A_j\}$  covers  $S$ . If every element of  $\{A_j\}$  is an open set, then  $\{A_j\}$  is called an **open cover** of  $S$ .

**Example 1.4.1** As shown in Figure 1.4.1, a circle can be covered by four arcs: arc  $A$ , arc  $B$ , arc  $C$ , and arc  $D$ . Each arc is an open set because it excludes its two endpoints. The union of these four arcs is the entire circle, so the collection of these four arcs constitutes an open cover of the circle:

$$\{A_j\} = \{\text{arc } A, \text{arc } B, \text{arc } C \text{ and arc } D\}.$$

### 2. Finite Subcover

**Definition 1.4.2** Let  $\{A_j\}$  be an open cover of a set  $S$ . If a finite set  $\{\tilde{A}_k\}$ , formed by selecting a finite number of elements from  $\{A_j\}$ , can also cover  $S$ , then the cover  $\{A_j\}$  is said to **have a finite subcover**.

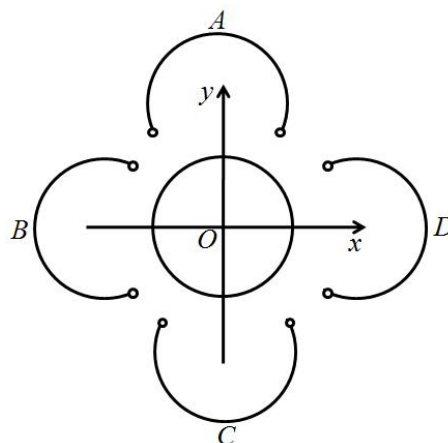
**Example 1.4.2** In Figure 1.4.1, the collection consisting of arc  $A$  and arc  $C$ ,

$$\{\tilde{A}_k\} = \{\text{arc } A, \text{arc } C\},$$

also covers the circle in the figure. Therefore, the open cover

$$\{A_j\} = \{\text{arc } A, \text{arc } B, \text{arc } C, \text{arc } D\}$$

has a finite subcover.



**Figure 1.4.1** Open cover of a circle

### 3.Compact Space

**Definition 1.4.3** A topological space is said to be **compact** if every open cover of it has a finite subcover.

**Theorem 1.4.1** A closed subset of a compact space is also compact.

**Definition 1.4.4** A set  $A \subset R^m$  is called bounded if there exists an open ball  $B \subset R^m$  such that  $A \subset B$ .

**Theorem 1.4.2** The set  $A \subset R^m$  is compact if and only if it is a closed and bounded set.

**Theorem 1.4.3** The finite union of compact subspaces is compact.

**Example 1.4.3** 1)  $R$  is not compact, but any closed interval in  $R$  is compact.

2)  $R^m$  is not compact.

3) A bounded closed subset in  $R^m$  is a compact subset of  $R^m$ .

4) Any finite set is compact.

**Theorem 1.4.4** The image of a compact space under a continuous mapping is also compact.

## §1.5 Connected Space

### 1.Complement

**Definition 1.5.1** Let set  $A$  be a subset of set  $S$ . The set consisting of all elements in  $S$  that do not belong to  $A$  is called the **complement** of  $A$ , denoted by  $A'$  or  $S - A$ .

$$A' = \{x \mid x \in S, x \notin A\}.$$

**Example 1.5.1** Suppose there is a set  $S = \{1, 2, 3\}$  and a set  $A = \{1, 2\}$ . The complement of  $A$  in  $S$  is  $A' = \{3\}$ .

### 2.Closed Set

**Definition 1.5.2** In a topological space  $S$ , a set  $A \subset S$  is called a **closed set** if its complement  $A'$  is an open set, i.e.,  $A' \in \tau$ .

**Example 1.5.2** Suppose a topological space  $S = \{1, 2\}$  has the topology  $\tau = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$ . The sets  $\{1\}$ ,  $\{2\}$ ,  $\{1, 2\}$  and  $\emptyset$  are all open sets.

The complement of the open set  $\{1\}$  in  $S$  is  $\{2\}$ , and since  $\{2\} \in \tau$ ,  $\{1\}$  is a closed set.  $\{1\}$  is both an open set and a closed set.

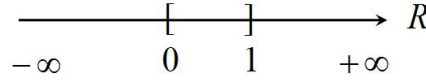
The complement of the open set  $\{2\}$  in  $S$  is  $\{1\}$ , and since  $\{1\} \in \tau$ ,  $\{2\}$  is a closed set.  $\{2\}$  is both an open set and a closed set.

The complement of  $\emptyset$  in  $S$  is  $\{1, 2\}$ , and since  $\{1, 2\} \in \tau$ ,  $\emptyset$  is a closed set.

The complement of  $\{1,2\}$  in  $S$  is  $\emptyset$ , and since  $\emptyset \in \tau$ ,  $\{1,2\}$  is a closed set.

Thus, in any topological space  $S$ , there are at least two sets that are both open and closed, namely  $S$  and  $\emptyset$ .

**Example 1.5.3** As shown in Figure 1.5.1, in  $R$ , the closed interval  $I=[0, 1]$  is a closed set. This is because its complement  $(-\infty, 0) \cup (1, +\infty)$  in  $R$  is an open set in  $R$ .



**Figure 1.5.1** A closed interval is a closed set

### 3.Continuous Mapping

**Definition 1.5.3** Let  $S$  and  $T$  be topological spaces, and let  $f : S \rightarrow T$  be a mapping. If for every open set  $U \subset T$ , the inverse image  $f^{-1}(U)$  is an open set in  $S$ , i.e., the preimage of an open set is open, then  $f$  is called a **continuous mapping**.

A continuous mapping does not necessarily map open sets to open sets. A mapping that maps open sets to open sets is called an **open mapping**.

**Theorem 1.5.1** Let  $f : S \rightarrow T$  be a mapping. Then the following conditions are equivalent:

- 1)  $f$  is a continuous mapping;
- 2) The preimage under  $f$  of any open set in  $T$  is an open set in  $S$ ;
- 3) The preimage under  $f$  of any closed set in  $T$  is a closed set in  $S$ .

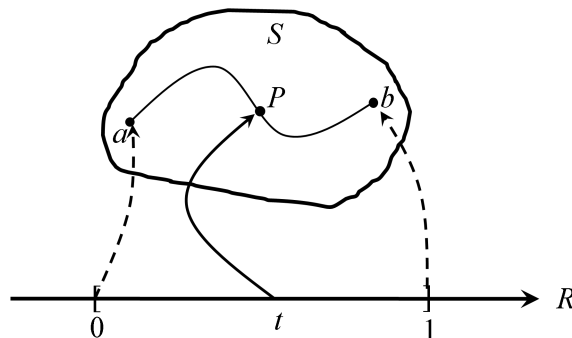
### 4.Connected Space

**Definition 1.5.4** A topological space  $S$  is called **connected** if the only subsets of  $S$  that are both open and closed are the empty set  $\emptyset$  and the whole space  $S$ .

**Example 1.5.4** The topological space  $S = \{1, 2\}$  with the discrete topology  $\tau = \{\{1\}, \{2\}, \{1,2\}, \emptyset\}$  is not connected, because every element in  $\tau$  (i.e., every subset of  $S$ ) is both an open set and a closed set.

### 5.Path-connected

**Definition 1.5.5** As shown in Figure 1.5.2, let  $I$  denote the closed interval  $[0, 1]$ , and let  $S$  be a topological space. A **path** in  $S$  from point  $a$  to point  $b$  is a continuous mapping  $\varphi : I \rightarrow S$  such that  $\varphi(0) = a$  and  $\varphi(1) = b$ .

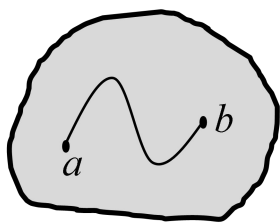


**Figure 1.5.2** A path is a continuous mapping

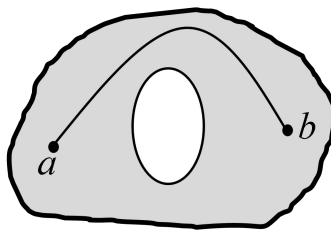
**Definition 1.5.6** A topological space  $S$  is said to be **path-connected** if for any two points  $a$  and  $b$  in  $S$ , there exists a path from  $a$  to  $b$ .

**Theorem 1.5.2** If a topological space is path-connected, then it is also connected.

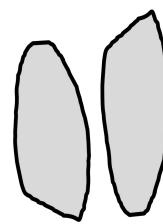
However, the converse of this theorem does not necessarily hold.



**Figure 1.5.3** Path-connected

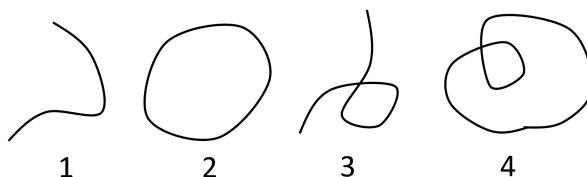


**Figure 1.5.4** Path-connected



**Figure 1.5.5** Non-path-connected

**Example 1.5.5** Figure 1.5.3 shows a path-connected topological space. Although Figure 1.5.4 has a hole in the middle, it still meets the requirements of the definition and is therefore a path-connected topological space. Figure 1.5.5 is not a path-connected topological space.



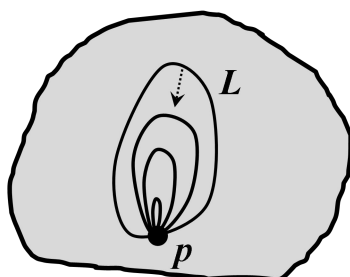
**Figure 1.5.6** Only curve 2 is a simple closed curve

## 6. Simply connected

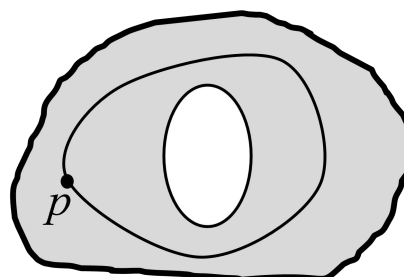
**Definition 1.5.7** A curve is called a **simple closed curve** if its two endpoints are joined together and there are no intersections along the curve.

As shown in Figure 1.5.6, curve 1 is not closed; curve 2 is closed; curve 3 has a closed segment in the middle, but its two endpoints are not joined; curve 4 has its two endpoints joined together, but there are intersections along the curve. Therefore, only curve 2 is referred to as a simple closed curve.

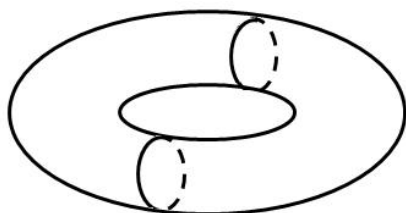
**Definition 1.5.8** Let  $S$  be a topological space. If every simple closed curve in  $S$  that passes through a given point can be continuously contracted to that point, then  $S$  is said to be **simply connected**.



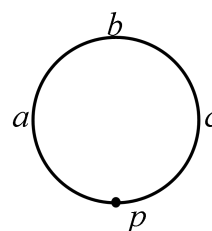
**Figure 1.5.7** Both path-connected and simply connected



**Figure 1.5.8** Path-connected but not simply connected



**Figure 1.5.9** A torus is not simply connected



**Figure 1.5.10** A circle is not simply connected

## Example 1.5.6

1) As shown in Figure 1.5.7, for any point  $p$  in the topological space, there exists a simple



closed curve  $L$  passing through  $p$ , and  $L$  can be continuously shrunk until it collapses to the point  $p$ . Therefore, this topological space is not only path-connected but also simply connected.

2) Figure 1.5.8 is path-connected but not simply connected, because the simple closed curve passing through point  $p$  is obstructed by a hole in the middle and cannot be contracted to the point  $p$ . Similarly, a torus is also not simply connected (Figure 1.5.9).

3) The circle in Figure 1.5.10 is path-connected because any two points on the circle can be connected by a path; for example, points  $a$  and  $b$  can be connected by a path along the circle between them. However, the circle is **not** simply connected because the simple closed curve passing through point  $p$  is the circle itself, which cannot be contracted to the single point  $p$ . Therefore, the circle is not simply connected. In contrast, the  $m$ -dimensional sphere  $S^m$ ,  $m \geq 2$  is simply connected.

4)  $R^m$  is simply connected. For  $m > 2$ ,  $R^m$  remains simply connected even after removing a finite number of points.

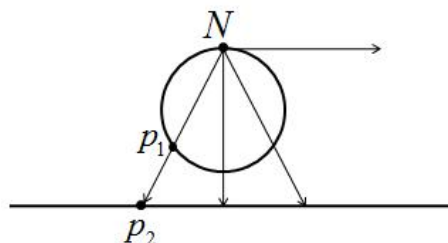
## §1.6 Homeomorphic Mapping

**Definition 1.6.1** Let  $S$  and  $T$  be topological spaces, and suppose there exists a mapping  $f: S \rightarrow T$ . If  $f$  is a bijection and both  $f$  and its inverse  $f^{-1}$  are continuous, then  $f$  is called a **homeomorphic mapping** or a **homeomorphism**.

A homeomorphic mapping  $f$  not only bijectively maps the points of  $S$  to the points of  $T$ , but also bijectively maps the open sets of  $S$  to the open sets of  $T$ . Therefore,  $S$  and  $T$  share the same topology and are essentially the same topological space, possessing identical topological properties. For example, if  $S$  and  $T$  are homeomorphic, then if  $S$  is connected,  $T$  is also connected; if  $S$  is compact,  $T$  is also compact.

**Theorem 1.6.1** Suppose there exists a homeomorphism  $f: S \rightarrow T$  between topological spaces  $S$  and  $T$ . If a point  $x$  is removed from  $S$  and the corresponding point  $f(x)$  is removed from  $T$ , then the remaining spaces  $S \setminus \{x\}$  and  $T \setminus \{f(x)\}$  are still homeomorphic.

**Example 1.6.1** There does not exist a homeomorphic mapping between a circle and a line. As shown in Figure 1.6.1, a point  $p_1$  on the circle can be mapped to a point  $p_2$  on the line, but the point  $N$  on the circle cannot be mapped to any point on the line. In other words, since the point  $N$  on the circle cannot be mapped to a point on the line—even though all other points can be mapped—there does not exist a bijective mapping between the circle and the line, and thus no homeomorphic mapping exists.

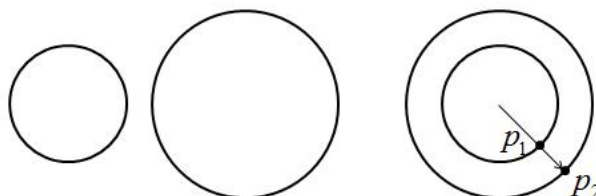


**Figure 1.6.1** A circle and a line are not homeomorphic

Therefore, the circle and the line are not homeomorphic. However, if a point at infinity is added to the line, it becomes homeomorphic to the circle. Generally,  $R^m$  space with a point at infinity added is homeomorphic to the  $m$ -dimensional sphere  $S^m$ , denoted as

$$S^m \cong R^m \cup \{\infty\}.$$

The symbol  $\cong$  denotes homeomorphism.



**Figure 1.6.2** A large circle and a small circle are homeomorphic

**Example 1.6.2** There exists a homeomorphic mapping between a large circle and a small

circle. As shown in Figure 1.6.2, if we align the centers of the large circle and the small circle, we can see that every point on the small circle can be mapped one-to-one to a point on the large circle, and conversely, every point on the large circle can be mapped one-to-one to a point on the small circle. Therefore, the two circles are homeomorphic.

Homeomorphic mappings are **transitive**. For example, a small circle with a radius less than 1 is homeomorphic to a circle with a radius equal to 1, and a circle with a radius equal to 1 is homeomorphic to a circle with a radius greater than 1. Consequently, a circle with a radius less than 1 is also homeomorphic to a circle with a radius greater than 1.

## §1.7 Hausdorff Space

**Definition 1.7.1** Let  $S$  be a topological space. If for any two distinct points  $s_1, s_2$ , there exist open sets  $U_j$ , such that  $s_1 \in U_1, s_2 \in U_2$  and  $U_1 \cap U_2 = \emptyset$ , then  $S$  is called a **Hausdorff space**.

**Example 1.7.1** The topological space consisting of the set  $S = \{1, 2\}$  and the topology  $\tau = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$  is a Hausdorff space, because for distinct points 1 and 2, we have  $1 \in \{1\}, 2 \in \{2\}$ , and  $\{1\} \cap \{2\} = \emptyset$ .

**Example 1.7.2** The set of real numbers  $R$  is a Hausdorff space. For instance, consider the two real numbers 3.1415 and 3.1416 in  $R$ . These two numbers are extremely close, but we can still take an open interval around 3.1415, say  $(3.1413, 3.14151)$ , such that  $3.1415 \in (3.1413, 3.14151)$ , and an open interval around 3.1416, say  $(3.14152, 3.14167)$ , such that  $3.1416 \in (3.14152, 3.14167)$  with  $(3.1413, 3.14151) \cap (3.14152, 3.14167) = \emptyset$ . This is possible because there are infinitely many real numbers between 3.1415 and 3.1416.

Similarly,  $R^m$  is also a Hausdorff space.

**Example 1.7.3** The trivial topological space mentioned earlier is not a Hausdorff space.

**Theorem 1.7.1** A subspace of a Hausdorff space is also a Hausdorff space.

## §1.8 Quotient space

As mentioned earlier, let  $\sim$  be an equivalence relation defined on a set  $S$ . The family of all distinct equivalence classes of  $S$  under the equivalence relation  $\sim$  is called the **quotient set** with respect to  $\sim$ , denoted as  $S/\sim$ .

**Definition 1.8.1** The mapping  $f: S \rightarrow S/\sim$ , is called the **identification mapping** or **quotient map**.

The identification mapping sends each element in  $S$  to its equivalence class. Let  $x \in S$ . The set of all elements in  $S$  equivalent to  $x$  forms the equivalence class of  $x$ , denoted as  $[x]$ . Then  $f: x \rightarrow [x]$ .

If  $S$  is a topological space, then an equivalence relation can be defined on  $S$  to form a quotient space.

**Definition 1.8.2** Let  $(S, \tau)$  be a topological space and  $\sim$  an equivalence relation on the set  $S$ . Taking the family of subsets in the quotient set  $S/\sim$ :

$$\tilde{\tau} = \{U \subset S/\sim \mid f^{-1}(U) \in \tau\}$$

as the topology on  $S/\sim$ , then  $(S/\sim, \tilde{\tau})$  becomes a topological space, called the **quotient space**, and this topology is called the **quotient topology** on  $S/\sim$  induced by  $f$ .

**Theorem 1.8.1** Let  $A$  be a subset of a topological space  $S$ . Define an equivalence relation in  $A$  as follows:

$$x \sim y \text{ for all } x, y \in A,$$

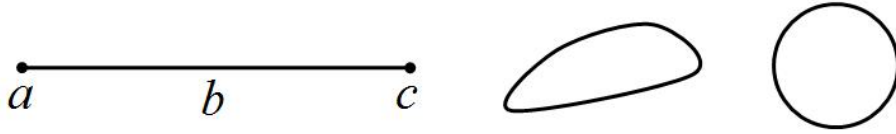
and for elements in  $S$  not belonging to  $A$ , define the equivalence relation as:

$$x \sim x \text{ for all } x \in S \text{ but } x \notin A$$

Then the resulting space  $S/A$  is a quotient space.

The essence of this theorem is to treat  $A$  as a single equivalence class—that is, to collapse  $A$  into a single point or regard it as the same element—while each element outside  $A$  forms its own equivalence class (each element being equivalent only to itself). The resulting space  $S/A$  is then a quotient space.

**Example 1.8.1** As shown in Figure 1.8.1, the line segment  $ac$  is a topological space  $S$ . We take the two endpoints  $a$  and  $c$  to form a subset  $A = \{a, c\}$ . Define an equivalence relation on  $A$  by letting  $a$  and  $c$  be equivalent, i.e.,  $a \sim c$ . Other points on the line segment  $ac$  are each equivalent only to themselves, such as point  $b$  satisfying  $b \sim b$ . The resulting space  $S/A$  is a quotient space. Since  $a$  and  $c$  are equivalent, they can be "glued" together, forming a simple closed curve  $C$  is homeomorphic to a circle  $S^1$ . Therefore, the quotient space  $S/A$  is also homeomorphic to  $S^1$ , i.e.,  $S/A \cong S^1$ .



**Figure 1.8.1** A simple closed curve is the quotient space of the line segment  $ac$

## §1.9 Product of Topological Spaces

Suppose there are two topological spaces  $(S_1, \tau_1)$  and  $(S_2, \tau_2)$ . They can be multiplied to form a **product topological space**, denoted as  $(S, \tau)$ . The method is as follows: the sets  $S_1$  and  $S_2$  are taken as the direct product:

$$S = S_1 \times S_2, \quad x_1 \in S_1, \quad x_2 \in S_2, \quad x = (x_1, x_2) \in S;$$

The topologies are also taken as the product topology:

$$\tau = \tau_1 \times \tau_2, \quad U_1 \in \tau_1, \quad U_2 \in \tau_2, \quad U = (U_1, U_2) \in \tau.$$

Here,  $\tau_1$  and  $\tau_2$  are the topologies of  $S_1$  and  $S_2$ , respectively.

**Example 1.9.1** Consider two topological spaces  $(S_1, \tau_1)$  and  $(S_2, \tau_2)$ :

$$S_1 = \{1, 2\}, \quad \tau_1 = \{\{\}, \{2\}, \{1, 2\}, \emptyset\}; \quad S_2 = \{a\}, \quad \tau_2 = \{\{a\}, \emptyset\}.$$

The product topology space can be constructed as follows:

$$S = S_1 \times S_2 = \{(1, a), (2, a)\}.$$

The product topology is

$$\tau = \tau_1 \times \tau_2 = \{(\{1\}, \{a\}), (\{2\}, \{a\}), (\{1, 2\}, \{a\}), (\emptyset, \{a\}), (\{1\}, \emptyset), (\{2\}, \emptyset), (\{1, 2\}, \emptyset), (\emptyset, \emptyset)\}.$$

**Theorem 1.9.1** The product topological space has the following properties:

- 1) The product topological space  $(S_1, \tau_1) \times (S_2, \tau_2)$  is a Hausdorff space if and only if both  $(S_1, \tau_1)$  and  $(S_2, \tau_2)$  are Hausdorff spaces;
- 2) If the topological spaces  $(S_1, \tau_1)$  and  $(S_2, \tau_2)$  are both compact, then the product topological space  $(S_1, \tau_1) \times (S_2, \tau_2)$  is also compact. Conversely, if  $(S_1, \tau_1) \times (S_2, \tau_2)$  is compact, then both  $(S_1, \tau_1)$  and  $(S_2, \tau_2)$  are compact;
- 3) The product of two connected spaces is also connected;
- 4) The product of two path-connected spaces is also path-connected.

Of course, a finite number of topological spaces can also be multiplied to form a product topological space. If there are  $m$  topological spaces  $S_1, \dots, S_k, \dots, S_m$ , their product becomes the product topological space  $S_1 \times \dots \times S_k \times \dots \times S_m$ .

The product operation on topological spaces satisfies the associative law, i.e.,

$$S_1 \times S_2 \times S_3 = (S_1 \times S_2) \times S_3 = S_1 \times (S_2 \times S_3).$$

Let

$$x_k \in S_k, \quad x = (x_1, \dots, x_k, \dots, x_m) \in S,$$

then the  $k$ -th **projection mapping**

$$\pi_k : S_1 \times \dots \times S_k \times \dots \times S_m \rightarrow S_k$$

is defined as

$$\pi_k(x_1, \dots, x_k, \dots, x_m) = x_k.$$

## §1.10 Euler characteristic

The **Euler characteristic** of a polyhedron is calculated using the formula:

$$\chi = V - E + F,$$

where  $V$ ,  $E$ , and  $F$  are the numbers of vertices, edges, and faces of the polyhedron, respectively.

**Example 1.10.1** As shown in Figure 1.10.1, a rectangular prism has 8 vertices, 12 edges, and 6 faces. Substituting these values into the formula gives

$$\chi = 8 - 12 + 6 = 2,$$

Thus, the Euler characteristic of the rectangular prism is 2.



Figure 1.10.1 Rectangular prism

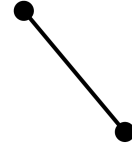


Figure 1.10.2 Line segment

**Example 1.10.2** As shown in Figure 1.10.2, a line segment can be regarded as a special polyhedron. It has 2 vertices, 1 edge, and 0 faces. Substituting these values into the formula gives

$$\chi = 2 - 1 + 0 = 1,$$

Thus, the Euler characteristic of a line segment is 1.

**Example 1.10.3** The Euler characteristic of the  $m$ -dimensional sphere is given by  $\chi = 1 + (-1)^m$ . For example:

- 1) A circle is 1-dimensional ( $m=1$ , odd), so its Euler characteristic is  $\chi = 1 + (-1)^1 = 0$ ;
- 2) A 2-dimensional sphere has Euler characteristic 2;
- 3) A 3-dimensional sphere  $S^3$  has Euler characteristic 0.

**Example 1.10.4** The Euler characteristic of a disk is equal to 1.

**Example 1.10.5** The Euler characteristic of the  $m$ -dimensional real projective space is given by  $\chi = \frac{1}{2}(1 + (-1)^m)$ . For example:

- 1) The 1-dimensional real projective line ( $m=1$ , odd) has Euler characteristic 0;
- 2) The 2-dimensional real projective plane ( $m=2$ , even) has Euler characteristic 1;
- 3) The 3-dimensional real projective space ( $m=3$ , odd) has Euler characteristic 0.

**Theorem 1.10.1** Let  $M$  and  $N$  be two topological spaces. Then:

- 1) The Euler characteristic of their union satisfies

$$\chi(M \cup N) = \chi(M) + \chi(N) - \chi(M \cap N),$$

In particular, when  $M \cap N = \emptyset$ , we have

$$\chi(M \cup N) = \chi(M) + \chi(N).$$

- 2) The Euler characteristic of their product space is

$$\chi(M \times N) = \chi(M) \cdot \chi(N).$$

**Example 1.10.6** If two disjoint 3-dimensional spheres  $S^3$  are combined, the Euler

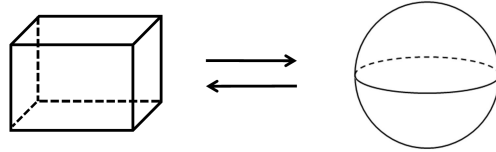
characteristic of the resulting union is

$$\chi(S^3 \cup S^3) = 0 + 0 = 0.$$

**Example 1.10.7** The  $m$ -dimensional torus  $T^m$  is the product space of  $m$  circles. Since the Euler characteristic of each circle is 0, the Euler characteristic of the  $m$ -dimensional torus is 0.

The Euler characteristic is a topological invariant. This means that if a topological space has Euler characteristic  $\chi_1$ , then any new topological space generated from it via a homeomorphism will also have Euler characteristic  $\chi_1$ .

**Example 1.10.8** As shown in Figure 1.10.3, suppose the surface of a rectangular prism is made of elastic rubber film and is hollow inside. If this prism is inflated, its surface will transform into a 2-dimensional sphere. Therefore, the rectangular prism is homeomorphic to the sphere. Since the Euler characteristic is a topological invariant, and the Euler characteristic of the rectangular prism is 2, the Euler characteristic of the sphere obtained via this homeomorphism is also 2. A homeomorphism is a reversible mapping, meaning that the 2-dimensional sphere can also be homeomorphically mapped back to a rectangular prism.



**Figure 1.10.3** The surface of a rectangular prism is homeomorphic to a 2-dimensional sphere

## §1.11 Group

### 1.Group

**Definition 1.11.1** Let  $G$  be a set with a binary operation, called "multiplication". For  $a \in G$ ,  $b \in G$ , the product of  $a$  and  $b$  is denoted as  $ab$ . If the set  $G$  together with this multiplication satisfies the following four axioms, we call  $G$  a **group**.

1)**Closure** If  $a \in G$ ,  $b \in G$ , then  $ab \in G$ .

2)**Associativity** For all  $a, b, c \in G$ ,

$$a(bc) = (ab)c.$$

3)**Existence of an identity element** There exists an element  $e \in G$  such that for every  $a \in G$

$$ae = ea = a \quad (\forall a \in G).$$

4)**Existence of inverse elements** For each  $a \in G$ , there exists an element  $a^{-1} \in G$  such that

$$aa^{-1} = a^{-1}a = e \quad (\forall a \in G).$$

**Example 1.11.1** It can be verified that the set of positive real numbers  $R^+ = \{x \mid x > 0, x \in R\}$  forms a group under ordinary multiplication. For example,  $3 \in R^+$ ,  $4 \in R^+$ , and  $3 \times 4 = 12 \in R^+$ , so the closure property holds. The inverse element of  $3 \in R^+$  is  $\frac{1}{3} \in R^+$ , and the identity element is  $1 \in R^+$ .

**Example 1.11.2** The multiplication operation in the definition of a group should be understood in a broad sense. For instance, the addition of two real numbers can also be interpreted as the multiplication operation in the group definition. If we regard the addition of two integers in the set

$$Z = \{\dots, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, \dots\}$$

as a form of multiplication, then the integer set  $Z$  forms a group. For example,  $-3 \in Z$ ,  $6 \in Z$ , and  $(-3) + 6 = 3 \in Z$ , so the closure property holds.

## 2.Subgroup

**Definition 1.11.2** Let  $H \subset G$ . If, under the multiplication structure defined for the group  $G$ ,  $H$  also forms a group, then  $H$  is called a **subgroup** of  $G$ .

The group  $G$  and its subgroup  $H$  must contain the same identity element.

## 3.Invariant Subgroup

**Definition 1.11.3** Let  $N$  be a subgroup of a group  $G$ . If for every element  $a \in G$ , we have  $aNa^{-1} = N$ , then the subgroup  $N$  is called an **invariant subgroup** or a **normal subgroup** of  $G$ .

The fact that  $N$  is a normal subgroup of  $G$  can be denoted as  $N \triangleleft G$ .

**Example 1.11.3** The set

$$N = 3Z = \{\dots, -12, -9, -6, -3, 0, 3, 6, 9, 12, \dots\}$$

is a subgroup of the integer group  $Z$  from Example 1.11.2. It can be verified that the subgroup  $3Z$  is also a normal subgroup of  $Z$ . For example, for  $2 \in Z$  and  $9 \in 3Z$ , we have  $2 + 9 + (-2) = 9 \in 3Z$ .

## 4.Quotient Group

**Definition 1.11.4** Let  $N$  be a normal subgroup of the group  $G$ . Define the multiplication operation as

$$(aN)(bN) = (ab)N, \quad \forall a, b \in G,$$

then the set consisting of the cosets  $aN$ , i.e.,

$$G/N = \{aN \mid a \in G, N \triangleleft G\}$$

forms a group, called the **quotient group** of  $G$  by  $N$ .

Let us verify that the set  $G/N$  is indeed a group.

1) Closure Let  $aN \in G/N$ ,  $bN \in G/N$ . Then  $(aN)(bN) = abN \in G/N$ .

2) Associativity Let  $cN \in G/N$ . Since

$$(aNbN)cN = (abN)cN = abcN, \quad aN(bNcN) = aN(bcN) = abcN,$$

we have

$$(aNbN)cN = aN(bNcN).$$

3) Identity element The identity element is  $N$ , because  $(aN)N = aN = N(aN)$ .

4) Existence of inverses For each element  $aN$ , the inverse is  $a^{-1}N$ , because

$$(aN)(a^{-1}N) = aa^{-1}N = N, \quad (a^{-1}N)(aN) = a^{-1}aN = N.$$

**Example 1.11.4** Partition the integer set

$$Z = \{\dots, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, \dots\}$$

by  $Z/nZ = \{a + nZ \mid a \in Z\}$ . Take  $n = 3$ , then the integer set  $Z$  is partitioned into three sets:

$$[0] = 3Z = \{\dots, -6, -3, 0, 3, 6, \dots\},$$

$$[1] = 1 + 3Z = \{\dots, -5, -2, 1, 4, 7, \dots\},$$

$$[2] = 2 + 3Z = \{\dots, -4, -1, 2, 5, 8, \dots\},$$

It can be verified that the quotient set  $Z/3Z = \{[0], [1], [2]\}$  forms a group, called the quotient group. For example,

$$[1] + [2] = 3 + 3Z = [0], \quad [2] + [2] = 4 + 3Z = [1],$$

so the closure property holds.

## Chapter 2 Tensors as Invariants Under Coordinate Transformations

### §2.1 Coordinate Transformation

#### 1. Transformation of coordinate differentials

Suppose in an  $m$ -dimensional Euclidean space, the coordinate transformation between two coordinate systems  $x^i$  and  $y^j$  ( $i, j = 1, 2, \dots, m$ ) is given by

$$y^j = f^j(x^1, x^2, \dots, x^m). \quad (2.1.1)$$

Differentiating both sides, we obtain the transformation rule for coordinate differentials:

$$dy^j = \sum_{i=1}^m \frac{\partial y^j}{\partial x^i} dx^i.$$

#### 2. Einstein Summation Convention

We adopt the **Einstein summation convention**: In a single term, if an index appears twice, it implies summation over that index within its range, and the summation symbol is omitted. For example, using this convention, the equation above can be written as

$$dy^j = \frac{\partial y^j}{\partial x^i} dx^i. \quad (2.1.2)$$

The repeated index to be summed over is called a **dummy index**, while the index that is not summed is called a **free index**. The letter representing a dummy index may be changed as needed. For instance, the same equation can also be written as

$$dy^j = \frac{\partial y^j}{\partial x^\alpha} dx^\alpha. \quad (2.1.3)$$

#### 3. Jacobian Matrix

The matrix

$$J = \left( \frac{\partial y^j}{\partial x^i} \right) = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} & \dots & \frac{\partial y^1}{\partial x^m} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} & \dots & \frac{\partial y^2}{\partial x^m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial y^m}{\partial x^1} & \frac{\partial y^m}{\partial x^2} & \dots & \frac{\partial y^m}{\partial x^m} \end{pmatrix}$$

is called the **Jacobian matrix** of the transformation function  $y^j$ .

If the transformation is **non-singular**, i.e., its Jacobian determinant satisfies

$$|J| \neq 0 \text{ (or } \infty),$$

then the inverse function of (2.1.1) exists, and consequently the inverse of transformation (2.1.2) exists:

$$dx^j = \frac{\partial x^j}{\partial y^\alpha} dy^\alpha. \quad (2.1.4)$$

Renaming the indices in (2.1.2) gives

$$dy^\alpha = \frac{\partial y^\alpha}{\partial x^i} dx^i. \quad (2.1.5)$$

Substituting (2.1.5) into (2.1.4) yields

$$\frac{\partial x^j}{\partial y^\alpha} \cdot \frac{\partial y^\alpha}{\partial x^i} = \frac{\partial x^j}{\partial x^i} = \delta_i^j. \quad (2.1.6)$$

Similarly, we obtain

$$\frac{\partial y^j}{\partial x^\alpha} \cdot \frac{\partial x^\alpha}{\partial y^i} = \frac{\partial y^j}{\partial y^i} = \delta_i^j. \quad (2.1.7)$$

The symbol  $\delta_i^j$  is the **Kronecker delta**, defined as

$$\delta_i^j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Equations (2.1.6) and (2.1.7) show that the transformation matrices in (2.1.2) and (2.1.4) are inverses of each other.

## §2.2 Definition of Tensors

A **tensor** is a collection of elements that satisfy specific transformation relations. The number of elements is determined by the dimension  $m$  of the space and the order  $n$  of the tensor.

### 1.Zero-Order Tensor

A zero-order tensor, also called a **scalar**, has only  $m^0 = 1$  component and remains invariant under coordinate transformations. Suppose it is denoted as  $\varphi(x^1, x^2, \dots, x^m)$  in the coordinate system  $x^i$  and as  $\tilde{\varphi}(y^1, y^2, \dots, y^m)$  in the coordinate system  $y^i$ . Then

$$\tilde{\varphi}(y^1, y^2, \dots, y^m) = \varphi(x^1, x^2, \dots, x^m).$$

A scalar is also referred to as an **invariant**. Scalar functions are invariants, and constants are also invariants.

### 2.Contravariant Tensor

#### (1)First-Order Contravariant Tensor

A first-order contravariant tensor is also called a **contravariant vector**. If a quantity consists of  $m^1 = m$  components, with components  $(V^1, V^2, \dots, V^m)$  in the coordinate system  $x^i$  and components  $(\tilde{V}^1, \tilde{V}^2, \dots, \tilde{V}^m)$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{V}^\alpha = \frac{\partial y^\alpha}{\partial x^\mu} V^\mu, \quad \alpha, \mu = 1, 2, \dots, m,$$

then it is called a **first-order contravariant tensor** in  $m$ -dimensional space, and  $V^\mu$  are called its **contravariant components**. The superscript index  $\mu$  is referred to as a **contravariant index**.

For convenience,  $V^\mu$  denotes both the first-order contravariant tensor itself and its components. This convention will be followed hereafter.

We write the transformation rule for a contravariant tensor in matrix form. Taking two-dimensional space as an example,  $m = 2$ . First, expand the transformation equations:

$$\tilde{V}^1 = \frac{\partial y^1}{\partial x^1} V^1 + \frac{\partial y^1}{\partial x^2} V^2,$$



$$\tilde{V}^2 = \frac{\partial y^2}{\partial x^1} V^1 + \frac{\partial y^2}{\partial x^2} V^2.$$

Thus, the matrix form of the transformation is

$$\begin{pmatrix} \tilde{V}^1 \\ \tilde{V}^2 \end{pmatrix} = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} \end{pmatrix} \begin{pmatrix} V^1 \\ V^2 \end{pmatrix}.$$

Therefore, the transformation matrix for a contravariant tensor is precisely the Jacobian matrix of the coordinate transformation.

**Example 2.2.1** Velocity and acceleration are first-order contravariant tensors.

Suppose the transformation between new coordinates  $y^\mu$  and old coordinates  $x^\nu$  is given by

$$y^\mu = f^\mu(x^1, x^2, \dots, x^m), \quad \mu = 1, 2, \dots, m, \quad (2.2.1)$$

and the Jacobian determinant satisfies  $|J| \neq 0$  (or  $\infty$ ). Then we have

$$dy^\mu = \frac{\partial y^\mu}{\partial x^\alpha} dx^\alpha.$$

Dividing both sides by the differential of time  $dt$  gives

$$\frac{dy^\mu}{dt} = \frac{\partial y^\mu}{\partial x^\alpha} \frac{dx^\alpha}{dt}.$$

Here  $\frac{dy^\mu}{dt}$  is the  $\mu$ -component of velocity in the new coordinate system, denoted as  $\tilde{V}^\mu$ , and

$\frac{dx^\alpha}{dt}$  is the  $\alpha$ -component of velocity in the old coordinate system, denoted as  $V^\alpha$ . Thus,

$$\tilde{V}^\mu = \frac{\partial y^\mu}{\partial x^\alpha} V^\alpha.$$

These components precisely satisfy the transformation rule for a first-order contravariant tensor. Therefore, velocity is a first-order contravariant tensor.

Differentiating the above equation with respect to time  $t$  yields

$$\frac{d\tilde{V}^\mu}{dt} = \frac{\partial y^\mu}{\partial x^\alpha} \frac{dV^\alpha}{dt}.$$

Clearly, the components of acceleration also satisfy the transformation rule for a first-order contravariant tensor. Hence, acceleration is also a first-order contravariant tensor.

### (2) Second-Order Contravariant Tensor

If a quantity consists of  $m^2$  components, with components  $T^{\mu\nu}$  in the coordinate system  $x^i$  and components  $\tilde{T}^{\alpha\beta}$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{T}^{\alpha\beta} = \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial y^\beta}{\partial x^\nu} T^{\mu\nu}, \quad \alpha, \beta, \mu, \nu = 1, 2, \dots, m,$$

then it is called a **second-order contravariant tensor** in  $m$ -dimensional space, and  $T^{\mu\nu}$  are called its **contravariant components**.

### (3) $n$ -th Order Contravariant Tensor

If a quantity consists of  $m^n$  components, with components  $T^{\mu_1\mu_2\cdots\mu_n}$  in the coordinate system  $x^i$  and components  $\tilde{T}^{\alpha_1\alpha_2\cdots\alpha_n}$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{T}^{\alpha_1\alpha_2\cdots\alpha_n} = \frac{\partial y^{\alpha_1}}{\partial x^{\mu_1}} \frac{\partial y^{\alpha_2}}{\partial x^{\mu_2}} \cdots \frac{\partial y^{\alpha_n}}{\partial x^{\mu_n}} T^{\mu_1\mu_2\cdots\mu_n}, \quad \alpha_1, \alpha_2, \dots, \alpha_n, \mu_1, \mu_2, \dots, \mu_n = 1, 2, \dots, m,$$

then it is called an  **$n$ -th order contravariant tensor** in  $m$ -dimensional space, and  $T^{\mu_1\mu_2\cdots\mu_n}$  are

called its **contravariant components**.

### 3. Covariant Tensor

#### (1) First-Order Covariant Tensor

A first-order covariant tensor is also called a **covariant vector**. If a quantity consists of  $m$  components, with components  $(V_1, V_2, \dots, V_m)$  in the coordinate system  $x^i$  and components  $(\tilde{V}_1, \tilde{V}_2, \dots, \tilde{V}_m)$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{V}_\alpha = \frac{\partial x^\mu}{\partial y^\alpha} V_\mu, \quad \alpha, \mu = 1, 2, \dots, m,$$

then it is called a **first-order covariant tensor** in  $m$ -dimensional space, and  $V_\mu$  are called its **covariant components**. The subscript index  $\mu$  is referred to as a **covariant index**.

We also write the transformation rule for a covariant tensor in matrix form. Taking two-dimensional space as an example,  $m = 2$ . First, expand the transformation equations:

$$\begin{aligned} \tilde{V}_1 &= \frac{\partial x^1}{\partial y^1} V_1 + \frac{\partial x^2}{\partial y^1} V_2, \\ \tilde{V}_2 &= \frac{\partial x^1}{\partial y^2} V_1 + \frac{\partial x^2}{\partial y^2} V_2. \end{aligned}$$

Thus, the matrix form of the transformation is

$$\begin{pmatrix} \tilde{V}_1 \\ \tilde{V}_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^2}{\partial y^1} \\ \frac{\partial x^1}{\partial y^2} & \frac{\partial x^2}{\partial y^2} \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix},$$

Therefore, the transformation matrix for a covariant tensor is the transpose of the Jacobian matrix of the **inverse** coordinate transformation (2.2.1).

**Example 2.2.2** Gradient is a first-order covariant tensor.

Let  $\tilde{\varphi}(y^1, y^2, \dots, y^m)$  and  $\varphi(x^1, x^2, \dots, x^m)$  be scalar functions in the new and old coordinate systems, respectively. Since they are scalars, they are equal:

$$\tilde{\varphi}(y^1, y^2, \dots, y^m) = \varphi(x^1, x^2, \dots, x^m).$$

The gradient of  $\varphi$ , denoted as  $A$ , is given by

$$A = \text{grad } \varphi.$$

$A$  is a vector, i.e., a first-order tensor. In the new coordinate system, the  $\alpha$ -component of  $A$  is

$$\tilde{A}_\alpha = \frac{\partial \tilde{\varphi}}{\partial y^\alpha}.$$

We can express this in terms of the old coordinates:

$$\frac{\partial \tilde{\varphi}}{\partial y^\alpha} = \frac{\partial \tilde{\varphi}}{\partial x^\mu} \frac{\partial x^\mu}{\partial y^\alpha} = \frac{\partial \varphi}{\partial x^\mu} \frac{\partial x^\mu}{\partial y^\alpha}.$$

Here  $\frac{\partial \varphi}{\partial x^\mu}$  is exactly the  $\mu$ -component of the gradient in the old coordinate system, i.e.,  $A_\mu$ .

Thus,

$$A_\mu = \frac{\partial \varphi}{\partial x^\mu}.$$

$\tilde{A}_\alpha$  and  $A_\mu$  satisfy the transformation rule for a first-order covariant tensor:

$$\tilde{A}_\alpha = \frac{\partial x^\mu}{\partial y^\alpha} A_\mu.$$

Therefore, the gradient is a first-order covariant tensor.

#### (2) Second-Order Covariant Tensor

If a quantity consists of  $m^2$  components, with components  $T_{\mu\nu}$  in the coordinate system  $x^i$  and components  $\tilde{T}_{\alpha\beta}$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{T}_{\alpha\beta} = \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} T_{\mu\nu}, \quad \alpha, \beta, \mu, \nu = 1, 2, \dots, m,$$

then it is called a **second-order covariant tensor** in  $m$ -dimensional space, and  $T_{\mu\nu}$  are called its **covariant components**.

### (3) $n$ -th Order Covariant Tensor

If a quantity consists of  $m^n$  components, with components  $T_{\nu_1\nu_2\cdots\nu_n}$  in the coordinate system  $x^i$  and components  $\tilde{T}_{\beta_1\beta_2\cdots\beta_n}$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{T}_{\beta_1\beta_2\cdots\beta_n} = \frac{\partial x^{\nu_1}}{\partial y^{\beta_1}} \frac{\partial x^{\nu_2}}{\partial y^{\beta_2}} \cdots \frac{\partial x^{\nu_n}}{\partial y^{\beta_n}} T_{\nu_1\nu_2\cdots\nu_n}, \quad \beta_1, \beta_2, \cdots, \beta_n, \nu_1, \nu_2, \cdots, \nu_n = 1, 2, \dots, m,$$

then it is called an  **$n$ -th order covariant tensor** in  $m$ -dimensional space, and  $T_{\nu_1\nu_2\cdots\nu_n}$  are called its **covariant components**.

## 4. Mixed Tensor

### (1) Second-Order Mixed Tensor

If a quantity consists of  $m^2$  components, with components  $T_\nu^\mu$  in the coordinate system  $x^i$  and components  $\tilde{T}_\beta^\alpha$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{T}_\beta^\alpha = \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial x^\nu}{\partial y^\beta} T_\nu^\mu, \quad \alpha, \beta, \mu, \nu = 1, 2, \dots, m,$$

then it is called a **second-order mixed tensor** in  $m$ -dimensional space. In this tensor, the upper index  $\mu$  is **contravariant**, and the lower index  $\nu$  is **covariant**.

**Example 2.2.3** The Kronecker delta

$$\delta_\nu^\mu = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu \end{cases}$$

is a second-order mixed tensor. This can be seen from its transformation behavior:

$$\delta_\beta^\alpha = \frac{\partial y^\alpha}{\partial x^\beta} = \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial x^\mu}{\partial y^\beta} = \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial x^\nu}{\partial y^\beta} \delta_\nu^\mu,$$

which satisfies the transformation rule for a second-order mixed tensor.

### (2) Higher-Order Mixed Tensor

If a quantity consists of  $m^{p+q}$  components, with components  $T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q}$  in the coordinate system  $x^i$  and components  $\tilde{T}^{\alpha_1\alpha_2\cdots\alpha_p}_{\beta_1\beta_2\cdots\beta_q}$  in the coordinate system  $y^i$ , and the transformation rule for the components is

$$\tilde{T}^{\alpha_1\alpha_2\cdots\alpha_p}_{\beta_1\beta_2\cdots\beta_q} = \frac{\partial y^{\alpha_1}}{\partial x^{\mu_1}} \frac{\partial y^{\alpha_2}}{\partial x^{\mu_2}} \cdots \frac{\partial y^{\alpha_p}}{\partial x^{\mu_p}} \frac{\partial x^{\nu_1}}{\partial y^{\beta_1}} \frac{\partial x^{\nu_2}}{\partial y^{\beta_2}} \cdots \frac{\partial x^{\nu_q}}{\partial y^{\beta_q}} T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q},$$

$$\mu_1, \mu_2, \cdots, \mu_p, \nu_1, \nu_2, \cdots, \nu_q = 1, 2, \dots, m,$$

then it is called a **mixed tensor of order  $p+q$**  in  $m$ -dimensional space. Among its indices, the upper  $p$  indices  $\mu_1, \mu_2, \cdots, \mu_p$  are **contravariant**, and the lower  $q$  indices  $\nu_1, \nu_2, \cdots, \nu_q$  are **covariant**.

If a tensor has  $p$  contravariant indices and  $q$  covariant indices, it is called a **tensor of type  $(p, q)$** . Thus, a mixed tensor of order  $p+q$  is also referred to as a  $(p, q)$ -type tensor. A

$(p,0)$ -type tensor is a  $p$ -th order **contravariant** tensor. A  $(0,q)$ -type tensor is a  $q$ -th order **covariant** tensor. A  $(1,0)$ -type tensor is a first-order contravariant tensor, i.e., a **contravariant vector**. A  $(0,1)$ -type tensor is a first-order covariant tensor, i.e., a **covariant vector**.

## §2.3 Symmetry and Antisymmetry of Tensors

For a second-order contravariant tensor, if exchanging its indices yields

$$T^{\mu\nu} = T^{\nu\mu},$$

then it is called a **symmetric second-order contravariant tensor**. Similarly, for a second-order covariant tensor, if

$$T_{\mu\nu} = T_{\nu\mu},$$

then it is called a **symmetric second-order covariant tensor**.

If exchanging the indices of a second-order contravariant tensor gives

$$T^{\mu\nu} = -T^{\nu\mu},$$

then it is called an **antisymmetric second-order contravariant tensor**. Similarly, for a second-order covariant tensor, if

$$T_{\mu\nu} = -T_{\nu\mu},$$

then it is called an **antisymmetric second-order covariant tensor**.

For tensors of order higher than two,  $T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q}$ , if swapping any two contravariant indices (e.g.,  $\mu_1, \mu_2$ ) yields

$$T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q} = T^{\mu_2\mu_1\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q},$$

then the tensor is said to be **symmetric** with respect to those **contravariant** indices. Similarly, if swapping any two covariant indices (e.g.,  $\nu_1, \nu_2$ ) yields

$$T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q} = T^{\mu_1\mu_2\cdots\mu_p}_{\nu_2\nu_1\cdots\nu_q},$$

then the tensor is said to be **symmetric** with respect to those **covariant** indices.

Index swapping can only be performed among contravariant indices or among covariant indices; it is not allowed between a contravariant index and a covariant index.

Similarly, if swapping any two contravariant indices (e.g.,  $\mu_1, \mu_2$ ) yields

$$T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q} = -T^{\mu_2\mu_1\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q},$$

then the tensor is said to be **antisymmetric** with respect to those **contravariant** indices. If swapping any two covariant indices (e.g.,  $\nu_1, \nu_2$ ) yields

$$T^{\mu_1\mu_2\cdots\mu_p}_{\nu_1\nu_2\cdots\nu_q} = -T^{\mu_1\mu_2\cdots\mu_p}_{\nu_2\nu_1\cdots\nu_q},$$

then the tensor is said to be **antisymmetric** with respect to those **covariant** indices.

## §2.4 Algebraic Operations on Tensors

### 1. Equality of Tensors

In the same coordinate system, two tensors of the same type are said to be equal if each corresponding component of the tensors is **equal**.

### 2. Zero Tensor

A tensor is called a **zero tensor** if all its components are zero.

### 3. Addition and Subtraction of Tensors

Only tensors of the same type can be added or subtracted, and the result is still a tensor of the same type. For two tensors  $A$  and  $B$  of the same type, their sum or difference yields a new tensor  $C$ :

$$A \pm B = C.$$

$C$  is a tensor of the same type as  $A$  and  $B$ . The components of  $C$  are the sum or difference of the corresponding components of  $A$  and  $B$ . For example,

$$C^{\mu\nu} = A^{\mu\nu} + B^{\mu\nu}, \quad T_\mu^\nu = D_\mu^\nu - E_\mu^\nu.$$

#### 4. Product of Tensors

The product of tensors is also called the **outer product**. Any two tensors, regardless of type, can be multiplied. For example,  $A_{\lambda\sigma\tau}^{\mu\nu}$  is a 5th-order tensor with three covariant indices and two contravariant indices, having a total of  $m^5$  components.  $B_\gamma^{\alpha\beta}$  is a 3rd-order tensor with one covariant index and two contravariant indices, having a total of  $m^3$  components. The tensor product of  $A_{\lambda\sigma\tau}^{\mu\nu}$  and  $B_\gamma^{\alpha\beta}$  yields a new tensor  $C$ :

$$C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta} = A_{\lambda\sigma\tau}^{\mu\nu} B_\gamma^{\alpha\beta}.$$

$C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta}$  is obtained as follows: each component of  $A_{\lambda\sigma\tau}^{\mu\nu}$  is multiplied successively by every component of  $B_\gamma^{\alpha\beta}$ , giving all components of  $C$ , totaling  $m^{5+3}$  components. The contravariant indices of  $A_{\lambda\sigma\tau}^{\mu\nu}$  and  $B_\gamma^{\alpha\beta}$  are combined to form the contravariant indices  $\mu, \nu, \alpha, \beta$  of  $C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta}$ , totaling 4; note that the contravariant indices of  $A_{\lambda\sigma\tau}^{\mu\nu}$  come first, followed by those of  $B_\gamma^{\alpha\beta}$ , because the tensor product is **not commutative**—the order cannot be swapped. Similarly, the covariant indices of  $A_{\lambda\sigma\tau}^{\mu\nu}$  and  $B_\gamma^{\alpha\beta}$  are combined, with the covariant indices of  $A_{\lambda\sigma\tau}^{\mu\nu}$  placed first and those of  $B_\gamma^{\alpha\beta}$  placed afterward, resulting in the covariant indices  $\lambda, \sigma, \tau, \gamma$  of  $C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta}$ , totaling 4. The resulting tensor  $C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta}$  is an 8th-order tensor of type (4,4).

The tensor product has the following properties:

- 1) The tensor product satisfies the **distributive law**, i.e.,  $(A + B)C = AC + BC$ ;
- 2) The tensor product satisfies the **associative law**, i.e.,  $(AB)C = A(BC)$ ;
- 3) The tensor product does **not** satisfy the commutative law.

For example, a matrix can be regarded as a tensor, and two matrices of the same order can be viewed as two tensors of the same order. As is well known, matrix multiplication does not commute; likewise, the tensor product does not commute.

#### 5. Contraction of Tensors

For mixed tensors, the operation of **contraction** can be performed. As an example, consider  $R_{\nu\lambda\sigma}^\mu$ , which is a fourth-order mixed tensor with one contravariant index and three covariant indices. We can choose the contravariant index  $\mu$  and one of the covariant indices, say  $\nu$ , to be the same index  $\alpha$ . Since an upper and a lower index with the same label implies summation over that repeated index, we obtain

$$R_{\nu\lambda\sigma}^\mu \rightarrow R_{\alpha\lambda\sigma}^\alpha = R_{1\lambda\sigma}^1 + R_{2\lambda\sigma}^2 + \cdots + R_{m\lambda\sigma}^m.$$

Through contraction, the tensor  $R_{\nu\lambda\sigma}^\mu$  becomes a second-order tensor. To summarize:

- 1) Contraction is applicable only to mixed tensors; other types of tensors cannot be contracted.
- 2) Contraction must specify which contravariant index and which covariant index are to be contracted; otherwise, the result may differ.
- 3) Contraction can be performed successively. Each contraction reduces the order of the tensor by two, decreasing the number of contravariant and covariant indices each by one.

## 6.Inner Product of Tensors

The **inner product** of tensors is not a new operation but a combination of first multiplying two tensors and then performing a contraction. The following example illustrates this.

The tensor  $A_{\lambda\sigma\tau}^{\mu\nu}$  and the tensor  $B_\gamma^{\alpha\beta}$  are multiplied to yield a new tensor  $C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta}$ , which is a mixed tensor and can be contracted. If we specify that the contravariant index  $\nu$  and the covariant index  $\gamma$  are to become the same index  $\eta$ , we obtain a new tensor  $D$ :

$$A_{\lambda\sigma\tau}^{\mu\nu} B_\gamma^{\alpha\beta} = C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta},$$

$$D = C_{\lambda\sigma\tau\eta}^{\mu\eta\alpha\beta} = C_{\lambda\sigma\tau 1}^{\mu 1\alpha\beta} + C_{\lambda\sigma\tau 2}^{\mu 2\alpha\beta} + \cdots + C_{\lambda\sigma\tau m}^{\mu m\alpha\beta}.$$

These two steps can be combined and written as

$$D = A_{\lambda\sigma\tau}^{\mu\eta} B_\eta^{\alpha\beta}.$$

The tensor  $A_{\lambda\sigma\tau}^{\mu\nu}$  is of order 5, and  $B_\gamma^{\alpha\beta}$  is of order 3. Their product gives an 8th-order tensor  $C_{\lambda\sigma\tau\gamma}^{\mu\nu\alpha\beta}$ . After contraction, it becomes a 6th-order tensor, with the number of contravariant and covariant indices each reduced by one.

Because the result of contraction depends on the specific contravariant and covariant indices chosen, the inner product of two tensors is not unique—different choices of indices to contract yield different results.

## §2.5 Flexible Use of Tensors

As we know, in a three-dimensional coordinate system, a contravariant vector can be expressed as

$$A = V^1 e_1 + V^2 e_2 + V^3 e_3,$$

where  $e_1, e_2, e_3$  are the basis vectors of the coordinate system. If a coordinate transformation is applied,

$$A = \frac{1}{2} V^1 2e_1 + \frac{1}{3} V^2 3e_2 + \frac{1}{4} V^3 4e_3,$$

let

$$\tilde{W}^1 = \frac{1}{2} V^1, \quad \tilde{W}^2 = \frac{1}{3} V^2, \quad \tilde{W}^3 = \frac{1}{4} V^3,$$

and

$$\tilde{e}_1 = 2e_1, \quad \tilde{e}_2 = 3e_2, \quad \tilde{e}_3 = 4e_3,$$

then, in the new coordinate system,

$$\tilde{A} = \tilde{W}^1 \tilde{e}_1 + \tilde{W}^2 \tilde{e}_2 + \tilde{W}^3 \tilde{e}_3,$$

Clearly,

$$A = \tilde{A}.$$

Therefore, the contravariant vector  $A$  is **invariant** under coordinate transformations—what change are its components  $A^i$  and the basis vectors  $e_i$  of the coordinate system. The same holds for other types of tensors: a tensor itself is invariant under coordinate transformations, while its components and the basis vectors of the coordinate system are what vary.

Any equation in which every term is a tensor is called a **tensor equation**. Since if a tensor is the zero tensor in one coordinate system, it must also be the zero tensor in any coordinate system, the form of a tensor equation is independent of the choice of coordinate system. In other words, the form of a tensor equation remains unchanged under coordinate transformations. For example, in the old coordinate system, a tensor equation may have the form

$$A - B = 0.$$

Then, after transforming to a new coordinate system, the tensor equation becomes

$$\tilde{A} - \tilde{B} = 0.$$

An observer or a laboratory exists in a specific coordinate system. Observers in different coordinate systems should obtain the same physical laws—that is, physical laws should be independent of the choice of coordinate system and possess objectivity. Because tensor equations written in terms of tensors are independent of the choice of coordinate system, they effectively embody this objectivity of physical laws.

This characteristic of tensor equations brings great convenience to our research. We can study certain physical phenomena under particularly simple conditions and in special coordinate systems. If, under these special circumstances, we can relatively easily derive a physical law and express it in the form of a tensor equation, then the obtained law can be applied to more complex situations and adapted to any coordinate system, because the form of a tensor equation remains unchanged under coordinate transformations.

This feature of tensor equations can also be used to test whether an existing physical law possesses universal significance. We can check whether a physical law, which has been summarized using a non-tensorial mathematical form, can be rewritten as a tensor equation. If it can be expressed as a tensor equation, it indicates that the law has universal significance. If it cannot, then it is worth reconsidering and re-examining the law.

## Chapter 3 Differentiable Manifolds

### §3.1 Functions in Euclidean Space

#### 1. $k$ -times Differentiable Functions

Let  $f$  be a real-valued function defined on an open set  $U \subseteq R^m$ . If all partial derivatives of  $f$  from order 1 to order  $k$  exist and are continuous, then  $f$  is said to be  **$k$ -times differentiable**, or  $f$  is of class  $C^k$ , or  $f$  is a  **$k$ -times differentiable function**, or simply a  $C^k$  function. A continuous function is called a 0-times differentiable function and is denoted as  $C^0$ .

#### 2. Smooth Functions

Let  $f$  be a real-valued function defined on an open set  $U \subseteq R^m$ . If  $f$  has continuous partial derivatives of **all orders**, then  $f$  is said to be **smooth** or of **class  $C^\infty$** , and  $f$  is called a **smooth function** or a  $C^\infty$  function.

**Example 3.1.1** The function  $f(x) = 3$  defined on the open set  $(-\infty, +\infty)$  is a smooth function.

Clearly, the function  $f(x)$  can be differentiated once:  $\frac{df}{dx} = 0$ , and its first derivative is

continuous—its graph is simply the  $x$ -axis. It can be differentiated again:  $\frac{d^2f}{dx^2} = 0$ , and this second derivative is also continuous, with its graph again being the  $x$ -axis. This process can be repeated indefinitely. In other words, the function  $f(x)$  possesses derivatives of all orders, and every derivative obtained is continuous. Therefore,  $f(x)$  is not only a  $C^k$  function but indeed a smooth function.

#### 3. Analytic Functions

Let  $f$  be a real-valued function defined on an open set  $U \subseteq R^m$ . If at every point of  $U$ ,  $f$  can be expanded into a convergent Taylor series within some neighborhood of that point, then  $f$  is called an **analytic function**, or  $f$  is said to be of **class  $C^\omega$** .

**Example 3.1.2** The exponential function  $f(x) = e^x$  can be expanded into a Taylor series:

$$f(x) = e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + \cdots,$$

whose interval of convergence is  $(-\infty, \infty)$ . Therefore,  $f(x) = e^x$  is an analytic function.

**Theorem 3.1.1** Establish a Cartesian coordinate system in the  $m$ -dimensional Euclidean space  $R^m$ . Let  $g$  be a smooth function in this coordinate system, and let point  $p$  be a point in the domain of  $g$  with coordinates  $(a^1, \dots, a^m)$ . Then the function  $g$  can be expressed as

$$g(x^1, \dots, x^m) = g(a^1, \dots, a^m) + \sum_{i=1}^m (x^i - a^i) \frac{\partial g(x^1, \dots, x^m)}{\partial x^i}.$$



## §3.2 Manifolds

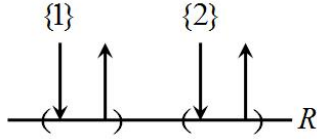
### 1. Topological Manifold

**Definition 3.2.1** Let  $M$  be a Hausdorff space. If for every point  $p \in M$ , there exists a neighborhood  $U$  of  $p$  in  $M$  that is homeomorphic to an open set of the  $m$ -dimensional Euclidean space  $R^m$ , then  $M$  is called an  **$m$ -dimensional manifold** or a **topological manifold**.

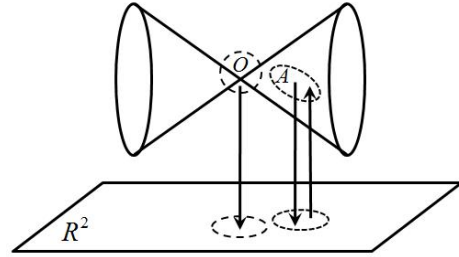
Here, the term “neighborhood” is sometimes written as “open neighborhood” or “open set” in other texts; they all mean the same.

All manifolds mentioned in this book are topological spaces that satisfy the second axiom of countability.

**Example 3.2.1** Consider a topological space  $(S, \tau)$ ,  $S = \{1, 2\}$ ,  $\tau = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$ . As shown in Figure 3.2.1, the open sets  $\{1\}$  and  $\{2\}$  are not in one-to-one correspondence with any open interval of the one-dimensional Euclidean space (the real line), and therefore they are not homeomorphic to such intervals. Consequently, the topological space  $(S, \tau)$  is **not** a manifold.



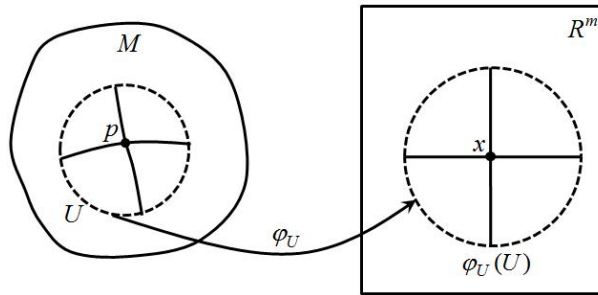
**Figure 3.2.1** The open sets  $\{1\}$  and  $\{2\}$  are not homeomorphic to any open interval



**Figure 3.2.2** A double cone is not a manifold

**Example 3.2.2** Figure 3.2.2 shows a two-dimensional figure formed by two cones joined at their tips. It is **not** a manifold because, although a neighborhood of point  $A$  is homeomorphic to an open set in the plane, a neighborhood of the intersection point  $O$  is not in one-to-one correspondence with any open set in the plane, and thus not homeomorphic to one.

### 2. Introducing Coordinates on a Manifold



**Figure 3.2.3** Establishing a coordinate system in a neighborhood  $U$

As shown in Figure 3.2.3, according to the definition of a manifold, a neighborhood  $U$  of a point  $p$  on the manifold  $M$  is homeomorphic to an open set in the  $m$ -dimensional Euclidean space  $R^m$ . Let this homeomorphism be  $\varphi: U \rightarrow \varphi_U(U)$ , where  $\varphi_U(U)$  is an open set in  $R^m$ . Because of the homeomorphism  $\varphi$ , every point  $p \in U$  corresponds one-to-one to a point  $\varphi_U(p)$ . Moreover, since every point in  $R^m$  corresponds to an  $m$ -tuple of real numbers  $x = (x^1, x^2, \dots, x^m)$ , the point  $p \in U$  also corresponds to an  $m$ -tuple in  $R^m$ . We can define this

array as the **coordinates** of the point  $p$ , i.e., we set

$$u^i = (\varphi_U(p))^i = x^i, \quad p \in U, \quad i = 1, 2, \dots, m.$$

We call  $u^i (1 \leq i \leq m)$  the **local coordinates** of the point  $p \in U$ . In this way, a local coordinate is assigned to every point in  $U$ , thus establishing a **coordinate system** on  $U$ , called the **local coordinate system** of  $U$  and denoted as  $(U; x^i)$ . The pair  $(U, \varphi_U)$  is called a **coordinate chart** on  $M$ .

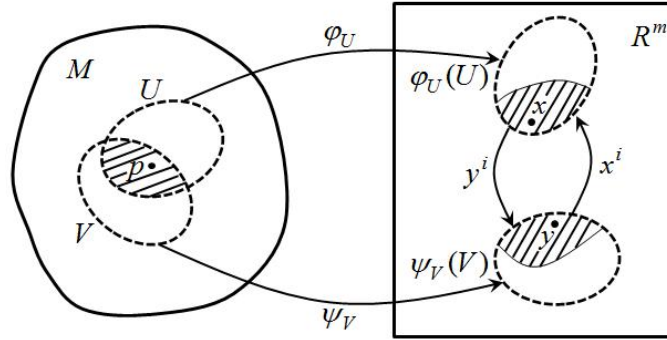
However, a point  $p$  does not have only one neighborhood; it may have several. As shown in Figure 3.2.4, let  $V$  be another neighborhood of  $p$  with  $U \cap V \neq \emptyset$ . Then there exists another homeomorphism  $\psi_V$ , mapping the neighborhood  $V$  to another open set  $\psi_V(V) \subset \mathbb{R}^m$ . Consequently, the point  $p$  also corresponds to another  $m$ -tuple  $y = (y^1, y^2, \dots, y^m)$  in  $\mathbb{R}^m$ , which can likewise serve as the coordinates of  $p$ , i.e., we set

$$v^i = (\psi_V(p))^i = y^i, \quad p \in V, \quad i = 1, 2, \dots, m.$$

This local coordinate system is denoted as  $(V; y^i)$ , and the corresponding coordinate chart is  $(V, \psi_V)$ .

Thus, the point  $p \in M$  possesses two local coordinate systems:  $(U; x^i)$  and  $(V; y^i)$ . The composite mapping  $\psi_V \circ \varphi_U^{-1}$  establishes a transformation relation between the coordinates  $x$  and  $y$ , namely

$$\begin{cases} y^1 = y^1(x^1, x^2, \dots, x^m), \\ y^2 = y^2(x^1, x^2, \dots, x^m), \\ \vdots \\ y^m = y^m(x^1, x^2, \dots, x^m). \end{cases} \quad (3.2.1)$$



**Figure 3.2.4** Compatible coordinate charts can be established at point  $p$

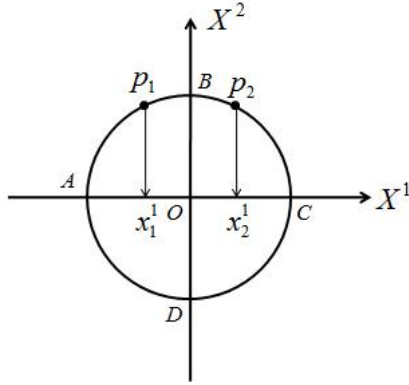
The inverse mapping  $\varphi_U \circ \psi_V^{-1}$  of the composite mapping  $\psi_V \circ \varphi_U^{-1}$  also gives an inverse transformation relation between the coordinates  $x$  and  $y$ , i.e.,

$$\begin{cases} x^1 = x^1(y^1, y^2, \dots, y^m), \\ x^2 = x^2(y^1, y^2, \dots, y^m), \\ \vdots \\ x^m = x^m(y^1, y^2, \dots, y^m). \end{cases} \quad (3.2.2)$$

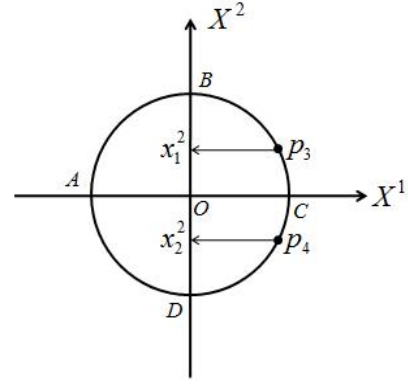
Since  $\psi_V \circ \varphi_U^{-1}$  and  $\varphi_U \circ \psi_V^{-1}$  are mutually inverse homeomorphisms, the functions (3.2.1) and (3.2.2) are both continuous functions.

$(U, \varphi_U)$  is one coordinate chart on  $M$ , and  $(V, \psi_V)$  is another coordinate chart. If  $U \cap V = \emptyset$ , or if  $U \cap V \neq \emptyset$  and the coordinate transformation functions  $y^i(x^1, x^2, \dots, x^m)$  and  $x^i(y^1, y^2, \dots, y^m)$  are both  $C^k$  functions, then we say the two charts  $(U, \varphi_U)$  and  $(V, \psi_V)$  are  **$C^k$ -compatible**.

**Example 3.2.3** As shown in Figure 3.2.5, the arc  $ABC$  (excluding the endpoints  $A$  and  $C$ ) is an open set of the circle  $S^1$ , denoted as  $U_{ABC}$ . There exists a one-to-one mapping (homeomorphism) between  $U_{ABC}$  and the line segment  $AC$  on the  $X^1$ -axis. Since point  $p_1$  maps to  $x_1^1$  and point  $p_2$  maps to  $x_2^1$ , we can take  $x_1^1$  as the coordinate of point  $p_1$  and  $x_2^1$  as the coordinate of point  $p_2$ .



**Figure 3.2.5** Coordinate system on the arc  $ABC$



**Figure 3.2.6** Coordinate system on the arc  $BCD$

Furthermore, as shown in Figure 3.2.6, the arc  $BCD$  (excluding the endpoints  $B$  and  $D$ ) is another open set of the circle  $S^1$ , denoted as  $U_{BCD}$ . There exists a one-to-one mapping (homeomorphism) between  $U_{BCD}$  and the line segment  $BD$  on the  $X^2$ -axis. Since point  $p_3$  maps to  $x_1^2$  and point  $p_4$  maps to  $x_2^2$ , we can take  $x_1^2$  as the coordinate of point  $p_3$  and  $x_2^2$  as the coordinate of point  $p_4$ .

### 3. Smooth Manifold

**Definition 3.2.2** Suppose an  $m$ -dimensional manifold  $M$  is given a collection  $\Sigma = \{(U, \varphi_U), (V, \psi_V), (W, \xi_W), \dots\}$  of coordinate charts satisfying the following conditions. Then  $\Sigma$  is called a  $C^k$  **differential structure** on  $M$ :

- 1)  $\{U, V, W, \dots\}$  forms an open cover of  $M$ ;
- 2) Any two charts in  $\Sigma$  are  $C^k$ -compatible;
- 3)  $\Sigma$  is **maximal**: if any coordinate chart  $(\tilde{U}, \varphi_{\tilde{U}})$  on  $M$  is  $C^k$ -compatible with every chart in  $\Sigma$ , then  $(\tilde{U}, \varphi_{\tilde{U}})$  itself belongs to  $\Sigma$ .

If a  $C^k$  differential structure  $\Sigma$  is given on  $M$ , then  $M$  is called a  $C^k$  **-differentiable manifold**.

If  $k = \infty$ , then  $\Sigma$  is called a **smooth differential structure**, and  $M$  is called a **smooth manifold**.

If  $k = \omega$  (i.e., the coordinate transformations are analytic), then  $M$  is called an **analytic manifold**.

The requirement in the definition that  $\Sigma$  be **maximal** merely means that  $\Sigma$  is open and inclusive. When constructing a differential structure  $\Sigma$  on a manifold  $M$ , we are not required to list every compatible coordinate chart. Many manifolds have infinitely many compatible charts, and trying to find all of them is often difficult and unnecessary—it suffices to find a compatible coordinate cover for the manifold. For example, suppose we know that a manifold  $M$  has two compatible charts  $(U, \varphi_U)$  and  $(V, \psi_V)$  with  $U \cup V = M$ . Then we can take the differential structure of  $M$  to be the chart set  $\Sigma = \{(U, \varphi_U), (V, \psi_V)\}$ . Later, if we discover another chart  $(W, \xi_W)$  that is compatible with every chart in  $\Sigma = \{(U, \varphi_U), (V, \psi_V)\}$ ,  $U \cup V \cup W = M$ , we may include it, updating the differential structure to  $\Sigma = \{(U, \varphi_U), (V, \psi_V), (W, \xi_W)\}$ . Alternatively, if

from the beginning we know that  $M$  has three compatible charts  $(U, \varphi_U)$ ,  $(V, \psi_V)$ , and  $(W, \xi_W)$  with  $U \cup V = M$ ,  $U \cup V \cup W = M$ , we may simply take the differential structure as  $\Sigma = \{(U, \varphi_U), (V, \psi_V)\}$  for convenience.

**Example 3.2.3** As shown in Figure 3.2.7, on the unit circle  $S^1$  we can take four coordinate charts:

Arc  $ABC$  (excluding endpoints  $A$  and  $C$ ) is denoted as  $U_{ABC}$ , with coordinate mapping  $\varphi_{U_{ABC}}(x) = x^1$ ;

Arc  $BCD$  (excluding endpoints  $B$  and  $D$ ) is denoted as  $U_{BCD}$ , with coordinate mapping  $\varphi_{U_{BCD}}(x) = x^2$ ;

Arc  $ADC$  (excluding endpoints  $A$  and  $C$ ) is denoted as  $U_{ADC}$ , with coordinate mapping  $\varphi_{U_{ADC}}(x) = x^1$ ;

Arc  $BDC$  (excluding endpoints  $B$  and  $D$ ) is denoted as  $U_{BAD}$ , with coordinate mapping  $\varphi_{U_{BAD}}(x) = x^2$ .

On the intersection  $U_{ABC} \cap U_{BCD}$ ,

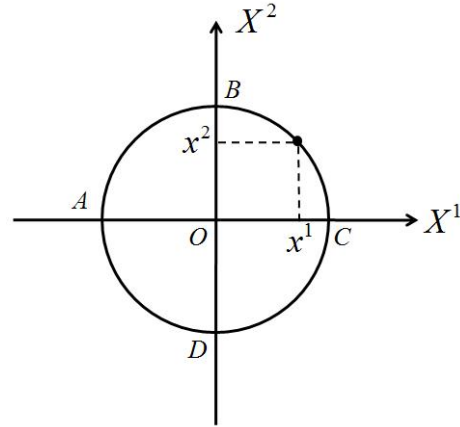
$$x^1 = \sqrt{1 - (x^2)^2}, \quad x^2 = \sqrt{1 - (x^1)^2}, \quad x^1 > 0, \quad x^2 > 0.$$

Here  $x^1$  is coordinate function on the open set  $U_{ABC}$ , and  $x^2$  is coordinate function on the open set  $U_{BCD}$ . Both functions are  $C^\infty$  (smooth). Therefore, the charts  $(U_{ABC}, \varphi_{U_{ABC}})$  and  $(U_{BCD}, \varphi_{U_{BCD}})$  are  $C^\infty$ -compatible. Similarly, every pair of the four charts is  $C^\infty$ -compatible. Thus, the differential structure formed by these four charts makes the unit circle  $S^1$  a **one-dimensional smooth manifold**.

Analogously, the  $m$ -dimensional unit sphere in  $R^{m+1}$ ,

$$S^m = \{x \in R^{m+1} \mid (x^1)^2 + \cdots + (x^{m+1})^2 = 1\},$$

is also a **smooth differentiable manifold**.



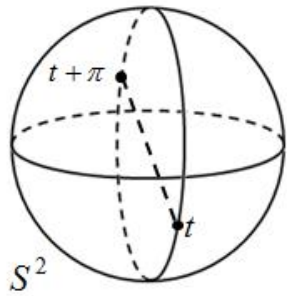
**Figure 3.2.7** Compatible coordinate charts on the circle  $S^1$

### §3.3 Real Projective Space

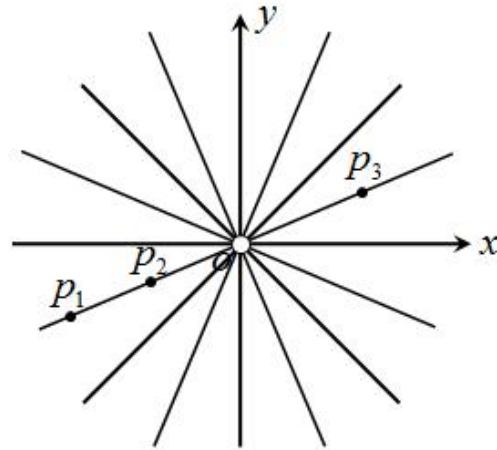
As shown in Figure 3.3.1, on a two-dimensional sphere, if every pair of antipodal points (two points symmetric about the center) are glued together, the resulting geometric object is called the **two-dimensional real projective space**. Similarly, on an  $m$ -dimensional sphere, gluing together every pair of antipodal points yields a geometric object called the  **$m$ -dimensional real projective space**, denoted as  $RP^m$ .

The  $m$ -dimensional real projective space  $RP^m$  can only be realized in the space  $R^{m+1}$  with the origin removed, denoted as  $R^{m+1} - \{0\}$  or  $R^{m+1} \setminus \{0\}$ . In fact,  $RP^m$  is the space formed by all lines passing through the origin in  $R^{m+1} - \{0\}$ . Specifically,  $RP^m$  consists of all lines through the origin in  $R^{m+1}$  (excluding the origin itself), and each such line is regarded as a single point in  $RP^m$ . Let us illustrate this feature with the one-dimensional real projective line  $RP^1$ . Although  $RP^1$  is one-dimensional, it must be represented using a two-dimensional plane. Figure 3.3.2 shows the one-dimensional real projective line. From the perspective of the Euclidean plane,

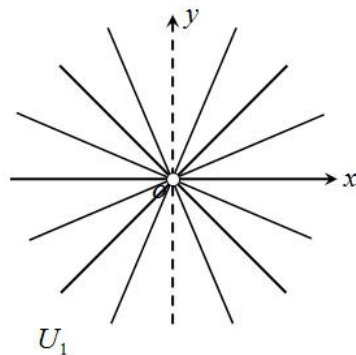
points  $p_1$ ,  $p_2$ , and  $p_3$  are distinct. However, from the viewpoint of  $RP^1$ , since these three points lie on the same line, they all belong to the same point. All points on the same line are considered the same point in  $RP^1$ ; that is, all points on the same line are **equivalent**. They are said to belong to the same equivalence class, denoted as  $[(x, y)]$ , where  $[(x, y)]$  represents all points on the same line in Figure 3.3.2.



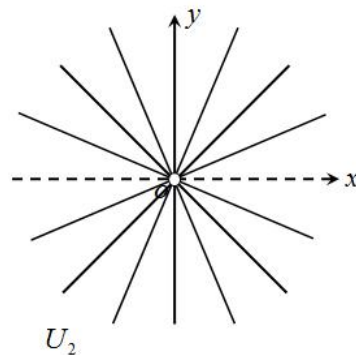
**Figure 3.3.1** Identifying antipodal points on  $S^2$  yields  $RP^2$



**Figure 3.3.2** The  $R^2$  model of  $RP^1$



**Figure 3.3.3** A coordinate chart  $U_1$  on  $RP^1$



**Figure 3.3.4** A coordinate chart  $U_2$  on  $RP^1$

Let us take the one-dimensional real projective line  $RP^1$  as an example to illustrate the smooth differential structure of  $RP^m$ . Figure 3.3.3 and Figure 3.3.4 can serve as two coordinate domains for  $RP^1$ . Removing the  $y$ -axis from Figure 3.3.2 yields Figure 3.3.3, which is an open set and can be taken as one coordinate domain of  $RP^1$ , denoted as

$$U_1 = \{[(x, y)] | (x, y) \in R^2 - \{0\}, x \neq 0\}.$$

Removing the  $x$ -axis from Figure 3.3.2 gives Figure 3.3.4, also an open set, which can serve as another coordinate domain of  $RP^1$ , denoted as

$$U_2 = \{[(x, y)] | (x, y) \in R^2 - \{0\}, y \neq 0\}.$$

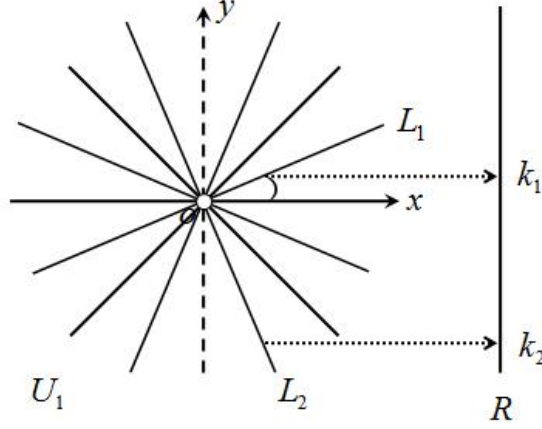
The union of these two coordinate domains equals  $RP^1$ :

$$U_1 \cup U_2 = RP^1.$$

Let us specify the coordinate mapping for the chart  $(U_1, \phi_1)$ . As shown in Figure 3.3.5, every line has a unique slope  $k$ . Since all points on the same line are equivalent, we take the slope  $k$  of

each line as the coordinate for all points on that line. Define

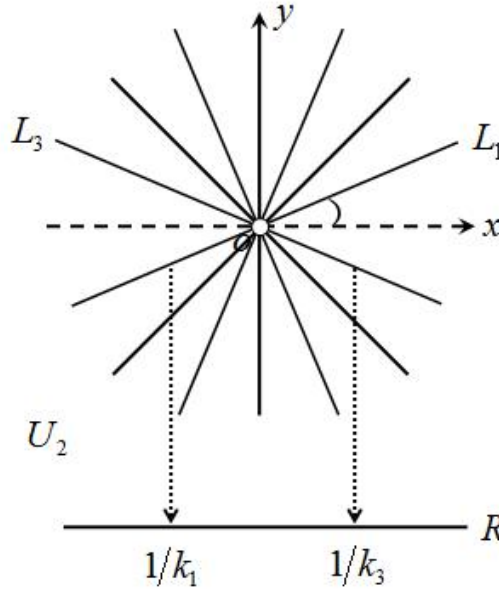
$$\varphi_1([(x, y)]) = \xi_1 = k = \frac{y}{x}, \quad \xi_1 \in \mathbb{R}.$$



**Figure 3.3.5** Coordinate mapping for the chart  $(U_1, \varphi_1)$

Now, let us specify the coordinate mapping for the chart  $(U_2, \varphi_2)$ . As shown in Figure 3.3.6, the  $y$ -axis is included in  $U_2$ , but the slope of the  $y$ -axis is infinite and cannot correspond to any point on the real number line  $\mathbb{R}$  (since  $\mathbb{R}$  does not include a point at infinity). Therefore, we take the **reciprocal of the slope** of each line in  $U_2$  as the coordinate for every point (i.e., all points on that line) in  $U_2$ . Define

$$\varphi_2([(x, y)]) = \xi_2 = \frac{1}{k} = \frac{x}{y}, \quad \xi_2 \in \mathbb{R}.$$



**Figure 3.3.6** Coordinate mapping for the chart  $(U_2, \varphi_2)$

Therefore, every point in the intersection  $U_1 \cap U_2$  of these two charts has two coordinates  $\xi_1$  and  $\xi_2$ , and the coordinate transformation between them is

$$\xi_1 = \frac{1}{\xi_2}.$$

Clearly, this transformation  $\xi_1$  is  $C^\infty$  (smooth). Hence, the differential structure

$\{(U_1, \varphi_1), (U_2, \varphi_2)\}$  we have constructed makes  $RP^1$  a **smooth differentiable manifold**.

Finally, note that  $RP^m$  can also be viewed as the union of  $R^m$  and  $RP^{m-1}$ , i.e.,  

$$RP^m = R^m \cup RP^{m-1}.$$

### §3.4 Product Manifolds

The direct product of two smooth manifolds yields a new manifold. Let  $M$  and  $N$  be smooth manifolds of dimensions  $m$  and  $n$ , respectively, with smooth differential structures

$$\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}, \quad \{(V_\beta, \psi_\beta)\}_{\beta \in B}.$$

The following procedure makes the topological product space  $M \times N$  into a new  $(m+n)$ -dimensional smooth manifold  $M \times N$ .

First, the open sets  $U_\alpha$  of  $M$  and  $V_\beta$  of  $N$  are taken as direct products, forming the collection  $\{(U_\alpha \times V_\beta)\}_{\alpha \in A, \beta \in B}$ , which provides an open cover of the product space  $M \times N$ .

Second, the coordinates of points from the original manifolds  $M$  and  $N$  are combined to serve as coordinates for points on  $M \times N$ . That is, we define the mapping

$$\varphi_\alpha \times \psi_\beta : U_\alpha \times V_\beta \rightarrow R^{m+n},$$

such that

$$\varphi_\alpha \times \psi_\beta(p, q) = (\varphi_\alpha(p), \psi_\beta(q)), \quad (p, q) \in U_\alpha \times V_\beta.$$

Each such pair  $(U_\alpha \times V_\beta, \varphi_\alpha \times \psi_\beta)$  then becomes a coordinate chart on  $M \times N$ . Because the original charts of the two manifolds are  $C^\infty$ -compatible, the newly constructed charts are also  $C^\infty$ -compatible. Therefore, the collection  $(U_\alpha \times V_\beta, \varphi_\alpha \times \psi_\beta)$  determines a smooth manifold structure on  $M \times N$ .

**Definition 3.4.1** The  $C^\infty$ -compatible coordinate cover

$$\{(U_\alpha \times V_\beta, \varphi_\alpha \times \psi_\beta)\}_{\alpha \in A, \beta \in B}$$

constructed on the topological product space  $M \times N$  determines a smooth manifold structure on  $M \times N$ , making  $M \times N$  an  $(m+n)$ -dimensional smooth manifold, called the **product manifold** of  $M$  and  $N$ .

From the product manifold  $M \times N$ , the factor manifolds  $M$  and  $N$  can be recovered. This is done by taking the **natural projections** from the product manifold onto each factor:

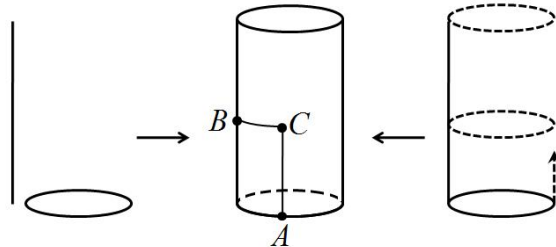
$$\pi_1 : M \times N \rightarrow M, \quad \pi_2 : M \times N \rightarrow N.$$

That is, for  $(x, y) \in M \times N$ , we have

$$\pi_1(x, y) = x, \quad \pi_2(x, y) = y.$$

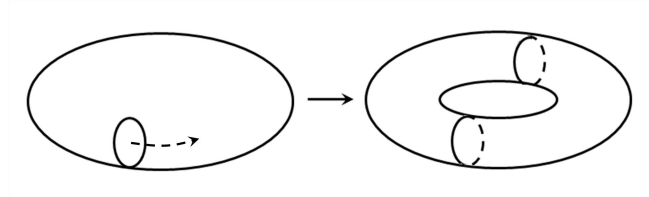
Both of these natural projections are **smooth mappings** (see §5.1).

**Example 3.4.1** The product of a line segment and a circle yields a cylinder. As shown in Figure 3.4.1, taking a point  $A$  on the circle and a point  $B$  on the line segment, their product corresponds to point  $C$ . The line segment and the circle are each one-dimensional, so the resulting cylinder is two-dimensional. A cylinder can also be viewed as a circle moving along a line segment, as illustrated in the right-hand side of Figure 3.4.1.



**Figure 3.4.1** A cylinder is the product manifold of a line segment and a circle

**Example 3.4.2** The product manifold of two circles is a torus, denoted as  $T^2 \cong S^1 \times S^1$ . As shown in Figure 3.4.2, a small circle revolving around a large circle generates a two-dimensional torus.



**Figure 3.4.2** A torus is the product manifold of two circles.

**Example 3.4.3** The product manifold of  $m$  circles yields an  $m$ -dimensional torus. Let

$$T^m = \{(z_1, \dots, z_m) \in C^m \mid |z_i| = 1, 1 \leq i \leq m\},$$

where  $C^m$  is a one-dimensional circle. Then  $T^m$  is an  $m$ -dimensional torus.

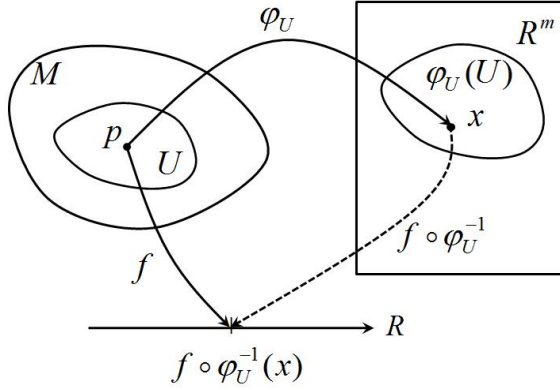


## Chapter 4 Tangent Vectors and Cotangent Vectors

### §4.1 Smooth Functions on Manifolds

**Definition 4.1.1** A **function**  $f$  on a manifold  $M$  is a mapping from  $M$  to  $R$ .

As a mapping,  $f$  assigns a real number to each point of  $M$ ; this real number is called the **value** of the function.



**Figure 4.1.1** The function  $f$  on  $M$  is transformed into a function on  $R^m$

Using a **compatible** coordinate chart on the manifold  $M$ , a function  $f$  defined on  $M$  can be transformed into a function defined on an open set of  $R^m$ .

As shown in Figure 4.1.1, let  $(U, \varphi_U)$  be a **compatible** coordinate chart of the manifold  $M$ , and let point  $p \in M$  correspond to  $p \in U$ . The function  $f \circ \varphi_U^{-1}$  defined on the open set  $\varphi_U(U) \subset R^m$ , which is a mapping

$$f \circ \varphi_U^{-1} : \varphi_U(U) \rightarrow R.$$

If the function  $f \circ \varphi_U^{-1}$  smooth ( $C^\infty$ ) at the point  $x = \varphi_U(p) \in R^m$ , then the function  $f$  is said to be **smooth** ( $C^\infty$ ) at the point  $p \in M$ .

The domains (open sets) of different functions defined at point  $p \in M$  may differ, because there is not a unique open set containing  $p \in M$  on the manifold  $M$ . For example, suppose  $U, V$ , and  $W$  are open sets containing  $p \in M$ , and the mappings  $f : U \rightarrow R$ ,  $g : V \rightarrow R$ , and  $h : W \rightarrow R$  are real-valued functions. Clearly, the domains of these three functions are different, but since  $p \in U \cap V \cap W$  belongs to all of them, each is considered a function defined at the point  $p \in M$ .

The set of all smooth ( $C^\infty$ ) functions defined at point  $p$  on the manifold  $M$  is denoted by  $C_p^\infty$ . Given two smooth functions  $f, g \in C_p^\infty$  defined at  $p$ , with domains  $U$  and  $V$  respectively, since  $U \cap V \neq \emptyset$ , both  $f$  and  $g$  can be regarded as  $C^\infty$  functions on the open set  $U \cap V$ . Because  $p \in U \cap V$ , we have  $f + g, fg \in C_p^\infty$ .

**Definition 4.1.2** Let  $f : M \rightarrow R$  be a real-valued function defined on a smooth manifold  $M$ . If  $f$  is smooth ( $C^\infty$ ) at **every** point of  $M$ , then  $f$  is called a **smooth** ( $C^\infty$ ) function on  $M$ .

The set of all smooth functions on  $M$  is denoted by  $C^\infty(M)$ .

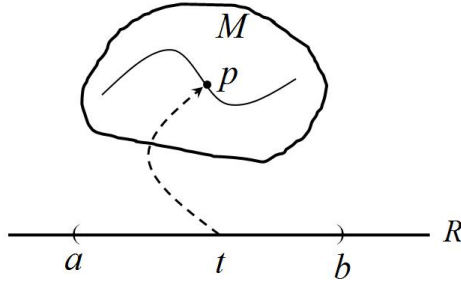
### §4.2 Tangent Vectors and Tangent Spaces

#### 1. Parametric Curves on a Manifold

Take an open interval  $I = (a, b)$ . A smooth mapping

$$f : I \rightarrow M$$

from  $I$  to the manifold  $M$  is called a **parametric curve** on  $M$ , as shown in Figure 4.2.1.



**Figure 4.2.1** A parametric curve on a manifold  $M$

## 2. Tangent Vector

**Definition 4.2.1** Let  $f, g \in C_p^\infty$ ,  $\alpha, \beta \in \mathbb{R}$ ,  $p \in M$ , and  $X : C_p^\infty \rightarrow \mathbb{R}$ . A mapping  $X$  is called a **tangent vector** at point  $p$  if it satisfies the following conditions:

- 1)  $X(\alpha f + \beta g) = \alpha(Xf) + \beta(Xg)$  ;
- 2)  $X(fg) = fXg + gXf$  .

## 3. Tangent Space

Let  $M$  be an  $m$ -dimensional smooth manifold. Tangent vectors at different points (e.g., at  $p \in M$  and at  $q \in M$ ) cannot be added or subtracted. However, tangent vectors at the **same** point  $p$  can be added and scaled. We denote the set of all tangent vectors at point  $p \in M$  by  $T_p M$  .

For any  $X, Y \in T_p M$  ,  $f \in C_p^\infty$  and  $\lambda \in \mathbb{R}$  , we define addition and scalar multiplication in  $T_p M$  as follows:

$$\begin{aligned} (X + Y)(f) &= X(f) + Y(f) ; \\ (\lambda X)(f) &= \lambda(X(f)) . \end{aligned}$$

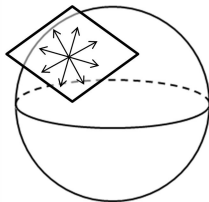
It can be verified that  $X + Y$  and  $\lambda Y$  are still tangent vectors at point  $p \in M$  . For example,

$$\begin{aligned} (X + Y)(\alpha f + \beta g) &= (X + Y)\alpha f + (X + Y)\beta g = \alpha((X + Y)f) + \beta((X + Y)g) , \\ (X + Y)(fg) &= f(X + Y)g + g(X + Y)f , \end{aligned}$$

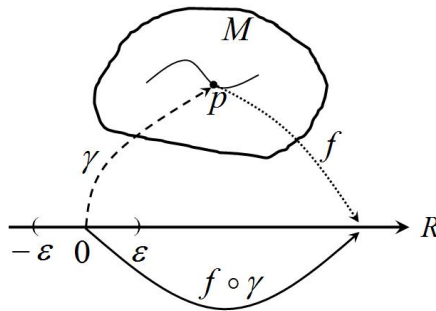
so  $X + Y$  is a tangent vector at  $p \in M$  , i.e.,  $X + Y \in T_p M$  . Similarly, one can verify  $\lambda Y \in T_p M$  . Therefore, with these definitions of addition and scalar multiplication,  $T_p M$  is closed under the operations, making it a **real vector space**.

**Definition 4.2.2** The set  $T_p M$  consisting of all tangent vectors at point  $p \in M$  of a smooth manifold  $M$  is called the **tangent space** of  $M$  at  $p$ .

The dimension of the tangent space  $T_p M$  is equal to the dimension of the smooth manifold  $M$ .



**Figure 4.2.2** Tangent space



**Figure 4.2.3**  $f \circ \gamma$  is the composite mapping of  $f$  and  $\gamma$

**Example 4.2.1** As shown in Figure 4.2.2, the tangent space at a point on the sphere  $S^2$  is

a linear space composed of infinitely many tangent vectors, all of which are tangent to the sphere at that same point. Since the sphere is two-dimensional, the tangent space is also two-dimensional—it is a plane.

#### 4. Tangent Vector of a Curve

Any tangent vector at point  $p \in M$  on a smooth manifold  $M$  can be expressed as the tangent vector of a smooth curve passing through  $p$  on  $M$ .

As shown in Figure 4.2.3, let  $\gamma: (-\varepsilon, \varepsilon) \rightarrow M$  be a smooth curve on the  $m$ -dimensional smooth manifold  $M$  passing through point  $p$ , with parameter  $t$  satisfying  $p = \gamma(0)$ . Let  $f \in C_p^\infty$  be a smooth function at  $p$ . The composite mapping  $f \circ \gamma$  is defined as

$$f \circ \gamma: R \rightarrow R.$$

Define the mapping  $X: C_p^\infty \rightarrow R$  by

$$X(f) = \left. \frac{d(f \circ \gamma(t))}{dt} \right|_{t=0}.$$

It is straightforward to verify that the mapping  $X: C_p^\infty \rightarrow R$  satisfies the conditions of Definition 4.2.1.

For example, let  $f, g \in C_p^\infty$ . Then

$$X(f+g) = \left. \frac{d}{dt}(f+g)(\gamma(t)) \right|_{t=0} = \left. \frac{d}{dt}f(\gamma(t)) \right|_{t=0} + \left. \frac{d}{dt}g(\gamma(t)) \right|_{t=0} = X(f) + X(g),$$

and

$$\begin{aligned} X(fg) &= \left. \frac{d(fg \circ \gamma)}{dt} \right|_{t=0} = \left. \frac{d((f \circ \gamma)(g \circ \gamma))}{dt} \right|_{t=0} \\ &= (f \circ \gamma(0)) \left. \frac{d(g \circ \gamma)}{dt} \right|_{t=0} + (g \circ \gamma(0)) \left. \frac{d(f \circ \gamma)}{dt} \right|_{t=0} = f(p)X(g) + g(p)X(f). \end{aligned}$$

Therefore,  $X$  is a tangent vector at point  $p$  on the smooth manifold  $M$ . Since  $p = \gamma(0)$ ,  $X$  is also called the **tangent vector of the curve  $\gamma$**  at point  $t = 0$ , denoted as  $X = \gamma'(0)$ . Thus,

$$\gamma'(0)(f) = \left. \frac{df(\gamma(t))}{dt} \right|_{t=0}.$$

Suppose a local coordinate system  $(U; x^i)$  established around  $p \in M$ , then the curve  $\gamma$  is described by the equations  $x^i = x^i(t)$  ( $i = 1, \dots, m$ ). Thus,  $f \circ \gamma$  is a function

$$g(t) = f(x^1(t), \dots, x^n(t)).$$

Differentiating with respect to  $t$  gives

$$\frac{dg}{dt} = \sum_{i=1}^m \frac{dx^i}{dt} \frac{\partial f}{\partial x^i}.$$

Since the function  $g(t)$  is arbitrary, we obtain

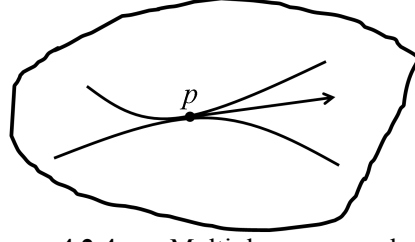
$$\frac{d}{dt} = \sum_{i=1}^m \frac{dx^i}{dt} \frac{\partial}{\partial x^i}.$$

Thus,  $\frac{d}{dt}$  is the tangent vector of the curve  $\gamma$  on the manifold  $M$ . At the point  $p$  ( $t = 0$ ), the tangent vector is

$$\gamma'(0) = \left. \frac{d}{dt} \right|_{t=0} = \sum_{i=1}^m \frac{dx^i(0)}{dt} \frac{\partial}{\partial x^i},$$

where the components are  $\frac{dx^i(0)}{dt}$ .

On the manifold  $M$ , there are usually more than one curve passing through point  $p \in M$ . Therefore, it is possible that multiple curves passing through  $p$  are tangent to one another and share the same tangent vector at  $p$ , as shown in Figure 4.2.4. Of course, there also exist different curves through  $p$  that have distinct tangent vectors.



**Figure 4.2.4** Multiple curves share a common tangent vector at point  $p$

Suppose there is another curve  $x^i = x^i(s)$  ( $i = 1, \dots, m$ ) on  $M$  that passes through point  $p$  at  $s = 0$ . Then its tangent vector at point  $p$  is

$$\frac{d}{ds} = \sum_{i=1}^m \frac{dx^i(0)}{ds} \frac{\partial}{\partial x^i}. \quad (4.2.1)$$

Let  $a$  and  $b$  be real numbers. At point  $p$ , the following holds:

$$a \frac{d}{dt} + b \frac{d}{ds} = \sum_{i=1}^m \left( a \frac{dx^i(0)}{dt} + b \frac{dx^i(0)}{ds} \right) \frac{\partial}{\partial x^i}.$$

The expression  $a \frac{d}{dt} + b \frac{d}{ds}$  is a linear combination of two tangent vectors at  $p$ , so it is also a tangent vector at  $p$ , and therefore corresponds to the tangent vector of some curve passing through  $p$ . Suppose this curve is parameterized by  $\lambda$ . At point  $p$ , we have

$$\frac{d}{d\lambda} = \sum_{i=1}^m \left( a \frac{dx^i(0)}{dt} + b \frac{dx^i(0)}{ds} \right) \frac{\partial}{\partial x^i},$$

that is,

$$a \frac{d}{dt} + b \frac{d}{ds} = \frac{d}{d\lambda}.$$

Thus, the set of all tangent vectors of curves at point  $p$  forms a vector space, which is isomorphic to the space of all tangent vectors at  $p$ . Consequently, a tangent vector at point  $p$  on manifold  $M$  is the tangent vector of some curve on  $M$  passing through  $p$ ; conversely, the tangent vector of a curve at point  $p$  on  $M$  is precisely a tangent vector at  $p$  on  $M$ .

It can be verified that  $\frac{\partial}{\partial x^i}$  is also a tangent vector. Equation (4.2.1) shows that any tangent vector  $\frac{d}{ds}$  can be expressed as a linear combination of  $\frac{\partial}{\partial x^i}$ . In fact,  $\frac{\partial}{\partial x^i}$  are the derivatives along the coordinate curves at point  $p$ ; they are linearly independent, so  $\left\{ \frac{\partial}{\partial x^i}; 1 \leq i \leq m \right\}$  can serve as a basis for the tangent vector space at  $p$ . The components of the tangent vector  $\frac{d}{ds}$  in this coordinate basis  $\frac{\partial}{\partial x^i}$  are precisely the values obtained by applying  $\frac{d}{ds}$  to the  $i$ -th coordinate function  $x^i$ , i.e.,  $\frac{dx^i(0)}{ds}$ .

**Theorem 4.2.1** Let  $(U; x^i)$  be a **compatible** local coordinate system around point  $p \in M$  on an  $m$ -dimensional smooth manifold  $M$ . Then the  $m$  tangent vectors

$$\frac{\partial}{\partial x^i}, \quad 1 \leq i \leq m$$

at point  $p$  form a basis of the tangent space  $T_p M$ , called the **natural basis** of  $T_p M$  with respect to the local coordinate system  $(U; x^i)$ .

Sometimes  $\frac{\partial}{\partial x^i}$  is written as  $\frac{\partial}{\partial x^i}\bigg|_p$  to emphasize that it is a basis vector at point  $p$ .

Using the natural basis, a tangent vector  $X$  at point  $p \in M$  can be expressed as

$$X = \sum_{i=1}^m \xi^i \frac{\partial}{\partial x^i}\bigg|_p,$$

where  $\xi^i$  are the **components** of the tangent vector  $X$  with respect to the natural basis, given by

$$\xi^i = \frac{d(x^i \circ \gamma)}{dt}.$$

**Example 4.2.2** As shown in Figure 4.2.5, on the plane, the parametric equation of a unit circle is

$$x = \cos t, \quad y = \sin t.$$

When  $t = 0$ , we have  $x = \cos 0 = 1$ ,  $y = \sin 0 = 0$ , corresponding to point  $p$  on the circle. Let us compute the tangent vector of the circle at point  $p$ .

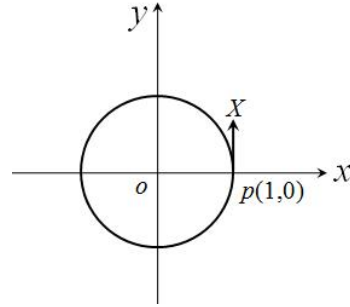
The basis vector along the  $x$ -axis is  $\frac{\partial}{\partial x}$ , and the corresponding component is

$$\frac{dx(0)}{dt} = \frac{d \cos t}{dt}\bigg|_{t=0} = -\sin t\bigg|_{t=0} = 0.$$

The basis vector along the  $y$ -axis is  $\frac{\partial}{\partial y}$ , and the corresponding component is

$$\frac{dy(0)}{dt} = \frac{d \sin t}{dt}\bigg|_{t=0} = \cos t\bigg|_{t=0} = 1.$$

Therefore, the tangent vector at point  $p$  is  $X = 0 \cdot \frac{\partial}{\partial x} + 1 \cdot \frac{\partial}{\partial y} = \frac{\partial}{\partial y}$ .



**Figure 4.2.5** Tangent vector on the unit circle

### §4.3 Cotangent Vectors and Cotangent Spaces

#### 1. Cotangent Vector

**Definition 4.3.1** Let  $T_p M$  be the tangent space at point  $p$  on a smooth manifold  $M$ . A mapping

$$\alpha : T_p M \rightarrow R$$

is called a **cotangent vector** of  $M$  at point  $p$ . The set of all cotangent vectors at point  $p$ , denoted by  $T_p^* M$ , is called the **cotangent space** of  $M$  at  $p$ .

The action of a cotangent vector  $\alpha \in T_p^* M$  on a tangent vector  $X \in T_p M$  is denoted by

$$\alpha(X) = \langle X, \alpha \rangle.$$

The pairing of a tangent vector  $X \in T_p M$  and a cotangent vector  $\alpha \in T_p^* M$  produces a real number  $\alpha(X)$ . Therefore,  $\langle X, \alpha \rangle$  can be viewed as a mapping

$$\langle \cdot, \cdot \rangle : T_p M \times T_p^* M \rightarrow R,$$

called the **pairing** between  $T_p M$  and  $T_p^* M$ . The value  $\langle X, \alpha \rangle$  is precisely the **contraction** (or inner product) of the tangent vector  $X$  and the cotangent vector  $\alpha$ .

Let  $f \in C_p^\infty$ ,  $X \in T_p M$ . Define the mapping  $(df)_p : T_p M \rightarrow R$  at point  $p$  by

$$(\mathrm{d}f)_p(X) = \langle X, (\mathrm{d}f)_p \rangle = X(f) \in \mathbb{R}. \quad (4.3.1)$$

Then  $(\mathrm{d}f)_p$  is a linear functional on  $T_p M$ . For example,

$$\begin{aligned} (\mathrm{d}f)_p(aX + bY) &= \langle aX + bY, (\mathrm{d}f)_p \rangle = (aX + bY)(f) \\ &= aX(f) + bY(f) = a(\mathrm{d}f)_p(X) + b(\mathrm{d}f)_p(Y). \end{aligned}$$

Therefore,  $(\mathrm{d}f)_p \in T_p^* M$ . Here,  $(\mathrm{d}f)_p$  is precisely the **gradient** (or **differential**) of the function  $f$  at point  $p$ .

## 2. Basis of the Cotangent Space

Suppose a **compatible** local coordinate system  $(U; x^i)$ ,  $p \in U$  is established at point  $p \in M$  on a smooth manifold  $M$ , with  $x^i \in C^\infty(U)$ . Since each coordinate function  $x^i$  is smooth on  $U$ , we have  $(\mathrm{d}x^i)_p \in T_p^* M$ . The set  $\{(\mathrm{d}x^i)_p; 1 \leq i \leq m\}$  linearly independent and forms a basis of  $T_p^* M$ ; any cotangent vector in  $T_p^* M$  can be expressed as a linear combination of them. For example, let  $f \in C_p^\infty$ . In the local coordinate system  $(U; x^i)$ , the differential of  $f$  is given by

$$(\mathrm{d}f)_p = \sum_{i=1}^m \left( \frac{\partial f}{\partial x^i} \right)_p (\mathrm{d}x^i)_p.$$

Thus, the cotangent vector  $(\mathrm{d}f)_p \in T_p^* M$  is exactly the differential of the function  $f$  at point  $p$ .

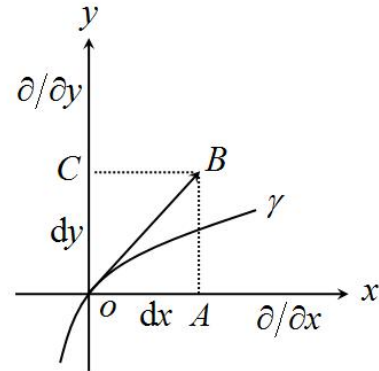
Now, let  $X_p = \left. \frac{\partial}{\partial x^i} \right|_p$ . From equation (4.3.1) we obtain

$$\left\langle \left. \frac{\partial}{\partial x^j} \right|_p, (\mathrm{d}x^i)_p \right\rangle = \left. \frac{\partial x^i}{\partial x^j} \right|_p = \delta_j^i,$$

Therefore,  $\{(\mathrm{d}x^i)_p; 1 \leq i \leq m\}$  is the **dual basis** of  $T_p^* M$  to the natural basis  $\left\{ \left. \frac{\partial}{\partial x^i} \right|_p; 1 \leq i \leq m \right\}$  of  $T_p M$ .

**Example 4.3.1** As shown in Figure 4.3.1, on the plane, the tangent vector  $\frac{d}{dt}$  has basis  $\frac{\partial}{\partial x}$  along the  $x$ -axis and basis  $\frac{\partial}{\partial y}$  along the  $y$ -axis. The cotangent vector  $\alpha$  has basis  $\mathrm{d}x$  along the  $x$ -axis (corresponding to segment  $OA$  in the figure) and basis  $\mathrm{d}y$  along the  $y$ -axis (corresponding to segment  $OC$  in the figure). The basis  $\frac{\partial}{\partial x}$  and  $\mathrm{d}x$  lie on the same  $x$ -axis, while  $\frac{\partial}{\partial y}$  and  $\mathrm{d}y$  lie on the same  $y$ -axis. We require that the “product” (pairing) of two bases on the same axis equals 1, and the “product” of bases on different axes equals 0. This is precisely defined by their pairing (see equation (4.3.1)):

$$\begin{aligned} \left\langle \frac{\partial}{\partial x}, (\mathrm{d}x)_o \right\rangle &= \frac{\partial}{\partial x}(x) = 1, & \left\langle \frac{\partial}{\partial y}, (\mathrm{d}y)_o \right\rangle &= \frac{\partial}{\partial y}(y) = 1, \\ \left\langle \frac{\partial}{\partial x}, (\mathrm{d}y)_o \right\rangle &= \frac{\partial}{\partial x}(y) = 0, & \left\langle \frac{\partial}{\partial y}, (\mathrm{d}x)_o \right\rangle &= \frac{\partial}{\partial y}(x) = 0. \end{aligned}$$



**Figure 4.3.1** Relationship between a tangent vector and a cotangent vector

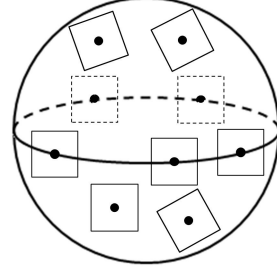
### 3. Tangent Bundle and Cotangent Bundle

Let  $M$  be an  $m$ -dimensional manifold. The tangent space and cotangent space at point  $p$  are denoted by  $T_p M$  and  $T_p^* M$ , respectively. The union of the tangent spaces at all points of  $M$  is called the **tangent bundle** of  $M$  (see Figure 4.3.2), denoted by

$$TM = \bigcup_{p \in M} T_p M.$$

Similarly, the union of the cotangent spaces at all points of  $M$  is called the **cotangent bundle** of  $M$ , denoted by

$$T^*M = \bigcup_{p \in M} T_p^* M.$$



**Figure 4.3.2** The union of tangent spaces forms the tangent bundle

## §4.4 Transformation of Vectors

Let  $(U; x^i)$  be a local coordinate system at point  $p$  on a smooth manifold  $M$ , called the **old coordinate system**, and let  $(V; y^i)$  be another local coordinate system at the same point  $p$ , called the **new coordinate system**. The coordinate transformation between them is given by

$$y^i = y^i(x^1, \dots, x^m), \quad 1 \leq i \leq m.$$

We discuss how tangent vectors and cotangent vectors transform under such a local coordinate transformation. We first consider the one-dimensional case and then generalize the results to  $m$  dimensions.

When  $m = 1$ , the coordinate transformation is

$$y = y(x). \quad (4.4.1)$$

Differentiating gives

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} \frac{\partial y}{\partial x},$$

hence

$$\frac{\partial}{\partial x} = \frac{\partial y}{\partial x} \frac{\partial}{\partial y}. \quad (4.4.2)$$

Suppose at point  $p \in M$  there is a tangent vector  $X$ , expressed in the old and new local coordinates as

$$X = \xi \frac{\partial}{\partial x} = \tilde{\xi} \frac{\partial}{\partial y}.$$

Substituting equation (4.4.2) yields

$$X = \xi \frac{\partial y}{\partial x} \frac{\partial}{\partial y} = \tilde{\xi} \frac{\partial}{\partial y}.$$

Therefore we obtain

$$\tilde{\xi} = \xi \frac{\partial y}{\partial x}.$$

This is the transformation rule for the components of a tangent vector under a change of local coordinates. This component transformation shows that a tangent vector is a contravariant vector.

Differentiating the coordinate transformation equation (4.4.1) gives

$$dy = \frac{\partial y}{\partial x} dx. \quad (4.4.3)$$

Suppose at point  $p \in M$  there is a cotangent vector  $A$ , expressed in the old and new local coordinates as

$$A = \tilde{a} dy = a dx.$$

Substituting equation (4.4.3) yields

$$A = \tilde{a} \frac{\partial y}{\partial x} dx = a dx .$$

Therefore we obtain

$$a = \tilde{a} \frac{\partial y}{\partial x} .$$

This is the transformation rule for the components of a cotangent vector under a change of local coordinates. This component transformation shows that a cotangent vector is a covariant vector.

Generalizing to the  $m$ -dimensional case. Let a tangent vector  $X$  be expressed in the local coordinate system  $(U; x^i)$  as

$$X = \sum_{i=1}^m \xi^i \frac{\partial}{\partial x^i} ,$$

and in another local coordinate system  $(V; y^i)$  as

$$X = \sum_{i=1}^m \tilde{\xi}^i \frac{\partial}{\partial y^i} .$$

Then the transformation rule for the components of  $X$  with respect to the natural bases is

$$\tilde{\xi}^j = \sum_{i=1}^m \xi^i \frac{\partial y^j}{\partial x^i} . \quad (4.4.4)$$

Let a cotangent vector  $A$  be expressed in the local coordinate system  $(U; x^i)$  as

$$A = \sum_{i=1}^m a_i dx^i ,$$

and in the other local coordinate system  $(V; y^i)$  as

$$A = \sum_{i=1}^m \tilde{a}_i dy^i .$$

Then the transformation rule for the components of  $A$  with respect to the natural bases is

$$a_i = \sum_{j=1}^m \tilde{a}_j \frac{\partial y^j}{\partial x^i} . \quad (4.4.5)$$



## Chapter 5 Mappings and Submanifolds

### §5.1 Mappings

#### 1. Smooth Mappings

**Definition 5.1.1** Let  $f: M \rightarrow N$  be a continuous mapping from an  $m$ -dimensional smooth manifold  $M$  to an  $n$ -dimensional smooth manifold  $N$ . If at a point  $p \in M$ , there exist **compatible** coordinate charts  $(U, \phi_U)$  about  $p$  and  $(V, \psi_V)$  about  $f(p)$  such that the induced mapping

$$\psi_V \circ f \circ \phi_U^{-1} : \phi_U(U) \rightarrow \psi_V(V)$$

is smooth ( $C^\infty$ ) at the point  $\phi_U(p)$ , then  $f$  is said to be **smooth at  $p$** . If  $f$  is smooth at every point of  $M$ , then  $f$  is called a **smooth mapping** from  $M$  to  $N$ .

**Theorem 5.1.1** Let  $M_1, \dots, M_k$  be smooth manifolds. Then each projection map

$$\pi_i : M_1 \times \dots \times M_k \rightarrow M_i$$

is a smooth mapping.

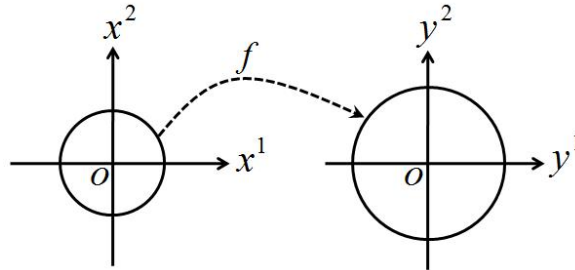
If the smooth manifolds  $M$  and  $N$  have the same dimension,  $f: M \rightarrow N$  is a homeomorphism, and both  $f$  and its inverse  $f^{-1}$  are smooth mappings, then  $f$  is called a **diffeomorphism** (or **smooth diffeomorphism**).

If two smooth manifolds are diffeomorphic, their smooth manifold structures are said to be **isomorphic**.

Suppose the mapping  $f$  is given by  $n$  smooth functions of  $m$  variables  $y^a = f^a(x^1, \dots, x^m)$  ( $a = 1, 2, \dots, n$ ). Then the **Jacobian matrix** of  $f$  is defined as

$$A = \left( \frac{\partial y^a}{\partial x^i} \right) = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \dots & \frac{\partial y^1}{\partial x^m} \\ \vdots & \ddots & \vdots \\ \frac{\partial y^n}{\partial x^1} & \dots & \frac{\partial y^n}{\partial x^m} \end{pmatrix}, \quad i = 1, 2, \dots, m.$$

The **rank** of the mapping  $f$  at point  $(x_0^1, \dots, x_0^m)$  is the rank of the Jacobian matrix  $A$  evaluated at  $(x_0^1, \dots, x_0^m)$ . The rank of a mapping is an important invariant.



**Figure 5.1.1** The mapping  $f$  is a diffeomorphism

**Example 5.1.1** As shown in Figure 5.1.1, consider a circle  $C_x$  in the plane  $x^1x^2$ :  $C_x : (x^1)^2 + (x^2)^2 = 1$ . The mapping

$$f: \begin{cases} y^1 = 2x^1, \\ y^2 = 2x^2 \end{cases} \quad (5.1.1)$$

maps the circle  $C_x$  to the circle  $C_y$  in the plane  $y^1 y^2$ . The equation of the circle  $C_y$  is

$$(y^1)^2 + (y^2)^2 = 4.$$

The Jacobian matrix of  $f$  is

$$A = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Since the determinant of the Jacobian matrix is

$$|A| = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 \neq 0,$$

the inverse transformation  $f^{-1}$  exists:

$$f^{-1}: \begin{cases} x^1 = \frac{1}{2} y^1, \\ x^2 = \frac{1}{2} y^2. \end{cases}$$

Both  $f$  and  $f^{-1}$  are smooth mappings, and  $f$  is a homeomorphism. Therefore,  $f$  is a **diffeomorphism** from  $C_x$  to  $C_y$ .

A smooth mapping between smooth manifolds naturally induces linear mappings between the tangent spaces and between the cotangent spaces at corresponding points, called the tangent map and the cotangent map, respectively. Let us first understand the cotangent map.

## 2. Cotangent Map

**Example 5.1.2** Assume that in Example 5.1.1, the smooth mapping from the circle  $C_x$  to the circle  $C_y$  has the following general form in local coordinates:

$$f: \begin{cases} y^1 = y^1(x^1, x^2), \\ y^2 = y^2(x^1, x^2). \end{cases} \quad (5.1.2)$$

The Jacobian matrix of  $f$  is

$$A = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} \end{pmatrix}.$$

Differentiating equation (5.1.2) gives

$$\begin{cases} dy^1 = \frac{\partial y^1}{\partial x^1} dx^1 + \frac{\partial y^1}{\partial x^2} dx^2, \\ dy^2 = \frac{\partial y^2}{\partial x^1} dx^1 + \frac{\partial y^2}{\partial x^2} dx^2. \end{cases} \quad (5.1.3)$$

In matrix form, this is

$$\begin{pmatrix} dy^1 \\ dy^2 \end{pmatrix} = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} \end{pmatrix} \begin{pmatrix} dx^1 \\ dx^2 \end{pmatrix}.$$

Here,  $dy^1, dy^2$  form a basis for cotangent vectors on  $C_y$ , and  $dx^1, dx^2$  form a basis for cotangent vectors on  $C_x$ . The coefficient matrix in (5.1.3) is precisely the Jacobian matrix of  $f$ . Equation (5.1.3) can be written as

$$dy^\alpha = \sum_{i=1}^2 \frac{\partial y^\alpha}{\partial x^i} dx^i, \quad \alpha = 1, 2.$$

Equation (5.1.3) expresses the basis  $dy^1, dy^2$  in terms of the basis  $dx^1, dx^2$ . Hence, if a cotangent vector on  $C_y$  is given by

$$d\theta = 2dy^1 + 3dy^2,$$

substituting (5.1.3) yields

$$\begin{aligned} d\theta &= 2 \frac{\partial y^1}{\partial x^1} dx^1 + 2 \frac{\partial y^1}{\partial x^2} dx^2 + 3 \frac{\partial y^2}{\partial x^1} dx^1 + 3 \frac{\partial y^2}{\partial x^2} dx^2 \\ &= \left( 2 \frac{\partial y^1}{\partial x^1} + 3 \frac{\partial y^2}{\partial x^1} \right) dx^1 + \left( 2 \frac{\partial y^1}{\partial x^2} + 3 \frac{\partial y^2}{\partial x^2} \right) dx^2. \end{aligned}$$

Through this transformation, the cotangent vector  $d\theta$  on  $C_y$  becomes a cotangent vector on  $C_x$ .

Based on this example, we define the following mapping:

**Definition 5.1.2** Let  $f: M \rightarrow N$  be a smooth mapping, where  $M$  and  $N$  are  $m$ - and  $n$ -dimensional smooth manifolds, respectively, and  $p \in M$ ,  $q = f(p)$ . The mapping  $f$  induces a mapping  $f^*: T_q^* N \rightarrow T_p^* M$ , such that for a cotangent vector  $dg$  on  $N$ , its image on the manifold  $M$  is the cotangent vector

$$f^*(dg) = d(g \circ f), \quad dg \in T_q^*.$$

The mapping  $f^*$  is called the **differential** of  $f$  at  $p$ , or the **cotangent map** of  $f$  at  $p$ .

Note that the direction of  $f^*$  is opposite to that of  $f$ ; therefore,  $f^*$  is also called the **pullback mapping** of  $f$  at  $p$ .

In the definition,  $g \circ f$  is the composite function of  $g$  and  $f$ , and  $d(g \circ f)$  is its total differential at  $p$ , analogous to equation (5.1.3).

Suppose the local coordinate system around point  $p$  is  $(U; x^i)$  and around point  $q$  is  $(V; y^i)$ . Then the mapping  $f$  can be expressed in local coordinates  $(U; x^i)$  as

$$y^\alpha = f^\alpha(x^1, \dots, x^m), \quad 1 \leq \alpha \leq n.$$

The action of the cotangent map  $f^*$  on the natural basis  $\{dy^\alpha, 1 \leq \alpha \leq n\}$  is given by

$$f^*(dy^\alpha) = d(y^\alpha \circ f) = \sum_{i=1}^m \left( \frac{\partial f^\alpha}{\partial x^i} \right)_p dx^i.$$

The coefficient matrix  $\left( \frac{\partial f^\alpha}{\partial x^i} \right)_p$  is precisely the **Jacobian matrix** of the mapping  $f$  at point  $p$ .

### 3. Tangent Map

**Example 5.1.3** Assume  $g(y^1, y^2)$  is a smooth function on the circle  $C_y$ . Using the mapping  $f$  (i.e., using equation (5.1.2)), we obtain

$$g(x^1, x^2) = g(y^1(x^1, x^2), y^2(x^1, x^2)), \quad (5.1.4)$$

so the function  $g$  becomes a function on the circle  $C_x$ . Taking partial derivatives of  $g(x^1, x^2)$  gives

$$\begin{cases} \frac{\partial g}{\partial x^1} = \frac{\partial g}{\partial y^1} \frac{\partial y^1}{\partial x^1} + \frac{\partial g}{\partial y^2} \frac{\partial y^2}{\partial x^1}, \\ \frac{\partial g}{\partial x^2} = \frac{\partial g}{\partial y^1} \frac{\partial y^1}{\partial x^2} + \frac{\partial g}{\partial y^2} \frac{\partial y^2}{\partial x^2}, \end{cases} \quad (5.1.5)$$

thus we have

$$\begin{cases} \frac{\partial}{\partial x^1} = \frac{\partial y^1}{\partial x^1} \frac{\partial}{\partial y^1} + \frac{\partial y^2}{\partial x^1} \frac{\partial}{\partial y^2}, \\ \frac{\partial}{\partial x^2} = \frac{\partial y^1}{\partial x^2} \frac{\partial}{\partial y^1} + \frac{\partial y^2}{\partial x^2} \frac{\partial}{\partial y^2}. \end{cases} \quad (5.1.6)$$

In matrix form,

$$\begin{pmatrix} \frac{\partial}{\partial x^1} \\ \frac{\partial}{\partial x^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^2}{\partial x^1} \\ \frac{\partial y^1}{\partial x^2} & \frac{\partial y^2}{\partial x^2} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial y^1} \\ \frac{\partial}{\partial y^2} \end{pmatrix}.$$

Here,  $\left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}\right)$  form the natural basis of tangent vectors on circle  $C_x$ , and  $\left(\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}\right)$  form the natural basis of tangent vectors on circle  $C_y$ . Equation (5.1.6) expresses the basis  $\left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}\right)$  in terms of the basis  $\left(\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}\right)$ . That is, given the tangent-vector basis  $\left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}\right)$  on circle  $C_x$ , what tangent vectors on circle  $C_y$  correspond to them under the smooth mapping  $f$ ? The answer is the linear combination of  $\frac{\partial}{\partial y^1}$  and  $\frac{\partial}{\partial y^2}$  shown on the right-hand side of (5.1.6). The coefficient matrix in (5.1.6) is the **transpose** of the Jacobian matrix  $A$  of  $f$ . Equation (5.1.6) can be written as

$$\frac{\partial}{\partial x^i} = \sum_{\beta=1}^2 \frac{\partial y^\beta}{\partial x^i} \frac{\partial}{\partial y^\beta}, \quad i=1, 2.$$

Inspired by this example, we define a mapping:

**Definition 5.1.3** Let  $f: M \rightarrow N$  be a smooth mapping, where  $M$  and  $N$  are  $m$ - and  $n$ -dimensional smooth manifolds, respectively, and  $p \in M$ . For any tangent vector  $X \in T_p M$ , the mapping  $f$  induces a mapping  $f_*: T_p M \rightarrow T_{f(p)} N$ , such that the image of  $X$  on the manifold  $N$  is the tangent vector  $f_*(X) \in T_{f(p)} N$  satisfying

$$f_*(X)(g) = X(g \circ f), \quad \forall g \in C_{F(p)}^\infty.$$

The mapping  $f_*$  is called the **tangent map**(or **differential**) of the smooth mapping  $f$  at point  $p$ .

The tangent map  $f_*$  has the **same direction** as  $f$ ; therefore, it is also called the **pushforward mapping** of  $f$  at  $p$ .

In the definition,  $g \circ f$  is the composite function of  $g$  and  $f$ , analogous to equation (5.1.4), and  $X(g \circ f)$  is the derivative of  $g \circ f$  along  $X$ , analogous to equation (5.1.5).

Suppose the local coordinate system around point  $p$  is  $(U; x^i)$  and around point  $q$  is  $(V; y^i)$ . Then the mapping  $f$  can be expressed in local coordinates  $(U; x^i)$  as

$$y^\alpha = f^\alpha(x^1, \dots, x^m), \quad 1 \leq \alpha \leq n.$$

The action of the tangent map  $f_*$  on the natural basis  $\left\{\frac{\partial}{\partial x^i}, 1 \leq i \leq m\right\}$  is given by

$$f_*\left(\frac{\partial}{\partial x^i}\right) = \sum_{\beta=1}^n \left(\frac{\partial f^\beta}{\partial x^i}\right)_p \frac{\partial}{\partial y^\beta}. \quad (5.1.7)$$

Thus, the matrix representing the tangent map  $f_*$  with respect to the natural bases

$\left\{\frac{\partial}{\partial x^i}, 1 \leq i \leq m\right\}$  and  $\left\{\frac{\partial}{\partial y^\beta}, 1 \leq \beta \leq n\right\}$  is precisely the **transpose** of the Jacobian matrix of  $f$  at point  $p$ .

**Example 5.1.4** Let  $X = -2\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y}$  is a tangent vector on  $R^2$ , and the map is

$$f: R^2 \rightarrow R^2, (x, y) \mapsto (u, v) = (2x + y, 3y).$$

We shall find the image of the tangent vector  $X$  under the tangent map  $f_*$ .

The Jacobian matrix of the map  $f$  is

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix},$$

and its transpose is

$$A^T = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}.$$

From equation (5.1.7) we obtain

$$\begin{pmatrix} f_* \frac{\partial}{\partial x} \\ f_* \frac{\partial}{\partial y} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix} = \begin{pmatrix} 2\frac{\partial}{\partial u} \\ \frac{\partial}{\partial u} + 3\frac{\partial}{\partial v} \end{pmatrix}.$$

The images of the natural basis vectors  $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$  under the tangent map  $f_*$  are as given above.

Consequently, the image of the tangent vector  $X$  under  $f_*$  is

$$\begin{aligned} Y = f_* X &= f_* \left( -2\frac{\partial}{\partial x} + 3\frac{\partial}{\partial y} \right) \\ &= -2f_* \left( \frac{\partial}{\partial x} \right) + 3f_* \left( \frac{\partial}{\partial y} \right) \\ &= -2 \left( 2\frac{\partial}{\partial u} \right) + 3 \left( \frac{\partial}{\partial u} + 3\frac{\partial}{\partial v} \right) \\ &= -\frac{\partial}{\partial u} + 9\frac{\partial}{\partial v}. \end{aligned}$$

**Theorem 5.1.2 (Inverse Function Theorem in  $R^m$ )** Let  $f$  be a smooth mapping defined on an open subset  $V \subset R^m$ , with  $p \in V$  and  $f: V \rightarrow R^m$ . If at point  $p \in V$  the Jacobian determinant of  $f$  satisfies

$$\det \left( \frac{\partial f^\alpha}{\partial x^i} \right) \bigg|_p \neq 0,$$

then there exists a neighborhood  $U \subset V$  of  $p$  such that  $f(U) = W$  is a neighborhood of  $f(p)$  and  $f$  has a smooth inverse

$$f^{-1}: W \rightarrow U.$$

**Theorem 5.1.3 (Inverse Function Theorem for Manifolds)** Let  $M$  and  $N$  be smooth manifolds of the same dimension  $m$ , and let  $f: M \rightarrow N$  be a smooth mapping. If at a point  $p \in M$  the tangent map  $f_*: T_p M \rightarrow T_{f(p)} N$  is an isomorphism, then there exists a neighborhood  $U$  of  $p$  in  $M$  such that  $f(U) = V$  is a neighborhood of  $f(p)$  in  $N$ , and the restriction  $f|_U: U \rightarrow V$  is a diffeomorphism.

Because the smooth manifolds  $M$  and  $N$  have the same dimension, the tangent map  $f_*: T_p M \rightarrow T_{f(p)} N$  being an isomorphism at point  $p \in M$  implies that the two tangent spaces  $T_p M$  and  $T_{f(p)} N$  are in one-to-one correspondence; hence the tangent map  $f_*$  is injective

(and surjective) at point  $p \in M$ .

If the smooth manifold  $M$  has dimension  $m$  and the smooth manifold  $N$  has dimension  $n$ , and there exists a smooth mapping  $f: M \rightarrow N$ , then when the tangent map  $f_*$  is **injective** at point  $p$ , we say the tangent map is **non-degenerate** at  $p$  (i.e.,  $f_*$  maps every nonzero tangent vector to a nonzero tangent vector). In this case,  $m \leq n$ , the dimension of the tangent space  $T_{f(p)}N$  is  $m$ , and the rank of the Jacobian matrix of the smooth mapping  $f$  at point  $p \in M$  is also  $m$ .

If the tangent map  $f_*$  is injective (non-degenerate) at point  $p$ , then the mapping  $f$  itself is also injective in a neighborhood of  $p$ . Therefore, the properties of the tangent map  $f_*$  at point  $p$  determine the properties of the mapping  $f$  in a neighborhood of  $p$ .

## §5.2 Submanifolds

### 1. Immersed Submanifolds and Embedded Submanifolds

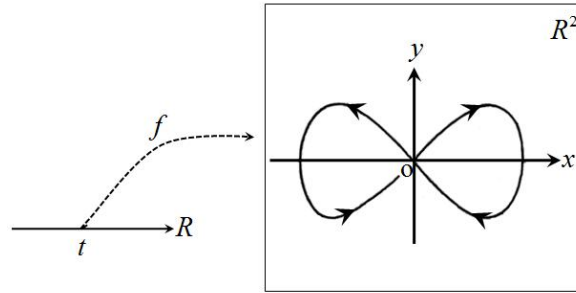
**Definition 5.2.1** Let  $f: M \rightarrow N$  be a smooth mapping from an  $m$ -dimensional smooth manifold  $M$  to an  $n$ -dimensional smooth manifold  $N$ , such that at every point  $p \in M$  the tangent map  $f_*: T_p M \rightarrow T_{f(p)}N$  is **non-degenerate** (i.e., injective). Then  $f$  is called an **immersion**, and  $(f, M)$  is called an **immersed submanifold** of  $N$ . If, in addition,  $f$  is **injective** (one-to-one), then  $(f, M)$  is called an **embedded submanifold** of  $N$ .

A mapping  $f$  being non-degenerate at point  $p$  means that  $f_*$  maps every nonzero tangent vector at  $p$  to a nonzero tangent vector.

An immersed submanifold is locally one-to-one (i.e., locally an embedding), but globally it may not be one-to-one. A practical way to distinguish between an immersed submanifold and an embedded submanifold is to check whether the image set  $f(M)$  has **self-intersections**.

**Example 5.2.1** As shown in Figure 5.2.1, consider the mapping  $f: \mathbb{R} \rightarrow \mathbb{R}^2$ ,

$$f(t) = \left( 2 \cos\left(t - \frac{\pi}{2}\right), \sin 2\left(t - \frac{\pi}{2}\right) \right).$$



**Figure 5.2.1** The mapping  $f$  is an immersion

The tangent vector at any point  $t \in \mathbb{R}$  is  $\frac{\partial}{\partial t}$ . The tangent map of  $f$  is

$$f_* \frac{\partial}{\partial t} = \frac{df}{dt} = \left( -2 \sin\left(t - \frac{\pi}{2}\right), 2 \cos 2\left(t - \frac{\pi}{2}\right) \right).$$

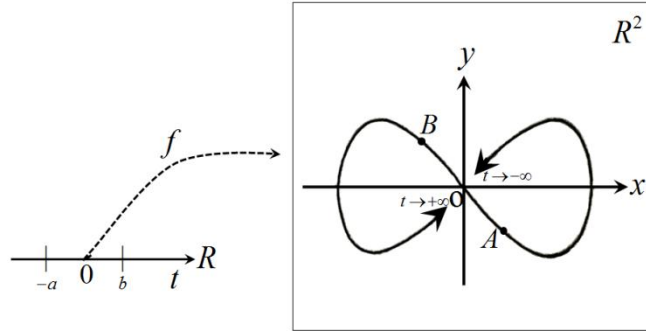
It is clear that  $f_* \frac{\partial}{\partial t}$  is nonzero everywhere, so  $(f, \mathbb{R})$  is an **immersed submanifold** of  $\mathbb{R}^2$ .

If we take a local interval of  $t$ , say  $t \in (0, \frac{\pi}{2})$ , the image  $f(t)$  has no self-intersections on this local interval. However, globally,  $t \in (-\infty, \infty) = \mathbb{R}$ , the image  $f(\mathbb{R})$  does have intersections (e.g.,  $f(\pi) = f(2\pi) = (0, 0)$ ). Therefore,  $f$  is an immersion, and  $(f, \mathbb{R})$  is an **immersed**

**submanifold**, but **not** an embedded submanifold.

**Example 5.2.2** Consider the mapping  $f: \mathbb{R} \rightarrow \mathbb{R}^2$ ,

$$f(t) = \left( 2 \cos(2 \arctg t + \frac{\pi}{2}), \sin 2(2 \arctg t + \frac{\pi}{2}) \right).$$



**Figure 5.2.2** The mapping  $f$  is an embedding

As shown in Figure 5.2.2, this mapping has the property that when  $t = 0$ ,  $f(0) = (0,0)$ , and when  $t \rightarrow \pm\infty$ ,  $f(t) \rightarrow (0,0)$ . Thus, the image  $f(\mathbb{R})$  does **not** intersect itself. Therefore,  $f$  is an **embedding**, and  $(f, \mathbb{R})$  is an **embedded submanifold** of  $\mathbb{R}^2$ .

## 2. Open Submanifold

**Definition 5.2.2** For an open subset  $U$  of a smooth manifold  $N$ , the identity mapping  $f = id: U \rightarrow N$  makes  $(f, U)$  an **embedded submanifold** of  $N$ , called an **open submanifold** of  $N$ .

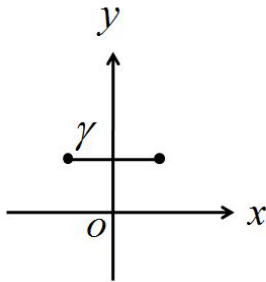
The dimension of the open submanifold  $(f, U)$  is the same as that of the smooth manifold  $N$ . Its smooth manifold structure is obtained by restricting the smooth manifold structure of  $N$  to  $U$ .

## 3. Closed Submanifold

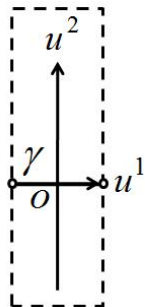
**Definition 5.2.3** Let  $(f, M)$  be an  $m$ -dimensional embedded submanifold of an  $n$ -dimensional smooth manifold  $N$ . If  $f(M)$  is a **closed subset** of  $N$  and for every point  $q \in f(M)$  there exists a local coordinate chart  $(U; u^i)$  of  $N$  around  $q \in f(M)$  such that the intersection  $f(M) \cap U$  is described by the equations

$$u^{m+1} = u^{m+2} = \dots = u^n = 0,$$

then  $(f, M)$  is called a **closed submanifold** of  $N$ .



**Figure 5.2.3**  
Embedded submanifold  $\gamma$



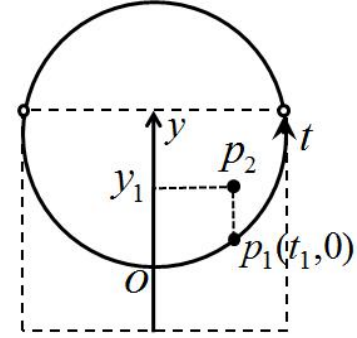
**Figure 5.2.4**  
Coordinate system on the open set  $U$

**Example 5.2.3** In Figure 5.2.3, the plane  $\mathbb{R}^2$  is the smooth manifold  $N$  referred to in the definition, and the line segment  $\gamma$  is  $(f, M)$ , which is an embedded submanifold of  $\mathbb{R}^2$ . The line segment is also a closed subset  $f(M)$ .

As shown in Figure 5.2.4, we set up a new Cartesian coordinate system, i.e., a local coordinate chart  $(U; u^i)$ . The region enclosed by the dashed box is the open set  $U$ . The intersection  $f(M) \cap U$  is precisely that part of the line segment inside the

dashed box (excluding the two endpoints). In this local coordinate system, the equation of  $f(M) \cap U$  is simply  $u^2 = 0$ . Therefore, the line segment  $\gamma$  is a **closed submanifold** of the plane.

**Example 5.2.4** In Figure 5.2.5, the plane  $R^2$  is the smooth manifold  $N$  mentioned in the definition. The circle shown is  $(f, M)$ , an embedded submanifold of  $R^2$ , and the circle  $f(M)$  is also a closed subset. For the circle, instead of a Cartesian coordinate system, we construct a curvilinear coordinate system  $(U; u^i)$  as a local chart. The region enclosed by the dashed box in the figure is the open set  $U$ , where  $u^1 = t$ ,  $u^2 = y$ . Here, the  $u^1$ -coordinate measures arc length along the circle from a reference point, and the  $u^2$ -coordinate measures the signed distance perpendicular to the circle.



**Figure 5.2.5** Circle is a closed submanifold of the plane

Within this local coordinate system, a point  $p_1$  on the circle has coordinates  $(t_1, 0)$ , where  $t_1$  is its arc-length from the origin. A nearby point  $p_2$  off the circle may have coordinates  $(t_1, y_1)$  with  $y_1 \neq 0$ .

The intersection  $f(M) \cap U$  is precisely the arc inside the dashed box (excluding the two endpoints). In the local coordinates  $(U; u^i)$ , the equation of  $f(M) \cap U$  is simply  $u^2 = y = 0$ . Therefore, according to Definition 5.2.3, the entire circle is a **closed submanifold** of the plane.

The circle is a one-dimensional manifold and a closed submanifold of the two-dimensional plane. Generalizing this, the unit sphere  $S^m \subset R^{m+1}$  is closed, and the identity mapping  $f: S^m \rightarrow R^{m+1}$  makes  $(f, S^m)$  a **closed submanifold** of  $R^{m+1}$ .

#### 4. Regular Submanifold

In Figure 5.2.2, the straight line on the left and the curve on the right are homeomorphic. Indeed, for an embedded submanifold  $(f, M)$ , because the smooth mapping  $f$  is injective,  $f$  can transfer the topological and differential structures from  $M$  onto the image set  $f(M)$ . For example, suppose on the left-hand side of Figure 5.2.2 the real line  $R$  has a neighborhood (open set)  $U = (-a, b)$  around the point  $t = 0$ , where  $a > 0$ ,  $b > 0$ . The embedding  $f$  maps this neighborhood  $U$  onto the curve segment  $AB$  near the origin  $(0, 0)$  (corresponding to  $t = 0$ ) on the curve; this image is a neighborhood  $U' = AB$ ,  $f(U) = U'$  in  $f(M)$ . In this way, the image set  $f(M)$  acquires the same topological and differential structure as  $M$ , making the embedding  $f: M \rightarrow f(M)$  a diffeomorphism.

However, the image set  $f(M)$  is also a subset of the smooth manifold  $N$  and therefore naturally inherits the **subspace topology** induced from  $N$ . In general, the topology transferred from  $M$  and the topology induced from  $N$  may **not** coincide. For instance, in Figure 5.2.2, the curve  $f(R)$  as a subset of  $R^2$  has the subspace topology induced from  $R^2$ . A neighborhood of the point  $(0, 0)$  (i.e.,  $f(0)$ ) in this induced topology is the intersection of a neighborhood of  $(0, 0)$  in  $R^2$  with  $f(R)$ ; such an intersection can contain unions of the form  $(-\alpha, \beta) \cup (\gamma, +\infty) \cup (-\infty, -\rho)$  with  $\alpha > 0, \beta > 0, \gamma > 0, \rho > 0$ . This differs from the neighborhood  $V = f(U) = f(-a, b) = U' = AB$  transferred from the line  $R$ .

If the topology transferred from  $M$  and the topology induced from  $N$  on  $(f, M)$  are **the same**, then such a submanifold is called a **regular submanifold**.

**Definition 5.2.4** Let  $(f, M)$  be a submanifold of a smooth manifold  $N$ . If the mapping  $f: M \rightarrow f(M) \subset N$  is a homeomorphism (where  $f(M)$  carries the subspace topology induced from  $N$ ), then  $f$  is called a **regular embedding** of  $M$  into  $N$ , and  $(f, M)$  is called a **regular**



**submanifold** of  $N$ .

In the definition, the condition “ $f: M \rightarrow f(M) \subset N$  is a homeomorphism” means that when  $f(M)$  is given the topology induced from  $N$ , the restricted map  $f: M \rightarrow f(M)$  is a homeomorphism.

A regular submanifold is a special case of an embedded submanifold.

**Theorem 5.2.1** Let  $(f, M)$  be a submanifold of a smooth manifold  $N$ . If  $M$  is **compact**, then  $f: M \rightarrow N$  is a regular embedding.

**Example 5.2.5** According to this theorem, the embedding of the sphere  $S^m$  into  $R^{m+1}$  is a **regular embedding**, and  $S^m$  is a **regular submanifold** of  $R^{m+1}$ , because  $S^m$  is compact.

**Example 5.2.6** A closed submanifold is a regular submanifold.

**Theorem 5.2.2** Let  $N$  be a smooth manifold and  $M$  a submanifold of  $N$ . Then a tangent vector defined on the submanifold  $M$  is also a tangent vector on the manifold  $N$ , and a cotangent vector defined on  $N$  is also a cotangent vector on the submanifold  $M$ .

However, the converse of this theorem does not hold.

Let the  $m$ -dimensional smooth submanifold  $M$  be a submanifold of the  $n$ -dimensional smooth manifold  $N$ , with  $m \leq n$ . Let the smooth mapping  $i: M \rightarrow N$  be the **inclusion map** and  $p \in M$ ,  $p = i(p) \in N$ . Then  $i$  is an embedding. The map  $i$  induces a tangent map  $i_*: T_p M \rightarrow T_{i(p)} N$ , so that a tangent vector  $X \in T_p M$  on  $M$  is mapped to a tangent vector  $i_*(X) \in T_{i(p)} N$  on  $N$ . Hence, a tangent vector defined on the submanifold  $M$  is also a tangent vector on the manifold  $N$ , but the converse is not true.

Similarly,  $i$  induces a cotangent map  $i^*: T_{i(p)}^* N \rightarrow T_p^* M$ , so that a cotangent vector  $\omega \in T_{i(p)}^* N$  on  $N$  is pulled back to a cotangent vector  $i^*(\omega) \in T_p^* M$  on  $M$ . Therefore, a cotangent vector defined on  $N$  is also a cotangent vector on the submanifold  $M$ , but the converse is not true.

## §5.3 Tangent Space of a Product Manifold

### 1. Direct Sum of Linear Spaces

**Definition 5.3.1** Let  $V_1, V_2$  be subspaces of a linear space  $V$ . Their **sum**  $V_1 + V_2$  is the subset consisting of all vectors that can be expressed as  $a_1 + a_2$  with  $a_1 \in V_1$ ,  $a_2 \in V_2$ :

$$V = V_1 + V_2 = \{(a_1 + a_2) | a_1 \in V_1, a_2 \in V_2\}.$$

$V_1$  and  $V_2$  are called **subspaces** of the linear space  $V$ .

**Definition 5.3.2** Let  $V_1, V_2$  be subspaces of a linear space  $V$ . If every vector  $a \in V_1 + V_2$  has a **unique** decomposition

$$a = a_1 + a_2, \quad a_1 \in V_1, \quad a_2 \in V_2,$$

then the sum is called a **direct sum**, denoted by  $V = V_1 \oplus V_2$ .

**Theorem 5.3.1** The sum  $V_1 + V_2$  is a direct sum if and only if the equation

$$a_1 + a_2 = 0, \quad a_1 \in V_1, \quad a_2 \in V_2$$

holds only when  $a_1 = 0$  and  $a_2 = 0$ .

**Theorem 5.3.2** The sum  $V_1 + V_2$  is a direct sum if and only if

$$V_1 \cap V_2 = \{0\}.$$

**Theorem 5.3.3** Let  $V_1, V_2$  be subspaces of  $V$ , and set  $W = V_1 + V_2$ . Then

$$W = V_1 \oplus V_2$$

if and only if

$$\dim(W) = \dim(V_1) + \dim(V_2).$$

The concept of a direct sum of subspaces can be extended to the case of finitely many

subspaces.

## 2. Tangent Space of a Product Manifold

Let the tangent space of smooth manifold  $M$  at point  $p$  be  $T_p M$ , and let the tangent space of smooth manifold  $N$  at point  $q$  be  $T_q N$ . Then the tangent space of the product manifold  $M \times N$  at point  $(p, q)$  is  $T_{(p,q)} M \times N$ . The tangent space  $T_{(p,q)} M \times N$  is the **direct sum** of  $T_p M$  and  $T_q N$ , i.e.,

$$T_{(p,q)} M \times N = T_p M \oplus T_q N.$$

## Chapter 6 Smooth Tangent Vector Fields

### §6.1 Tangent Vector Fields and Smooth Tangent Vector Fields

#### 1. Tangent Vector Field

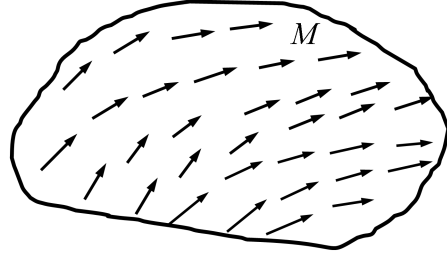
**Definition 6.1.1** If at each point  $p$  of a smooth manifold  $M$ , a tangent vector  $X_p$  is assigned, then  $X$  is called a **tangent vector field** on the manifold  $M$ .

As shown in Figure 6.1.1, a tangent vector field  $X$  on the manifold  $M$  is a collection of selected tangent vectors, one from the tangent space at each point of  $M$ . Since there are infinitely many tangent vectors in the tangent space at each point, there are also infinitely many possible tangent vector fields on  $M$ .

We know that a tangent vector  $X_p$  is a real-valued function  $X_p : C_p^\infty \rightarrow \mathbb{R}$  defined on the set of smooth functions at point  $p$  on  $M$ . Therefore, for a tangent vector field  $X$ , if for every smooth function  $f \in C^\infty(M)$  we define

$$(Xf)_{(p)} = X_p f,$$

then  $Xf$  becomes a real-valued function defined on the manifold  $M$ .



**Figure 6.1.1** Tangent vector field on manifold  $M$

#### 2. Smooth Tangent Vector Field

**Definition 6.1.2** If a tangent vector field  $X$  on a smooth manifold  $M$  satisfies that for every  $f \in C^\infty(M)$ , the function  $Xf \in C^\infty(M)$ , then  $X$  is called a **smooth tangent vector field** on  $M$ .

According to this definition, a smooth tangent vector field  $X$  is an **operator** from  $C^\infty(M)$  to  $C^\infty(M)$ . Since a smooth tangent vector field  $X$  assigns a tangent vector at each point of  $M$ , it inherits the properties specified in the definition of a tangent vector. Let  $f, g \in C^\infty(M)$  and  $\alpha, \beta \in \mathbb{R}$ ; then

- 1)  $X(\alpha f + \beta g) = \alpha(Xf) + \beta(Xg)$ ;
- 2)  $X(fg) = f(Xg) + g(Xf)$ .

The set of all smooth tangent vector fields on  $M$  is denoted by  $\chi(M)$ .

We define addition and scalar multiplication in  $\chi(M)$  (adding two smooth vector fields, and multiplying a vector field by a real number), making  $\chi(M)$  a **vector space**. Let  $X, Y \in \chi(M)$  and  $\alpha \in \mathbb{R}$ . At any point  $p \in M$ , define

$$\begin{aligned}(X + Y)(p) &= X(p) + Y(p), \\ (aX)(p) &= aX(p),\end{aligned}$$

then  $X + Y$  and  $aX$  are also smooth tangent vector fields on  $M$ .

A smooth tangent vector field can also be multiplied by a smooth function. Let  $X \in \chi(M)$  and  $f \in C^\infty(M)$ . At any point  $p \in M$ , define

$$(fX)(p) = fX(p),$$

then  $fX$  is also a smooth tangent vector field.

**Theorem 6.1.1** The set  $\chi(M)$  of all smooth tangent vector fields on a smooth manifold  $M$  forms a **vector space over the real numbers**.

**Theorem 6.1.2** Let  $X$  be a tangent vector field on a smooth manifold  $M$ .  $X$  is a smooth

tangent vector field **if and only if** for every point  $p \in M$ , there exists a local coordinate chart  $(U; x^i)$  about  $p$  such that the restriction of  $X$  to  $U$  can be expressed as

$$X|_U = \sum_{i=1}^m X^i \frac{\partial}{\partial x^i}, \quad (6.1.1)$$

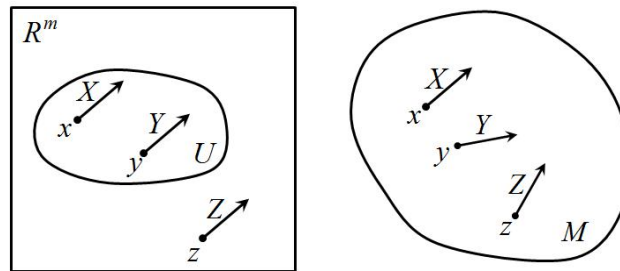
and the components  $X^i (1 \leq i \leq m)$  are **smooth functions** on  $U$ .

### 3.Parallelizability

**Definition 6.1.3** An  $m$ -dimensional smooth manifold  $M$  is called **parallelizable** if its tangent bundle  $TM$  can be expressed as the direct product of  $M$  and an  $m$ -dimensional vector space  $R^m$ , i.e.,

$$TM = M \times R^m.$$

**Example 6.1.1** If a manifold is parallelizable, one can define tangent vectors that are everywhere equal in magnitude and parallel in direction. The  $m$ -dimensional Euclidean space  $R^m$  and any open subset  $U \subset R^m$  are parallelizable. However, most smooth manifolds are **not** parallelizable. As shown in Figure 6.1.2, tangent vectors  $X$  and  $Y$  inside an open set  $U \subset R^m$  are parallel to each other and also parallel to the tangent vector  $Z$  outside  $U$ ; their magnitudes can also be chosen to be equal. In contrast, on a general manifold  $M$ , the tangent vectors are not necessarily parallel, and their magnitudes are not necessarily equal. The tangent vectors  $X, Y, Z$  in the manifold  $M$  shown in the figure are pairwise non-parallel.



**Figure 6.1.2** Parallelizable and non-parallelizable manifolds

**Example 6.1.2** 1) The two-dimensional cylinder is parallelizable, whereas the Möbius strip is not parallelizable.

2) The  $m$ -dimensional torus  $T^m$  is **parallelizable**.

3) Among spheres, only three are parallelizable:  $S^1$ ,  $S^3$ , and  $S^7$  are parallelizable, while  $S^m$  for other  $n$  are not.

4) Lie groups, which will be introduced later, are parallelizable.

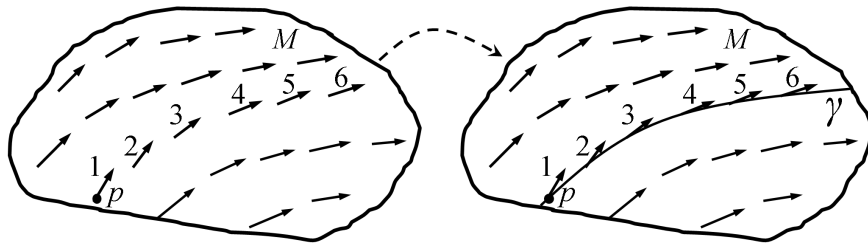
The concept that a manifold  $M$  is parallelizable has an alternative equivalent definition:

**Definition 6.1.4** An  $m$ -dimensional smooth manifold  $M$  is called **parallelizable** if there exist  $m$  smooth tangent vector fields  $\{X_1, X_2, \dots, X_m\}$  on  $M$  such that at every point  $p \in M$ , the tangent vectors  $\{X_1(p), X_2(p), \dots, X_m(p)\}$  form a **basis** of the tangent space  $T_p M$ . Such a set  $\{X_1, X_2, \dots, X_m\}$  is called a **global frame** for the manifold  $M$  (or for the tangent bundle  $TM$ ).

## §6.2 Integral Curves

We know that at every point of a curve on a smooth manifold  $M$ , there is a tangent vector. Conversely, suppose a smooth tangent vector field  $X$  is given on  $M$ . Then  $X$  assigns a tangent vector at each point of  $M$ , as shown in the left-hand side of Figure 6.2.1. We ask: Can we start from a given point  $p$  on the manifold and construct a curve such that the tangent vector at any point of the curve equals the tangent vector assigned by  $X$  at that point, as illustrated in the

right-hand side of Figure 6.2.1?



**Figure 6.2.1** Relation between a tangent vector field and a curve on manifold  $M$

The answer depends on two situations:

- 1) If at point  $p \in M$  we have  $X_p = 0$  (the zero vector), then it is **not possible** to construct such a curve starting from  $p$ .
- 2) If at point  $p \in M$  we have  $X_p \neq 0$ , then we **can** construct such a curve.

Now consider the case where  $X_p \neq 0$ .

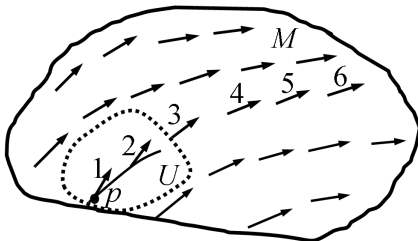
**Definition 6.2.1** Let  $M$  be a smooth manifold and  $X$  a smooth tangent vector field on  $M$ . A smooth curve  $\gamma: (a, b) \rightarrow M$  is called an **integral curve** of  $X$  if at every point  $\gamma(t)$ , the tangent vector  $\dot{\gamma}(t)$  equals  $X_{\gamma(t)}$ .

Suppose a local coordinate chart  $(U; x^i)$  is established around point  $p \in M$ , and the tangent vector field  $X$  is expressed as in equation (6.1.1). Then the system of equations

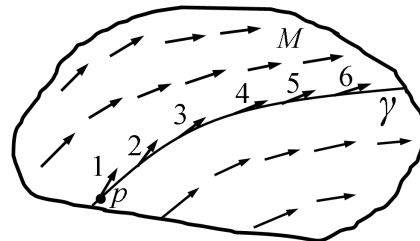
$$\frac{dx^i}{dt} = X^i, \quad 1 \leq i \leq m \quad (6.2.1)$$

is a system of first-order ordinary differential equations. Its solution exists and is unique under usual smoothness conditions, so solving this system yields an integral curve of  $X$ .

The integral curve obtained in this way exists **locally** within the open set  $U$ , as shown in Figure 6.2.2. For the integral curve to extend over the entire manifold  $M$ , additional conditions are required; one sufficient condition is that  $M$  be **compact**, as illustrated in Figure 6.2.3.



**Figure 6.2.2** Local integral curve



**Figure 6.2.3** Global integral curve

**Example 6.2.1** Consider a smooth tangent vector field on  $R^2$  given by  $X = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}$ , where  $(x, y) \neq (0, 0)$  (the origin is a singular point of  $X$ ). From equation (6.2.1) we obtain the system

$$\frac{dx}{dt} = -y, \quad \frac{dy}{dt} = x. \quad (6.2.2)$$

Differentiating the first equation gives  $\frac{d^2x}{dt^2} = -\frac{dy}{dt}$ . Substituting into the second equation yields

$$\frac{d^2x}{dt^2} + x = 0.$$

Solving the second-order equation  $\frac{d^2x}{dt^2} + x = 0$  gives

$$x = C_1 \cos t + C_2 \sin t, \quad (6.2.3)$$

and from the first equation,

$$y = C_1 \sin t - C_2 \cos t. \quad (6.2.4)$$

Equations (6.2.3) and (6.2.4) constitute the general solution of the system (6.2.2).

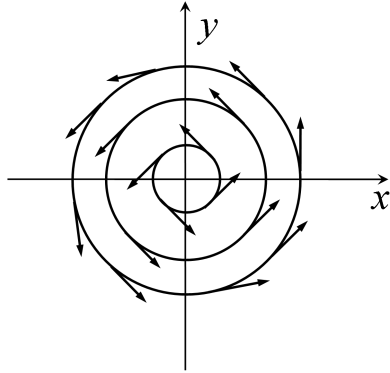
From (6.2.3) and (6.2.4) we obtain

$$\begin{aligned} x^2 + y^2 &= C_1^2 \cos^2 t + C_2^2 \sin^2 t + 2C_1 C_2 \cos t \sin t \\ &\quad + C_1^2 \sin^2 t + C_2^2 \cos^2 t - 2C_1 C_2 \sin t \cos t \\ &= C_1^2 + C_2^2 = C^2. \end{aligned}$$

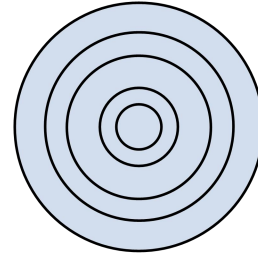
Therefore, the integral curves of the tangent vector field  $X$  are concentric circles centered at the origin in the plane  $\mathbb{R}^2$ , as shown in Figure 6.2.4.

As can be seen from Figure 6.2.4, integral curves starting from different points (except the origin) do not intersect, because the solution of the system (6.2.2) is unique.

The constants  $C_1$  and  $C_2$  in the solution are determined by the coordinates of the initial point  $p$ . Different points yield different values of  $C_1$  and  $C_2$ , so through each point  $p$  there passes exactly one integral curve. Thus, the integral curves fill the entire plane  $\mathbb{R}^2$ . If we consider a disk region in the plane, it can be viewed as being composed of infinitely many concentric circles, as illustrated in Figure 6.2.5. This collection of infinitely many concentric circles is called a **foliation** of the disk.



**Figure 6.2.4** Integral curves of the tangent vector field  $X$



**Figure 6.2.5** A disk composed of infinitely many concentric circles

**Theorem 6.2.1** Every point of a smooth manifold  $M$  lies on exactly one integral curve of a given smooth tangent vector field.

## §6.3 Poisson Bracket

### 1. Poisson Bracket

Let  $X$  and  $Y$  be two smooth tangent vector fields on a manifold  $M$ . Their **Poisson bracket**(or **Lie bracket**) is defined as

$$[X, Y] = XY - YX.$$

One can regard  $[X, Y]$  as an operator acting on  $C^\infty(M)$ , because for any  $f \in C^\infty(M)$ ,

$$[X, Y]f = X(Yf) - Y(Xf).$$

Further verification shows that  $[X, Y]$  satisfies the properties of a smooth tangent vector field:

for any  $f, g \in C^\infty(M)$  and  $\alpha, \beta \in \mathbb{R}$ ,

$$1) [X, Y](\alpha f + \beta g) = \alpha [X, Y]f + \beta [X, Y]g,$$

$$2) [X, Y](fg) = f[X, Y]g + g[X, Y]f.$$

Therefore,  $[X, Y]$  is also a smooth tangent vector field on  $M$ . This means the set  $\chi(M)$  is

**closed** under the bracket operation  $[\cdot, \cdot]$ .

It is interesting to note that the ordinary product  $XY$  does **not** define a smooth tangent vector field on  $M$ .

## 2. Algebraic Properties of the Poisson Bracket

**Theorem 6.3.1** Let  $X, Y, Z$  be smooth tangent vector fields on a manifold  $M$ , and let  $f, g \in C^\infty(M)$ ,  $\alpha, \beta \in \mathbb{R}$ . Then

- 1) Bilinearity (distributivity)  $[\alpha X + \beta Y, Z] = \alpha[X, Z] + \beta[Y, Z]$ ,
- 2) Anticommutativity  $[X, Y] = -[Y, X]$ ,
- 3) Jacobi identity  $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ ,
- 4)  $[fX, gY] = f(Xg)Y - g(Yf)X + fg[X, Y]$ .

Comparing properties 1) and 4), we observe that the Poisson bracket is **linear** with respect to real numbers  $\alpha, \beta \in \mathbb{R}$ , but not with respect to smooth functions  $f, g \in C^\infty(M)$ .

A real vector space equipped with a multiplication operation satisfying bilinearity, anticommutativity, and the Jacobi identity is called a **Lie algebra**. Since the set  $\chi(M)$  of all smooth tangent vector fields on  $M$  satisfies these three properties under the Poisson bracket,  $\chi(M)$  forms a Lie algebra—indeed, an infinite-dimensional Lie algebra.

## 3. Geometric Interpretation of the Poisson Bracket

Suppose the manifold  $M$  is an **analytic manifold**  $C^\omega$ , and assume the coordinates  $x^i(s)$  of points on an integral curve of a tangent vector field  $X = \frac{d}{ds}$  are analytic functions of the parameter  $s$ , i.e., they can be expanded into Taylor series. Expanding  $x^i(s)$  in a neighborhood  $-\varepsilon < s_0 < \varepsilon$  of the parameter value  $s = s_0$ , we obtain

$$\begin{aligned} x^i(s_0 + \varepsilon) &= x^i(s_0) + \varepsilon \left( \frac{dx^i}{ds} \right)_{s_0} + \frac{1}{2!} \varepsilon^2 \left( \frac{d^2 x^i}{ds^2} \right)_{s_0} + \cdots \\ &= \left( 1 + \varepsilon \frac{d}{ds} + \frac{1}{2} \varepsilon^2 \frac{d^2}{ds^2} + \cdots \right) x^i \Big|_{s_0} \\ &= \exp \left( \varepsilon \frac{d}{ds} \right) x^i \Big|_{s_0}. \end{aligned} \quad (6.3.1)$$

From this expansion we see that  $\exp \left( \varepsilon \frac{d}{ds} \right)$  is a **differential operator**; applying it to  $x^i(s)$  and evaluating at  $s_0$  yields the coordinate  $x^i$  of a neighboring point  $s_0 + \varepsilon$ . The operator  $\exp \left( \varepsilon \frac{d}{ds} \right)$  is called the **exponential operator**.

If  $f$  is a multivariate function,  $x^i = x^i(s^1, \dots, s^m)$ , its Taylor series expansion is

$$x^i = x^i(s^1, \dots, s^m) = \exp \left( \sum_{i=1}^m \varepsilon^i \frac{\partial}{\partial s^i} \right) x^i \Big|_{s_0}. \quad (6.3.2)$$

The exponential operator possesses the following important property:

**Theorem 6.3.2** For all  $a, b \in \mathbb{R}$ ,

$$\exp \left[ a \frac{d}{ds} + b \frac{d}{dt} \right] = \exp \left[ a \frac{d}{ds} \right] \exp \left[ b \frac{d}{dt} \right]$$

holds **if and only if**  $\left[ \frac{d}{ds}, \frac{d}{dt} \right] = 0$ .

Using Taylor series, we can provide a geometric interpretation of the Poisson bracket.

Suppose there are two arbitrary vector fields,  $X = \frac{d}{ds}$  and  $Y = \frac{d}{dt}$ , on the manifold  $M$ , with their integral curves as shown in Figure 6.3.1. In a local coordinate system  $(U; x^i)$  around point  $p$ , starting from  $p$  and moving along the integral curve of  $X$  through  $p$  by a parameter increment  $\Delta s = \sqrt{\varepsilon}$ , we reach point  $p_1$ . The coordinates of point  $p_1$  are

$$x^i(p_1) = \exp\left(\sqrt{\varepsilon} \frac{d}{ds}\right) x^i \Big|_p.$$

Next, starting from point  $p_1$  and moving along the integral curve of  $Y$  through  $p_1$  by a parameter increment  $\Delta t = \sqrt{\varepsilon}$ , we reach point  $p_2$ . The coordinates of point  $p_2$  are

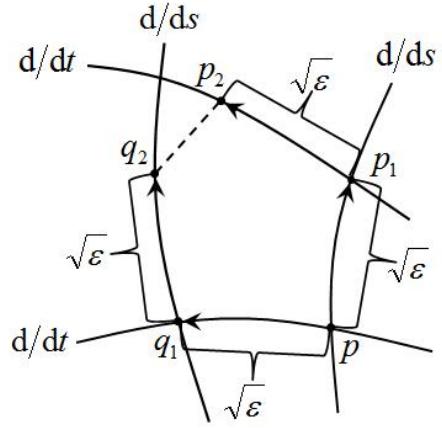
$$x^i(p_2) = \exp\left(\sqrt{\varepsilon} \frac{d}{dt}\right) x^i \Big|_{p_1} = \exp\left(\sqrt{\varepsilon} \frac{d}{dt}\right) \exp\left(\sqrt{\varepsilon} \frac{d}{ds}\right) x^i \Big|_p.$$

Similarly, starting from point  $p$  and moving along the integral curve of  $Y$  through  $p$  by a parameter increment  $\Delta t = \sqrt{\varepsilon}$ , we reach point  $q_1$ . The coordinates of point  $q_1$  are

$$x^i(q_1) = \exp\left(\sqrt{\varepsilon} \frac{d}{dt}\right) x^i \Big|_p,$$

Then, starting from point  $q_1$  and moving along the integral curve of  $X$  through  $q_1$  by a parameter increment  $\Delta s = \sqrt{\varepsilon}$ , we reach point  $q_2$ . The coordinates of point  $q_2$  are

$$x^i(q_2) = \exp\left(\sqrt{\varepsilon} \frac{d}{ds}\right) \exp\left(\sqrt{\varepsilon} \frac{d}{dt}\right) x^i \Big|_p.$$



**Figure 6.3.1** Geometric interpretation of the Poisson bracket

The coordinate difference between points  $q_2$  and  $p_2$  is

$$x^i(q_2) - x^i(p_2) = \left[ \exp\left(\sqrt{\varepsilon} \frac{d}{ds}\right), \exp\left(\sqrt{\varepsilon} \frac{d}{dt}\right) \right] x^i \Big|_p.$$

We compute the value of the Poisson bracket in the above expression:

$$\begin{aligned} & \left[ \exp\left(\sqrt{\varepsilon} \frac{d}{ds}\right), \exp\left(\sqrt{\varepsilon} \frac{d}{dt}\right) \right] \\ &= \left[ 1 + \sqrt{\varepsilon} \frac{d}{ds} + \frac{1}{2} (\sqrt{\varepsilon})^2 \frac{d^2}{ds^2} + O((\sqrt{\varepsilon})^3), 1 + \sqrt{\varepsilon} \frac{d}{dt} + \frac{1}{2} (\sqrt{\varepsilon})^2 \frac{d^2}{dt^2} + O((\sqrt{\varepsilon})^3) \right] \\ &= \varepsilon \left[ \frac{d}{ds}, \frac{d}{dt} \right] + O((\sqrt{\varepsilon})^3). \end{aligned}$$

Thus, the coordinate difference between points  $q_2$  and  $p_2$  is

$$x^i(q_2) - x^i(p_2) = \varepsilon \left[ \frac{d}{ds}, \frac{d}{dt} \right] x^i \Big|_p + O((\sqrt{\varepsilon})^3) x^i \Big|_p.$$

Let the parameter difference between points  $q_2$  and  $p_2$  along their respective curves be  $\lambda$ . Taking the limit:

$$\lim_{\varepsilon \rightarrow 0} \frac{x^i(q_2) - x^i(p_2)}{\varepsilon} = \lim_{\lambda \rightarrow 0} \frac{x^i(q_2) - x^i(p_2)}{\lambda} = \left[ \frac{d}{ds}, \frac{d}{dt} \right] x^i \Big|_p.$$



This shows that the Poisson bracket  $\left[ \frac{d}{ds}, \frac{d}{dt} \right] x^i \Big|_p$  precisely corresponds to the tangent vector from point  $p_2$  to point  $q_2$ .

## §6.4 Mapping of Smooth Tangent Vector Fields

Let  $M$  and  $N$  be smooth manifolds of dimensions  $m$  and  $n$  respectively, and let  $f : M \rightarrow N$  be a smooth mapping. As we have seen earlier,  $f$  induces a **tangent map**  $f_*$ , which maps a tangent vector  $X_p$  at point  $p \in M$  to a tangent vector  $f_*(X_p)$  at point  $f(p) \in N$ . A natural question arises: Can the tangent map  $f_*$  transport **all** smooth tangent vector fields from  $M$  to  $N$ ? The answer is **not always**. If  $f$  is a **diffeomorphism** (smooth with a smooth inverse), then  $M$  and  $N$  are essentially the same as smooth manifolds, so every smooth tangent vector field on  $M$  can indeed be transported entirely to  $N$ . However, if  $f$  is not a diffeomorphism, in general  $f_*X$  does **not** define a smooth tangent vector field on  $N$ , for two reasons:

**1)  $f$  may not be injective (one-to-one).**

Suppose there exist distinct points  $p_1, p_2 \in M$  with  $f(p_1) = f(p_2) = q$ . Then  $f_*$  maps the tangent vectors  $X_{p_1}$  and  $X_{p_2}$  tangent vectors at the same point  $q \in N$ . These images  $f_*(X_{p_1})$  and  $f_*(X_{p_2})$  may be different, resulting in **two distinct tangent vectors** at the same point  $q \in N$ , which contradicts the definition of a single-valued vector field.

**2)  $f$  may not be surjective (onto).**

If  $f$  is not surjective, then  $f_*(X)$  is not defined at points of  $N$  that are not in the image  $f(M)$ . Consequently, on  $N$  there would be points with **no assigned tangent vector** from  $f_*(X)$ , again violating the definition of a vector field on the whole manifold.

Thus, unless  $f$  is a diffeomorphism, the pushforward of a smooth tangent vector field via  $f_*$  generally does **not** yield a well-defined smooth tangent vector field on  $N$ .

**Definition 6.4.1** Let  $f : M \rightarrow N$  be a smooth mapping from smooth manifold  $M$  to smooth manifold  $N$ . If  $X \in \chi(M)$ ,  $\tilde{X} \in \chi(N)$ , and at every point  $p \in M$  we have

$$f_*(X_p) = \tilde{X}_{f(p)},$$

then the vector fields  $\tilde{X}$  and  $X$  are said to be  **$f$ -related**.

**Example 6.4.1** Suppose a tangent vector field in the plane is given by  $X = 2x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$ , and let the mapping be

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto (u, v) = (x + y^2, y^2).$$

We aim to find the image of the tangent vector field  $X$  under the action of the tangent map  $f_*$ .

The Jacobian matrix of the mapping  $f$  is

$$A = \begin{pmatrix} 1 & 2y \\ 0 & 2y \end{pmatrix},$$

and its transpose is

$$A^T = \begin{pmatrix} 1 & 0 \\ 2y & 2y \end{pmatrix}.$$

From equation (5.1.7), we obtain

$$\begin{pmatrix} f_* \frac{\partial}{\partial x} \\ f_* \frac{\partial}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2y & 2y \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial u} \\ 2y \frac{\partial}{\partial u} + 2y \frac{\partial}{\partial v} \end{pmatrix}.$$

The image of the tangent vector field  $X$  under the tangent map  $f_*$  is

$$\begin{aligned} Y = f_*X &= f_*\left(2x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right) = 2xf_*\left(\frac{\partial}{\partial x}\right) + yf_*\left(\frac{\partial}{\partial y}\right) \\ &= 2x\frac{\partial}{\partial u} + y\left(2y\frac{\partial}{\partial u} + 2y\frac{\partial}{\partial v}\right) = (2x + 2y^2)\frac{\partial}{\partial u} + 2y^2\frac{\partial}{\partial v} \\ &= 2u\frac{\partial}{\partial u} + 2v\frac{\partial}{\partial v}. \end{aligned}$$

**Theorem 6.4.1** Let  $f: M \rightarrow N$  be a smooth mapping,  $X, Y \in \chi(M)$ , and  $\tilde{X}, \tilde{Y} \in \chi(N)$ . If  $\tilde{X}$  is  $f$ -related to  $X$  and  $\tilde{Y}$  is  $f$ -related to  $Y$ , then the Lie bracket  $[\tilde{X}, \tilde{Y}]$  is  $f$ -related to  $[X, Y]$ .

**Theorem 6.4.2** If  $f: M \rightarrow N$  is a **diffeomorphism** of smooth manifolds, then the tangent map  $f_*$  establishes an **isomorphism** between the Lie algebras of smooth vector fields  $\chi(M)$  and  $\chi(N)$ .

**Theorem 6.4.3 (Restriction of a vector field to a submanifold)** Let  $M$  be a smooth manifold,  $S \subseteq M$  an **immersed submanifold**, and  $f: S \rightarrow M$  the inclusion map. If a vector field  $X \in \chi(M)$  is **tangent** to  $S$  (i.e.,  $X_p \in T_p(S)$  for all  $p \in S$ ), then there exists a **unique** smooth vector field  $X|_S$  such that  $X|_S$  and  $X$  are  $f$ -related.

## §6.5 Coordinate Basis

**Theorem 6.5.1** Suppose there exist  $h$  smooth tangent vector fields  $X_1, \dots, X_h$  on a manifold  $M$  that are **everywhere linearly independent** in a neighborhood  $U$ . Then at each point  $p \in U$ , there exists a local coordinate system  $(W; w^i)$  such that

$$X_\alpha|_w = \frac{\partial}{\partial w^\alpha}, \quad 1 \leq \alpha \leq h$$

if and only if

$$[X_\alpha, X_\beta] = 0, \quad 1 \leq \alpha, \beta \leq h.$$

We know that at each point  $p$  of an  $m$ -dimensional manifold  $M$ , there is a tangent space  $T_p M$  of dimension  $m$ . If for every point  $p \in M$ , there exist  $h$  smooth tangent vector fields  $X_1, \dots, X_h$  in a neighborhood  $U$  of  $p$  that are everywhere linearly independent, then at each point  $q \in U$ , the tangent space  $T_q M$  contains a distinguished subspace  $L^h(q)$ . This subspace is spanned by the tangent vectors  $X_1, \dots, X_h$ —that is, the subspace  $L^h(q)$  is generated by the values of the vector fields  $X_1, \dots, X_h$  at  $q$ . Such subspaces can be defined consistently across the entire manifold  $M$ . We call  $L^h$  a **smooth  $h$ -dimensional tangent subspace field** on  $M$ , or a **smooth  $h$ -dimensional distribution** on  $M$ , and denote it by

$$L^h|_U = \text{Span}\{X_1, \dots, X_h\}.$$

**Definition 6.5.1** Let  $L^h$  be a smooth  $h$ -dimensional distribution on a manifold  $M$ . In any coordinate domain  $U$ ,  $L^h$  is spanned by smooth tangent vector fields  $X_1, \dots, X_h$  that are everywhere linearly independent. If the Lie bracket  $[X_\alpha, X_\beta]$  ( $1 \leq \alpha, \beta \leq h$ ) of any two of these vector fields can also be expressed as a linear combination of the  $X_\gamma$  within  $U$ , i.e.,

$$[X_\alpha, X_\beta] = \sum_{\gamma=1}^h C_{\alpha\beta}^\gamma X_\gamma,$$

where the coefficients  $C_{\alpha\beta}^\gamma$  are smooth functions on  $U$  (i.e.,  $C_{\alpha\beta}^\gamma \in C^\infty(U)$ ), then the distribution

$L^h$  is said to satisfy the **Frobenius condition**.

**Theorem 6.5.2 (Frobenius Theorem)** Let  $L^h$  be a smooth  $h$ -dimensional distribution defined on an open set  $U$  of a manifold  $M$ . Then for any point  $p \in U$ , there exists a local coordinate system  $(W; w^i)$  with  $W \subset U$  such that

$$L^h|_W = \text{Span}\left(\frac{\partial}{\partial w^1}, \dots, \frac{\partial}{\partial w^h}\right)$$

if and only if  $L^h$  satisfies the **Frobenius condition**.

The Frobenius theorem has an important application. If  $h = m$  and the conditions of the Frobenius theorem are satisfied, then the coordinate basis vectors of the local coordinate system  $(W; w^i)$  are

$$\left(\frac{\partial}{\partial w^1}, \dots, \frac{\partial}{\partial w^m}\right),$$

Every point  $q$  in the neighborhood of a point  $p \in W$  can be expressed with coordinates

$$w^i(q) = w^i(a_1, \dots, a_m) = \exp\left(\sum_{j=1}^m a_j \frac{\partial}{\partial w^j}\right) w^i \Big|_p.$$

**Definition 6.5.2** Let  $U$  be an open subset of a smooth manifold  $M$ . If at each point  $p \in U$  an  $h$ -dimensional subspace  $L^h(p)$  of the tangent space  $T_p M$  is assigned, then  $L^h$  is called an  **$h$ -dimensional distribution** on  $U$ . If for each point  $p \in U$  there exists a neighborhood  $V \subset U$  and  $h$  smooth tangent vector fields  $X_1, \dots, X_h$  defined on  $V$  that are everywhere linearly independent, such that at each point  $q \in V$ ,  $L^h(q)$  is spanned by  $X_1(q), \dots, X_h(q)$ , then the distribution  $L^h$  is said to be **smooth**.

**Example 6.5.1** On  $R^3 \setminus \{0\}$ , consider the three smooth tangent vector fields

$$X = -z \frac{\partial}{\partial y} + y \frac{\partial}{\partial z}, \quad Y = z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z}, \quad Z = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}.$$

They satisfy the following commutation relations:

$$[X, Y] = -Z, \quad [Y, Z] = -X, \quad [Z, X] = -Y.$$

Thus, the distribution  $L^h|_V = \text{Span}\{X, Y, Z\}$  satisfies the Frobenius condition.

This distribution  $L^h$  is **two-dimensional**, because

$$xX + yY + zZ = 0,$$

where  $x, y, z$  are not all zero, so among  $X, Y, Z$ , they are linearly dependent; only two of them are independent.

## §6.6 One-Parameter Differentiable Transformation Groups

**Definition 6.6.1** Let  $M$  be an  $m$ -dimensional smooth manifold, and let  $\varphi: R \times M \rightarrow M$  be a smooth mapping. For any point  $p \in M$ , i.e.,  $(t, p) \in R \times M$ , denote

$$\varphi_t(p) = \varphi(t, p).$$

If the mapping  $\varphi$  satisfies the following conditions:

- 1) Identity mapping:  $\varphi_0(p) = p$ ,
- 2) Group property: for  $s, t \in R$ ,  $\varphi_s \circ \varphi_t = \varphi_{s+t}$ ,

then  $\varphi_t$  is called a **one-parameter differentiable transformation group** acting on the smooth manifold  $M$ , or we say that  $R$  **acts smoothly (from the left)** on  $M$ .

If the parameter  $t$  is fixed, then  $\varphi_t: M \rightarrow M$  is a smooth mapping. Because  $\varphi_t \circ \varphi_{-t} = \varphi_0 = id$ , each  $\varphi_t$  is invertible,  $(\varphi_t)^{-1} = \varphi_{-t}$ ; hence every  $\varphi_t$  is a **diffeomorphism**

from  $M$  onto itself.

**Example 6.6.1** Let  $\varphi(t, x, y) = (xe^{\lambda t}, ye^{\mu t})$ , where  $\lambda, \mu$  is a constant. Then  $\varphi_t$  is a one-parameter differentiable transformation group on  $R^2$ . This is because:

First, when  $t = 0$ , we have  $\varphi_0(x, y) = \varphi(0, x, y) = (x, y)$ , maps the point on  $R^2$  to its original position, so  $\varphi_0(x, y)$  is the identity map, satisfying the first condition of the definition.

Second, for any real numbers  $s, t$ ,

$$\varphi_t = \varphi(t, x, y) = (xe^{\lambda t}, ye^{\mu t}),$$

$$\varphi_s \circ \varphi_t = \varphi(s, xe^{\lambda t}, ye^{\mu t}) = (xe^{\lambda t} e^{\lambda s}, ye^{\mu t} e^{\mu s}) = (xe^{\lambda(t+s)}, ye^{\mu(t+s)}) = \varphi(s+t, x, y) = \varphi_{s+t},$$

which satisfies the second condition of the definition.

**Example 6.6.2** In Example 6.6.1,

$$\varphi(t, x, y) = (xe^{\lambda t}, ye^{\mu t}), \quad \varphi(-t, x, y) = (xe^{-\lambda t}, ye^{-\mu t}),$$

$$\begin{aligned} \varphi_t \circ \varphi_{-t} &= \varphi(t, xe^{-\lambda t}, ye^{-\mu t}) = \varphi(xe^{-\lambda t} e^{\lambda t}, ye^{-\mu t} e^{\mu t}) \\ &= \varphi(xe^{(t-t)\lambda}, ye^{(t-t)\mu}) = \varphi(0, x, y) = \varphi_0(x, y), \end{aligned}$$

so  $\varphi_{-t} = (\varphi_t)^{-1}$ .

If we fix a point  $p \in M$  and define

$$\gamma_p(t) = \varphi_t(p) = \varphi(t, p),$$

then  $\gamma_p$  is a parametric curve on the smooth manifold  $M$  passing through point  $p$ , called the **orbit (or trajectory)** of the one-parameter transformation group  $\varphi_t$  through  $p$ .

At point  $p$ , we have  $t = 0$ . If we denote the tangent vector of the orbit  $\gamma_p$  at point  $p$  by  $X_p$ , and write it as  $X_p = \gamma'_p(0)$ , we obtain a tangent vector field  $X$  on the smooth manifold  $M$ , called the **induced tangent vector field** of the one-parameter differentiable transformation group  $\varphi_t$  on  $M$ .

Establish a local coordinate system  $(U; x^i)$  at point  $p$ . Then  $\varphi$  can be expressed as

$$x^i = \varphi^i(t, x_p^1, \dots, x_p^m),$$

where  $(x_p^1, \dots, x_p^m)$  are the coordinates of point  $p$ , and the  $\varphi^i$  are smooth functions of  $t, x_p^1, \dots, x_p^m$ . The expression for the tangent vector field  $X$  is

$$X_p = \sum_{i=1}^m \frac{\partial \varphi^i(t, x_p^1, \dots, x_p^m)}{\partial t} \bigg|_{t=0} \frac{\partial}{\partial x^i} \bigg|_p,$$

The components  $\frac{\partial \varphi^i}{\partial t} \bigg|_{t=0}$  are smooth functions on  $U$ , so  $X$  is smooth.

The integral curve of the tangent vector field  $X$  is precisely the orbit  $\gamma_p$ , because if we take any point  $q = \gamma_p(s)$  on the orbit  $\gamma_p$ , then the tangent vector of the orbit at  $t = s$  (i.e., at point  $q$ ) is  $X_q$ , denoted by  $X_q = \gamma'_q(s)$ .

**Example 6.6.3** The one-parameter transformation group in Example 6.6.1

$$\varphi_t(x, y) = \varphi(t, x, y) = (xe^{\lambda t}, ye^{\mu t})$$

induces a tangent vector field. At point  $p$  (i.e.,  $t = 0$ ), it is

$$\begin{aligned} X_p = \gamma'_p(0) &= \frac{\partial(xe^{\lambda t})}{\partial t} \bigg|_{t=0} \frac{\partial}{\partial x} \bigg|_p + \frac{\partial(ye^{\mu t})}{\partial t} \bigg|_{t=0} \frac{\partial}{\partial y} \bigg|_p \\ &= \lambda xe^{\lambda t} \bigg|_{t=0} \frac{\partial}{\partial x} \bigg|_p + \mu ye^{\mu t} \bigg|_{t=0} \frac{\partial}{\partial y} \bigg|_p = \lambda x \frac{\partial}{\partial x} \bigg|_p + \mu y \frac{\partial}{\partial y} \bigg|_p. \end{aligned}$$

If we consider an arbitrary point  $q = \gamma_p(s)$  on the orbit  $\gamma_p$ , then at that point we have

$$\begin{aligned} X_q = \gamma'_q(s) &= \left. \frac{\partial(xe^{\lambda t})}{\partial t} \right|_{t=s} \frac{\partial}{\partial x} \Big|_q + \left. \frac{\partial(ye^{\mu t})}{\partial t} \right|_{t=s} \frac{\partial}{\partial y} \Big|_q \\ &= \lambda x e^{\lambda s} \frac{\partial}{\partial x} \Big|_q + \mu y e^{\mu s} \frac{\partial}{\partial y} \Big|_q. \end{aligned}$$

**Theorem 6.6.1** If a smooth manifold  $M$  is **compact** and  $X$  is a smooth tangent vector field on  $M$ , then  $X$  determines a **one-parameter differentiable transformation group** on  $M$ .

Establish a local coordinate system  $(U; x^i)$  at point  $p$ . Then the tangent vector field  $X$  can be expressed as

$$X_U = \sum_{i=1}^m X^i \frac{\partial}{\partial x^i},$$

where the components  $X^i$  are smooth functions, i.e.,  $X^i \in C^\infty$ . Solving the system of ordinary differential equations

$$\frac{dx^i}{dt} = X^i, \quad 1 \leq i \leq m,$$

with initial condition  $x^i(0) = x^i(p)$ , yields the one-parameter differentiable transformation group generated by  $X$ .

**Example 6.6.4** Consider a smooth tangent vector field on  $R^2$  given by  $X = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}$ .

In §6.2 we already solved the integral-curve equations for  $X$ , obtaining

$$x = C_1 \cos t + C_2 \sin t, \quad y = C_1 \sin t - C_2 \cos t.$$

Now we determine the constants  $C_1$  and  $C_2$ . According to the requirement of a one-parameter transformation group:  $\varphi_0$  must be the identity map, i.e., when  $t=0$ , we must have  $\varphi_0(x, y) = \varphi(0, x, y) = (x, y)$ . Substituting  $t=0$  gives

$$x = C_1 \cos 0 + C_2 \sin 0, \quad y = C_1 \sin 0 - C_2 \cos 0,$$

hence

$$C_1 = x, \quad C_2 = -y.$$

Therefore, the one-parameter differentiable transformation group generated by  $X$  is

$$\varphi_t(x, y) = \varphi(t, x, y) = (x \cos t - y \sin t, x \sin t + y \cos t).$$

In matrix form,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

This transformation group is precisely the  $SO(2)$  **group** (the group of rotations in the plane).

**Theorem 6.6.2** Let  $\varphi_t$  be a one-parameter differentiable transformation group acting on a smooth manifold  $M$ , and let  $X$  be the tangent vector field induced by  $\varphi_t$  on  $M$ . If  $\psi: M \rightarrow M$  is a diffeomorphism, then  $\psi \circ \varphi_t \circ \psi^{-1}$  is also a one-parameter transformation group, and the tangent vector field it induces on  $M$  is  $\psi_* X$ .

**Definition 6.6.2** Let  $X$  be a smooth tangent vector field on a smooth manifold  $M$ , and let  $\psi: M \rightarrow M$  be a diffeomorphism. If

$$\psi_* X = X,$$

then the tangent vector field  $X$  is said to be **invariant under  $\psi$** .

**Theorem 6.6.3** Let the local one-parameter transformation group determined by the tangent vector field  $X$  be  $\varphi_t$ . Then  $X$  is invariant under a diffeomorphism  $\psi: M \rightarrow M$  if and only if  $\psi$  commutes with  $\varphi_t$ , i.e.,

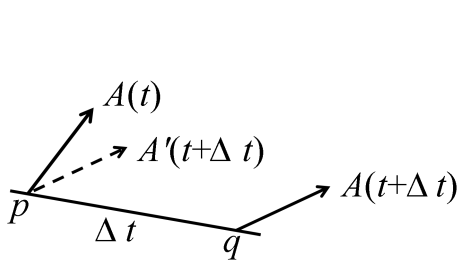
$$\varphi_t \circ \psi = \psi \circ \varphi(t), \quad \forall t.$$

## §6.7 Lie Derivative

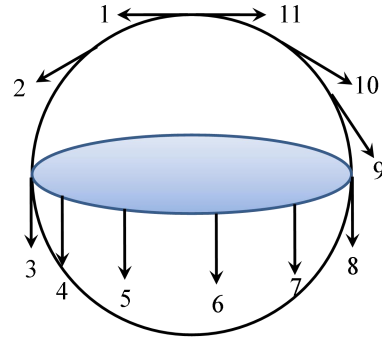
Using the one-parameter differentiable transformation group determined by a smooth tangent vector field  $X$ , we can differentiate smooth tangent vector fields on the manifold.

In Euclidean space, as shown in Figure 6.7.1, the derivative of a vector is computed by **parallel transporting** the vectors corresponding to different values of the parameter  $t$  to a common starting point, taking their difference, dividing by the parameter increment  $\Delta t$ , and then taking the limit as  $\Delta t \rightarrow 0$ . That is,

$$\frac{dA(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{A(t + \Delta t) - A(t)}{\Delta t}.$$



**Figure 6.7.1** Parallel transport of a tangent vector in Euclidean space



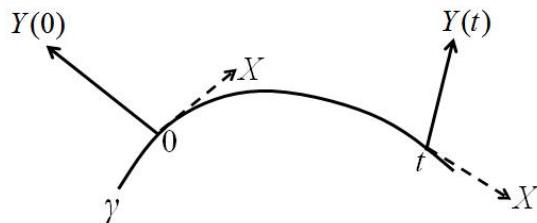
**Figure 6.7.2** Parallel transport of a tangent vector on a two-dimensional sphere

As we know, in Euclidean space, if the corresponding angles of two lines are equal, the lines are parallel. As shown in Figure 6.7.1, when a vector  $A(t + \Delta t)$  is transported along a straight line while maintaining a constant angle between the vector and the line, it eventually reaches position  $A'(t + \Delta t)$ . The starting points of  $A(t + \Delta t)$  and  $A'(t + \Delta t)$  coincide. During this transport, the magnitude and direction of the vector  $A(t + \Delta t)$  remain unchanged; therefore, the vector  $A'(t + \Delta t)$  at point  $p$  equals the vector  $A(t + \Delta t)$  at point  $q$ . Comparing the vector  $A'(t + \Delta t)$  with the vector  $A(t)$  at point  $p$  is equivalent to comparing the vector  $A(t)$  at point  $p$  with the vector  $A(t + \Delta t)$  at point  $q$ .

However, on a sphere, as illustrated in Figure 6.7.2, if we transport a vector along a closed curve (e.g., a great circle) while always keeping the angle between the vector and the curve constant at each step, we eventually find that when the vector returns to its starting position, its direction is opposite to the original direction. During this transport, the vector changes; it is no longer equal to the original vector. The reason is that Euclidean space is **flat**, whereas the sphere is **curved**.

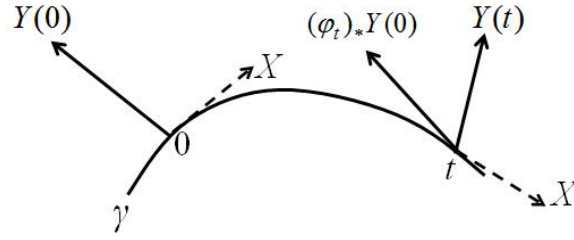
Then how do we compute the derivative of a tangent vector on a manifold like a sphere? We can make use of a one-parameter transformation group. As shown in Figure 6.7.3, let  $\varphi_t$  be the one-parameter differentiable transformation group determined by the tangent vector field  $X$ , and let  $\gamma$  be an integral curve of  $X$ . We want to compute the derivative of the tangent vector field  $Y$  along the curve  $\gamma$ . There are two approaches:

**First method:** As illustrated in Figure 6.7.4, using the one-parameter transformation group  $\varphi_t$ , we map the tangent vector  $Y(0)$  to the point on the curve  $\gamma$  with parameter value  $t$ , obtaining  $(\varphi_t)_* Y(0)$ . Then we subtract  $Y(t)$  from it, divide by the distance  $t$  along  $\gamma$  between the two points, and take the following limit:



**Figure 6.7.3** Change of vector field  $Y$  along curve  $\gamma$

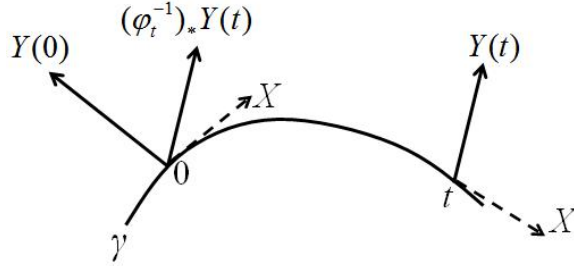
$$\lim_{t \rightarrow 0} \frac{Y(t) - (\varphi_t)_* Y(0)}{t}.$$



**Figure 6.7.4** First method for computing the derivative of vector field  $Y$  along curve  $\gamma$

**Second method:** As shown in Figure 6.7.5, using the inverse of the one-parameter transformation group,  $\varphi_t^{-1}$ , we map the tangent vector  $Y(t)$  to the point on the curve  $\gamma$  with parameter value 0, obtaining  $(\varphi_t^{-1})_* Y(t)$ . Then we subtract it from  $Y(0)$ , divide by the distance  $t$  along  $\gamma$ , and take the limit:

$$\lim_{t \rightarrow 0} \frac{(\varphi_t^{-1})_* Y(t) - Y(0)}{t}.$$



**Figure 6.7.5** Second method for computing the derivative of vector field  $Y$  along curve  $\gamma$

These two computation methods are equivalent. We have the following theorem.

**Theorem 6.7.1** Let  $X$  and  $Y$  be any two smooth tangent vector fields on a smooth manifold  $M$ . If the local one-parameter transformation group generated by  $X$  is  $\varphi_t$ , then

$$\lim_{t \rightarrow 0} \frac{Y(t) - (\varphi_t)_* Y(0)}{t} = \lim_{t \rightarrow 0} \frac{(\varphi_t^{-1})_* Y(t) - Y(0)}{t} = [X, Y].$$

Define

$$L_X Y = [X, Y],$$

and call  $L_X Y$  the **Lie derivative** of the tangent vector field  $Y$  with respect to  $X$ .  $L_X Y$  measures the rate of change of  $Y$  along the orbits of  $X$ .

We also define the **Lie derivative of a scalar field**  $f$  with respect to the tangent vector field  $X$  as the directional derivative of  $f$  along  $X$ , i.e.,

$$L_X f = X(f).$$

**Example 6.7.1** In  $R^2$ , the smooth tangent vector field  $X = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}$  generates the one-parameter differentiable transformation group

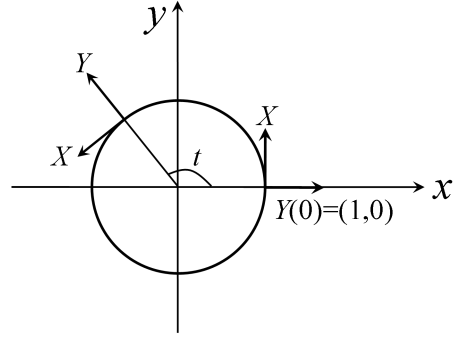
$$\varphi_t(x, y) = \varphi(t, x, y) = (x \cos t - y \sin t, x \sin t + y \cos t).$$

This transformation group is the rotation group in  $R^2$ , which can be written in matrix form as

$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

The smooth tangent vector field  $Y = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$  is

such that at each point, the tangent vector of  $Y$  is perpendicular to the tangent vector of  $X$  on the integral curve of  $X$ , as shown in Figure 6.7.6. We now compute the Lie derivative  $L_X Y$  using the definition.



**Figure 6.7.6** Computing the derivative of vector field  $Y$  along a circle

**First method:** At the point  $t = 0$ , we have  $Y(0) = (1, 0)$ . Using the one-parameter group, we push  $Y(0)$  forward to the point with parameter value  $t$ :

$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix},$$

namely,

$$(\varphi_t)_* Y(0) = (\cos t, \sin t).$$

The vector field  $Y = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$  expressed in terms of the parameter  $t$  is  $Y(t) = (\cos t, \sin t)$ .

Hence

$$L_X Y = \lim_{t \rightarrow 0} \frac{Y(t) - (\varphi_t)_* Y(0)}{t} = \lim_{t \rightarrow 0} \frac{(\cos t - \cos t, \sin t - \sin t)}{t} = (0, 0).$$

**Second method:** The inverse of the one-parameter transformation group is

$$(\varphi_t)^{-1} = \varphi_{-t} = \begin{pmatrix} \cos(-t) & -\sin(-t) \\ \sin(-t) & \cos(-t) \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

At the point with parameter  $t$ , we have  $Y(t) = (\cos t, \sin t)$ . Using  $(\varphi_t)^{-1}$  to pull  $Y(t)$  back to the point with parameter  $t = 0$ :

$$(\varphi_t^{-1})_* Y(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

namely,

$$(\varphi_t^{-1})_* Y(t) = (1, 0).$$

At  $t = 0$ ,  $Y(0) = (1, 0)$ . Therefore

$$L_X Y = \lim_{t \rightarrow 0} \frac{(\varphi_t^{-1})_* Y(t) - Y(0)}{t} = \lim_{t \rightarrow 0} \frac{(1, 0) - (1, 0)}{t} = (0, 0).$$

Now we compute  $[X, Y]$  to verify the result above. Let  $f$  be a smooth function defined on  $R^2$ . Then

$$\begin{aligned} [X, Y]f &= \left[ -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}, x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right] f \\ &= \left( -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right) \left( x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) f - \left( x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) \left( -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right) f \\ &= -y \frac{\partial}{\partial x} \left( x \frac{\partial f}{\partial x} \right) - y \frac{\partial}{\partial x} \left( y \frac{\partial f}{\partial y} \right) + x \frac{\partial}{\partial y} \left( x \frac{\partial f}{\partial x} \right) + x \frac{\partial}{\partial y} \left( y \frac{\partial f}{\partial y} \right) \\ &\quad + x \frac{\partial}{\partial x} \left( y \frac{\partial f}{\partial x} \right) - x \frac{\partial}{\partial x} \left( x \frac{\partial f}{\partial y} \right) + y \frac{\partial}{\partial y} \left( y \frac{\partial f}{\partial x} \right) - y \frac{\partial}{\partial y} \left( x \frac{\partial f}{\partial y} \right) \end{aligned}$$



$$\begin{aligned}
&= -y \frac{\partial f}{\partial x} - xy \frac{\partial^2 f}{\partial x^2} - y^2 \frac{\partial^2 f}{\partial x \partial y} + x^2 \frac{\partial^2 f}{\partial x \partial y} + x \frac{\partial f}{\partial y} + xy \frac{\partial^2 f}{\partial y^2} \\
&\quad + xy \frac{\partial^2 f}{\partial x^2} - x \frac{\partial f}{\partial y} - x^2 \frac{\partial^2 f}{\partial x \partial y} + y \frac{\partial f}{\partial x} + y^2 \frac{\partial^2 f}{\partial x \partial y} - xy \frac{\partial^2 f}{\partial y^2} \\
&= 0.
\end{aligned}$$

Therefore, the two computation methods described above are correct.

Based on the algebraic properties of the Poisson bracket, we can derive the following operational rules for the Lie derivative.

**Theorem 6.7.2** Let  $X, Y, Z$  be arbitrary smooth tangent vector fields on a manifold  $M$ , and let  $f \in C^\infty$ ,  $\lambda \in R$ . Then:

- 1)  $L_X Y = -L_Y X$  ;
- 2)  $L_X (Y + \lambda Z) = L_X Y + \lambda L_X Z$  ;
- 3)  $L_{X+Y} Z = L_X Z + L_Y Z$  ;
- 4)  $L_X (fY) = X(f) \cdot Y + fL_X Y$  ;
- 5)  $L_X ([Y, Z]) = [L_X Y, Z] + [Y, L_X Z]$ .

## §6.8 Singular Points

**Definition 6.8.1** Let  $X$  be a smooth tangent vector field on a smooth manifold  $M$ . If at a point  $p \in M$  we have  $X_p = 0$ , then  $p$  is called a **singular point**(or zero) of the smooth tangent vector field  $X$ .

**Example 6.8.1** As shown in Figure 6.2.4, at the origin the smooth tangent vector field  $X$  equals zero; thus the origin is a singular point of  $X$ .

**Example 6.8.2** On a unit sphere  $S^{2m-1} \subset R^{2m}$  with **odd** dimension, there exists a smooth tangent vector field that is **nowhere zero**(e.g., the “hair-combing” construction).

**Example 6.8.3** On a unit sphere  $S^{2m-1} \subset R^{2m}$  with **even** dimension, there does **not** exist a smooth tangent vector field that is nowhere zero (a consequence of the hairy ball theorem).

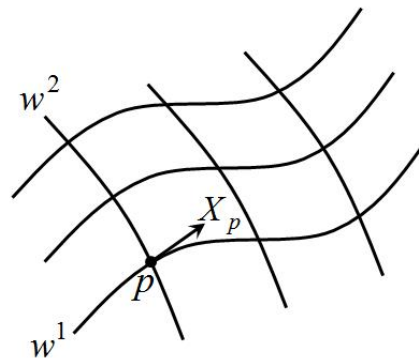
Generally, if a smooth manifold  $M$  is **compact** and its **Euler characteristic** is nonzero, then  $M$  does not admit a continuous tangent vector field that is nowhere zero.

**Example 6.8.4** On an  $m$ -dimensional torus  $T^m$ , there exist smooth tangent vector fields that are **nowhere zero**.

**Theorem 6.8.1** If a smooth tangent vector field  $X$  on a manifold  $M$  satisfies  $X_p \neq 0$  at a point  $p \in M$ , then there exists a local coordinate system  $(W; w^i)$  around  $p$  such that

$$X|_W = \frac{\partial}{\partial w^1}. \quad (6.8.1)$$

Take a two-dimensional smooth manifold as an example. As shown in Figure 6.8.1, since  $X_p \neq 0$  at point  $p \in M$ , there exists exactly one integral curve of  $X$  passing through  $p$ , and the tangent vector of this integral curve at  $p$  equals  $X_p$ . We thus take the family of integral curves of  $X$  as one set of coordinate curves, with the parameter chosen as  $w^1$ . Then we select another family of curves that intersect the integral-curve family as the other set of coordinate curves, with parameter  $w^2$ . This yields a two-dimensional curvilinear coordinate system  $(W; w^i)$  in a



**Figure 6.8.1** Adapted coordinate system

neighborhood  $W$  of  $p$ . In this coordinate system, the smooth tangent vector field  $X$  can be expressed as in equation (6.8.1). Such a coordinate system is called an **adapted coordinate system** for the smooth tangent vector field  $X$ .

## §6.9 Phase Plane

### 1. Phase Plane

We consider the system of differential equations:

$$\begin{cases} \frac{dx}{dt} = X(x, y), \\ \frac{dy}{dt} = Y(x, y), \end{cases} \quad (6.9.1)$$

where  $X(x, y)$  and  $Y(x, y)$  are assumed to have continuous partial derivatives, guaranteeing the existence and uniqueness of solutions.

If we regard  $t$  as a parameter and view a solution of (6.9.1) as a curve in the Euclidean plane with coordinates  $(x, y)$ , then this plane is called the **phase plane** of the system. Curves in the phase plane representing solutions are called **trajectories** or **integral curves**.

When  $X(x, y) \neq 0$ , the system (6.9.1) can be rewritten as

$$\frac{dy}{dx} = \frac{Y(x, y)}{X(x, y)}. \quad (6.9.2)$$

Alternatively, when  $Y(x, y) \neq 0$ , we can write

$$\frac{dx}{dy} = \frac{X(x, y)}{Y(x, y)}. \quad (6.9.3)$$

Since  $X(x, y)$  and  $Y(x, y)$  have continuous partial derivatives, the right-hand sides of (6.9.2) and (6.9.3) also have continuous partial derivatives. Therefore, solutions of (6.9.2) or (6.9.3) exist and are unique. Consequently, through every point  $(x, y)$  in the phase plane there passes **exactly one** integral curve (trajectory) of equation (6.9.2) or (6.9.3). That is, the integral curves (trajectories) of the solutions of (6.9.2) or (6.9.3) **do not intersect** in the phase plane.

### 2. Singular Points of the System

**Definition 6.9.1** A point  $(x^*, y^*)$  that simultaneously satisfies

$$X(x, y) = 0, \quad Y(x, y) = 0$$

is called a **singular point** (or equilibrium point) of system (6.9.1).

If  $X(x^*, y^*) = 0$ ,  $Y(x^*, y^*) = 0$  is a singular point, then

$$\begin{cases} \frac{dx^*}{dt} = X(x^*, y^*) = 0, \\ \frac{dy^*}{dt} = Y(x^*, y^*) = 0, \end{cases}$$

so

$$x = x^*, \quad y = y^*$$

is a solution of the system.

In the following, we assume  $X(x, y)$  and  $Y(x, y)$  are **linear** and study the types and behavior of singular points. System (6.9.1) becomes

$$\begin{cases} \frac{dx}{dt} = ax + by, \\ \frac{dy}{dt} = cx + dy. \end{cases} \quad (6.9.4)$$

where  $a, b, c, d$  are constants. One can verify that the origin  $x = 0, y = 0$  is a singular point. If

the determinant of the coefficient matrix is nonzero, i.e.,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0, \quad (6.9.5)$$

then this singular point is **unique**. We henceforth assume condition (6.9.5) always holds.

Since a linear transformation does not change the location of the singular point nor the qualitative behavior of the integral curves—and thus does not alter the type of the singular point—we perform a linear transformation on the system. Let this linear transformation be

$$\begin{cases} \xi = k_{11}x + k_{12}y, \\ \eta = k_{21}x + k_{22}y. \end{cases} \quad (6.9.6)$$

The characteristic equation of system (6.9.4) is

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0,$$

or written explicitly,

$$\lambda^2 - (a + d)\lambda + ad - bc = 0. \quad (6.9.7)$$

Next, we classify the type of singular point according to the roots of equation (6.9.7).

### 3.Types of Singular Points

#### (1)Distinct real roots

Suppose equation (6.9.7) has two distinct real roots  $\lambda_1$  and  $\lambda_2$ . When  $b \neq 0$ , define the linear transformation (6.9.6) as

$$\begin{cases} \xi = (d - \lambda_1)x - by, \\ \eta = (d - \lambda_2)x - by. \end{cases} \quad (6.9.8)$$

Then

$$\begin{aligned} \frac{d\xi}{dt} &= (d - \lambda_1) \frac{dx}{dt} - b \frac{dy}{dt} = (d - \lambda_1)(ax + by) - b(cx + dy) \\ &= (d - \lambda_1)ax - \lambda_1 by - bcx = (d - \lambda_1)ax + \lambda_1 \xi - \lambda_1 dx + \lambda_1^2 x - bcx \\ &= \lambda_1 \xi + [\lambda_1^2 - (a + d)\lambda_1 + ad - bc]x = \lambda_1 \xi, \\ \frac{d\eta}{dt} &= (d - \lambda_2) \frac{dx}{dt} - b \frac{dy}{dt} = (d - \lambda_2)(ax + by) - b(cx + dy) = \lambda_2 \eta. \end{aligned}$$

Thus system (6.9.4) reduces to the canonical form

$$\frac{d\xi}{dt} = \lambda_1 \xi, \quad \frac{d\eta}{dt} = \lambda_2 \eta. \quad (6.9.9)$$

If  $c \neq 0$ , then let the linear transformation (6.9.6) be

$$\begin{cases} \xi = -cx + (a - \lambda_1)y, \\ \eta = -cx + (a - \lambda_2)y, \end{cases}$$

Following similar computational steps as above, system (6.9.4) can also be reduced to the standard form (6.9.9).

If  $b = c = 0$ , then from the characteristic equation (6.9.7) we obtain

$$\begin{aligned} (a - \lambda)(d - \lambda) &= 0, \\ \lambda_1 &= a, \quad \lambda_2 = d. \end{aligned}$$

Take the linear transformation (6.9.6) as

$$\begin{cases} \xi = x, \\ \eta = y, \end{cases}$$

Then we get

$$\frac{d\xi}{dt} = \frac{dx}{dt} = ax = \lambda_1 \xi, \quad \frac{d\eta}{dt} = \frac{dy}{dt} = dy = \lambda_2 \eta.$$

Thus, system (6.9.4) can also be transformed into the standard form (6.9.9).

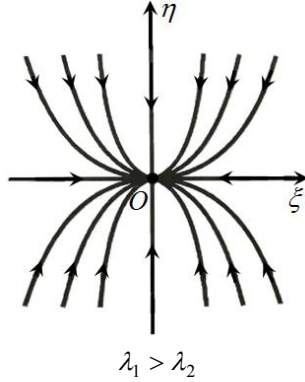
The solution of system (6.9.9) is

$$\xi(t) = Ae^{\lambda_1 t}, \quad \eta(t) = Be^{\lambda_2 t}, \quad (6.9.10)$$

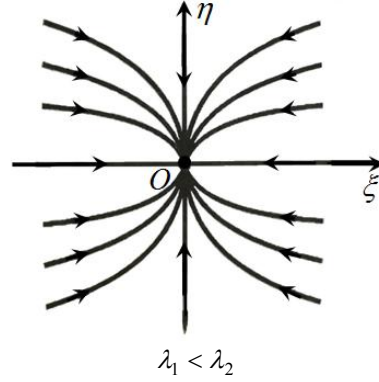
where  $A$  and  $B$  are arbitrary real constants. Next, we discuss the solution (6.9.10).

1) Real roots with the same sign but unequal values

Assume  $\lambda_1$  and  $\lambda_2$  are real numbers with the same sign but unequal. This can be divided into two cases.



**Figure 6.9.1** Double node



**Figure 6.9.2** Double node

① Both  $\lambda_1$  and  $\lambda_2$  are negative real numbers.

Suppose  $\lambda_1 < 0$ ,  $\lambda_2 < 0$ . When  $A = 0$ ,  $\xi(t) = 0$ . The positive and negative halves of the  $\eta$ -axis are trajectories (or integral curves; the same term applies below). When  $B = 0$ ,  $\eta(t) = 0$ . The positive and negative halves of the  $\xi$ -axis are trajectories.

If both  $A$  and  $B$  are nonzero, we consider the following two cases:

**First case:** If  $\lambda_1 > \lambda_2$ , then from the solution (6.9.10) we obtain the slope  $k$  of the tangent line to the trajectory at parameter  $t$  as

$$k = \frac{\eta(t)}{\xi(t)} = \frac{B}{A} e^{(\lambda_2 - \lambda_1)t}.$$

When  $t \rightarrow +\infty$ , we have  $k \rightarrow 0$ , so the tangent lines of the trajectories all become parallel to the  $\xi$ -axis, and the contact point is the coordinate origin. On the phase plane, the shape of the trajectories is as shown in Figure 6.9.1: all trajectories converge to the origin.

**Second case.** If  $\lambda_1 < \lambda_2$ , then from the solution (6.9.10) we obtain the reciprocal of the slope  $k$  of the tangent line to the trajectory at parameter  $t$  as

$$\frac{1}{k} = \frac{\xi(t)}{\eta(t)} = \frac{A}{B} e^{(\lambda_1 - \lambda_2)t}.$$

When  $t \rightarrow +\infty$ , we have  $\frac{1}{k} \rightarrow 0$ , so the tangent lines of the trajectories all become parallel to the  $\eta$ -axis, and the contact point is the coordinate origin. On the phase plane, the shape of the trajectories is as shown in Figure 6.9.2.

From the figure, we can see that all trajectories approach the origin, and except for a few special trajectories, they share the same tangent line at the singular point. A singular point whose neighboring trajectories exhibit such behavior is called a **double node**.

② Both  $\lambda_1$  and  $\lambda_2$  are positive real numbers

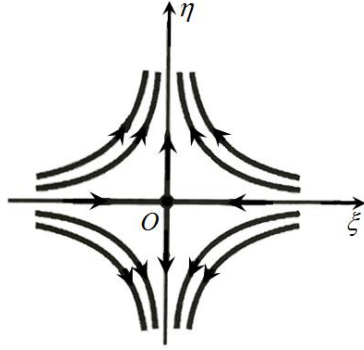
If  $\lambda_1$  and  $\lambda_2$  are both positive, we can still use an analysis similar to the one above, simply by replacing  $t \rightarrow +\infty$  with  $t \rightarrow -\infty$ ; that is, reversing the direction of the trajectories in the figure.

2) Real roots with opposite signs

Assume  $\lambda_1$  and  $\lambda_2$  are real numbers with opposite signs. When  $A = 0$ ,  $\xi(t) = 0$ ; the positive and negative halves of the  $\eta$ -axis are trajectories. When  $B = 0$ ,  $\eta(t) = 0$ ; the positive and negative halves of the  $\xi$ -axis are trajectories.

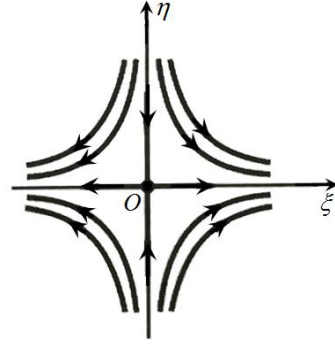
If both  $A$  and  $B$  are nonzero, we consider the following two cases:

First case: If  $\lambda_1 < 0$  and  $\lambda_2 > 0$ , then from solution (6.9.10) we see that as  $t \rightarrow +\infty$ ,  $\xi(t) \rightarrow 0$  and  $\eta(t) \rightarrow +\infty$ . On the phase plane, the shape of the trajectories is as shown in Figure 6.9.3.



$$\lambda_1 < 0, \quad \lambda_2 > 0$$

**Figure 6.9.3** Saddle point



$$\lambda_1 > 0, \quad \lambda_2 < 0$$

**Figure 6.9.4** Saddle point

Second case: If  $\lambda_1 > 0$  and  $\lambda_2 < 0$ , then from solution (6.9.10) we see that when  $t \rightarrow +\infty$ ,  $\xi(t) \rightarrow +\infty$  and  $\eta(t) \rightarrow 0$ . On the phase plane, the shape of the trajectories is as shown in Figure 6.9.4.

From the figure, we can see that in the neighborhood of the singular point, the solution trajectories of the system form a saddle shape. Such a singular point is called a **saddle point**.

### (2) Equal real roots

Suppose equation (6.9.7) has two equal real roots,  $\lambda_1 = \lambda_2 = \lambda$ . Then

$$\lambda_1 = \lambda_2 = \lambda = \frac{a+d}{2}.$$

Now we discuss in two cases below.

1) If  $b \neq 0$  or  $c \neq 0$ , define the linear transformation (6.9.6) as

$$\begin{cases} \xi = x, \\ \eta = (\lambda - d)x + by. \end{cases}$$

Then

$$\begin{aligned} \frac{d\xi}{dt} &= \frac{dx}{dt} = ax + by = ax + \eta - (\lambda - d)x = \eta + (a + d - \lambda)x = \eta + \lambda\xi, \\ \frac{d\eta}{dt} &= (\lambda - d)\frac{dx}{dt} + b\frac{dy}{dt} = (\lambda - d)(ax + by) + b(cx + dy) \\ &= (\lambda a - ad + bc)x + \lambda by = (\lambda a - ad + bc)x + \lambda\eta - \lambda(\lambda - d)x \\ &= -(\lambda^2 - \lambda a - \lambda d + ad - bc)x + \lambda\eta = \lambda\eta, \end{aligned}$$

Thus system (6.9.4) reduces to the canonical form

$$\frac{d\xi}{dt} = \lambda\xi + \eta, \quad \frac{d\eta}{dt} = \lambda\eta. \quad (6.9.11)$$

The solution of system (6.9.11) is

$$\xi(t) = (At + B)e^{\lambda t}, \quad \eta(t) = Ae^{\lambda t}, \quad (6.9.12)$$

where  $A, B$  are arbitrary real constants. We now discuss the solution (6.9.12).

①  $\lambda < 0$

Since  $\lambda < 0$ , when  $t \rightarrow +\infty$ , we have  $\xi(t) \rightarrow 0$ ,  $\eta(t) \rightarrow 0$ .

when  $A = 0$ , we have  $\eta(t) = 0$ , so the positive and negative halves of the  $\xi$ -axis are trajectories.

When  $A \neq 0$ ,

$$k = \frac{\eta(t)}{\xi(t)} = \frac{A}{At + B}.$$

When  $t \rightarrow +\infty$ , we have  $k \rightarrow 0$ . Hence, the tangents of the trajectories all approach the  $\xi$ -axis, and the contact point is the coordinate origin, as shown in Figure 6.9.5. When  $t = -\frac{B}{A}$ ,  $\xi(t) = 0$  but  $\eta(t) \neq 0$ , so the trajectories cross the  $\eta$ -axis. All trajectories approach the singular point along the direction of the  $\xi$ -axis. A singular point whose neighboring trajectories exhibit this behavior is called a **degenerate node**.

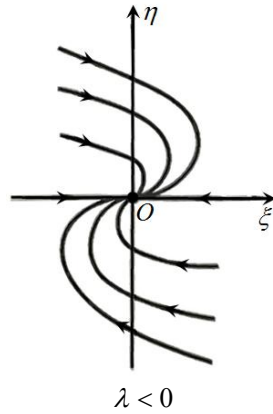


Figure 6.9.5 Degenerate node

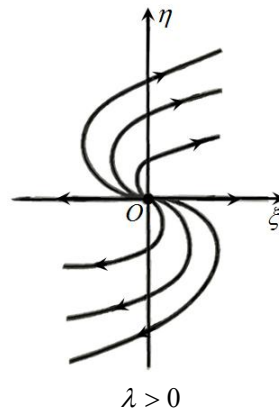


Figure 6.9.6 Degenerate node

②  $\lambda > 0$

When  $\lambda > 0$ , we still follow a similar analysis as before, but replace  $t \rightarrow +\infty$  with  $t \rightarrow -\infty$ . On the phase plane, the shape of the trajectories is as shown in Figure 6.9.6. All trajectories move away from the singular point, and this singular point is still called a **degenerate node**.

2) When  $b = c = 0$ , from the characteristic equation (6.9.7) we obtain

$$(a - \lambda)(d - \lambda) = 0,$$

$$\lambda = \lambda_1 = \lambda_2 = a = d.$$

System (6.9.4) is already in canonical form:

$$\frac{dx}{dt} = \lambda x, \quad \frac{dy}{dt} = \lambda y.$$

Its solution is

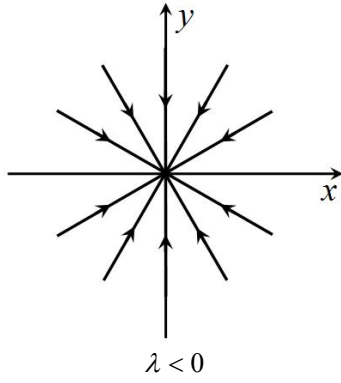
$$x(t) = Ae^{\lambda t}, \quad y(t) = Be^{\lambda t}. \quad (6.9.13)$$

Hence

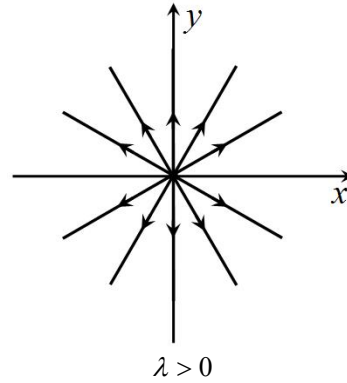
$$y = \frac{B}{A}x.$$

When  $\lambda < 0$ , if  $t \rightarrow +\infty$ , then  $x(t) \rightarrow 0$ ,  $y(t) \rightarrow 0$ . As shown in Figure 6.9.7, the trajectories are half-lines converging to the singular point. Such a singular point is called a **sink**(or stable node).

When  $\lambda > 0$ , if  $t \rightarrow +\infty$ , then  $x(t) \rightarrow \infty$ ,  $y(t) \rightarrow \infty$ . As shown in Figure 6.9.8, the trajectories are half-lines diverging from the singular point. Such a singular point is called a **source**(or unstable node). In both cases, the trajectories approach or leave the singular point along fixed directions, and the tangent directions differ for different trajectories.



**Figure 6.9.7** Sink



**Figure 6.9.8** Source

### (3)Complex roots with nonzero real part

When equation (6.9.7) has complex roots  $\lambda_1, \lambda_2$ , then  $\lambda_1$  must be the complex conjugate of  $\lambda_2$ . Let  $\lambda_1 = \mu + i\nu$ , then  $\lambda_2 = \mu - i\nu$ . From Vieta's formulas,

$$\lambda_1 + \lambda_2 = a + d = 2\mu.$$

Since  $\lambda_1$  is a root of (6.9.7), we have

$$\begin{aligned}(\mu + i\nu)^2 - (a + d)(\mu + i\nu) + ad - bc &= 0, \\ \mu^2 - \nu^2 + 2\mu i\nu - (a + d)\mu - (a + d)i\nu + ad - bc &= 0, \\ \mu^2 - (a + d)\mu + ad - bc &= \nu^2 - 2\mu i\nu + (a + d)i\nu = \nu^2.\end{aligned}$$

Define the linear transformation (6.9.6) as

$$\begin{cases} \xi = (d - \mu)x - by, \\ \eta = \nu x. \end{cases}$$

Then

$$\begin{aligned}\frac{d\xi}{dt} &= (d - \mu)\frac{dx}{dt} - b\frac{dy}{dt} = (d - \mu)(ax + by) - b(cx + dy) \\ &= adx - a\mu x + bdy - b\mu y - bcx - bdy \\ &= adx - a\mu x + \mu\xi - \mu(d - \mu)x - bcx \\ &= [\mu^2 - (a + d)\mu + ad - bc]x + \mu\xi = \nu^2 x + \mu\xi = \nu\eta + \mu\xi, \\ \frac{d\eta}{dt} &= \nu\frac{dx}{dt} = \nu(ax + by) = a\nu x + \nu by = a\nu x + \nu(d - \mu)x - \nu\xi \\ &= (a\nu + d\nu - \mu\nu)x - \nu\xi = (a + d - \mu)\nu x - \nu\xi = (2\mu - \mu)\nu x - \nu\xi \\ &= \mu\nu x - \nu\xi = -\nu\xi + \mu\eta.\end{aligned}$$

Thus system (6.9.4) reduces to the canonical form

$$\frac{d\xi}{dt} = \mu\xi + \nu\eta, \quad \frac{d\eta}{dt} = -\nu\xi + \mu\eta. \quad (6.9.14)$$

$\xi, \eta$  are Cartesian coordinates; we transform them into polar coordinates. Let  $\xi = r \cos \theta$ ,  $\eta = r \sin \theta$ . Then

$$\xi \frac{d\xi}{dt} + \eta \frac{d\eta}{dt} = r \frac{dr}{dt}, \quad \xi \frac{d\eta}{dt} - \eta \frac{d\xi}{dt} = r^2 \frac{d\theta}{dt}.$$

Transforming system (6.9.14) yields

$$\begin{aligned}\frac{dr}{dt} &= \frac{1}{r} \left( \xi \frac{d\xi}{dt} + \eta \frac{d\eta}{dt} \right) = \frac{1}{r} [\xi(\mu\xi + \nu\eta) + \eta(-\nu\xi + \mu\eta)] \\ &= \frac{1}{r} \mu(\xi^2 + \eta^2) = \mu r, \\ \frac{d\theta}{dt} &= \frac{1}{r^2} \left( \xi \frac{d\eta}{dt} - \eta \frac{d\xi}{dt} \right) = \frac{1}{r^2} [\xi(-\nu\xi + \mu\eta) - \eta(\mu\xi + \nu\eta)]\end{aligned}$$

$$= -\frac{1}{r^2} \nu (\xi^2 + \eta^2) = -\nu ,$$

which are

$$\frac{dr}{dt} = \mu r , \quad \frac{d\theta}{dt} = -\nu .$$

Solving this system gives

$$r = Ae^{\mu t}, \quad \theta = -\nu t + B, \quad (6.9.15)$$

where  $A > 0$  and  $B$  are arbitrary constants.

From the solution (6.9.15), we see that the trajectories are a family of **logarithmic spirals**. When  $\mu < 0$  and  $\nu > 0$ , as  $t$  increases, the trajectory spirals clockwise toward the origin, as shown in Figure 6.9.9. When  $\mu > 0$  and  $\nu > 0$ , as  $t$  increases, the trajectory spirals clockwise away from the origin, as shown in Figure 6.9.10. Such a singular point is called a **focus**(or spiral point).

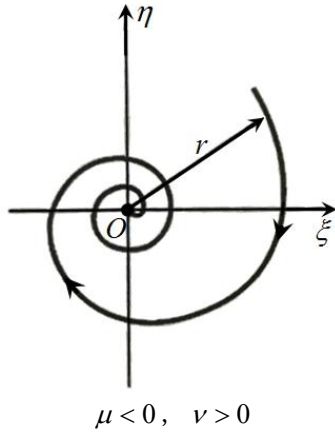


Figure 6.9.9 Focus (stable)

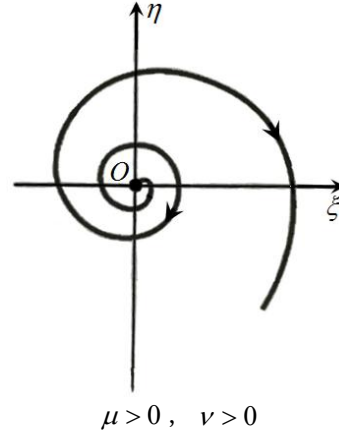


Figure 6.9.10 Focus (stable)

#### (4) Pure imaginary roots

If the complex roots of equation (6.9.7),  $\lambda_1, \lambda_2$  are **pure imaginary**, then  $\mu = 0$  in the canonical form (6.9.14). The solution (6.9.15) becomes

$$r = A, \quad \theta = -\nu t + B, \quad (6.9.16)$$

where  $A > 0$  and  $B$  are constants. From the solution (6.9.16), we see that the trajectories are a family of **concentric circles** centered at the origin, as shown in Figure 6.9.11. Such a singular point is called a **center**. When  $\nu > 0$ , the trajectories rotate clockwise. When  $\nu < 0$ , the trajectories rotate counterclockwise.

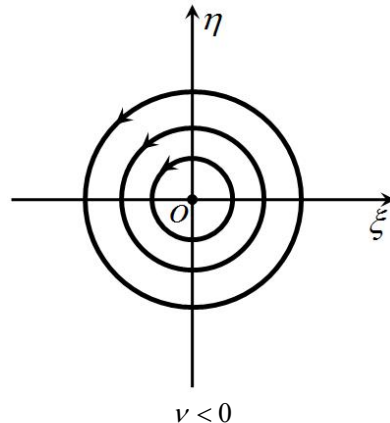
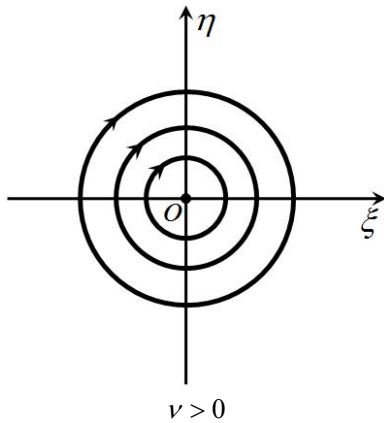


Figure 6.9.11 Center point



## Chapter 7 Tensors as Linear Operators

In this chapter, the underlying field is the real number field  $R$ . However, the conclusions remain valid if the real field is replaced by the complex field  $C$ .

### §7.1 Vector Spaces

**Definition 7.1.1** Let  $V$  be a non-empty set, and let  $k, l \in R, \alpha, \beta, \gamma, \lambda \in V$ . If the elements of  $V$  satisfy the following two sets of axioms, then  $V$  is called a **vector space over the real field  $R$** .

#### 1. Axioms of Addition

An operation called **addition** is defined between elements of  $V$ : for any two elements  $\alpha, \beta \in V$ , there exists a unique element  $\gamma \in V$  corresponding to them, denoted by  $\gamma = \alpha + \beta$ , called the **sum** of  $\alpha$  and  $\beta$ .

Addition satisfies the following rules:

1) Commutativity  $\alpha + \beta = \beta + \alpha$  ;

2) Associativity  $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$  ;

3) **Zero element** There exists an element  $0 \in V$ , called the **zero vector**, such that for every  $\alpha \in V$ ,

$$\alpha + 0 = \alpha ;$$

4) **Negative element** For each  $\alpha \in V$ , there exists an element  $\beta \in V$  ( $\beta$  is called the **negative element** of  $\alpha$ ) such that

$$\alpha + \beta = 0 .$$

#### 2. Axioms of Scalar Multiplication

An operation called **scalar multiplication** is defined between the real field  $R$  and the elements of  $V$ : for any real number  $k \in R$  and any element  $\alpha \in V$ , there exists a unique element  $\delta \in V$  corresponding to them, denoted by  $\delta = k\alpha$ , called the **scalar product** of  $k$  and  $\alpha$ .

Scalar multiplication satisfies the following rules:

$$1\alpha = \alpha ;$$

$$k(l\alpha) = (kl)\alpha, l \in R ;$$

$$(k + l)\alpha = k\alpha + l\alpha ;$$

$$k(\alpha + \beta) = k\alpha + k\beta .$$

**Definition 7.1.2** If every element  $v \in V$  can be **uniquely** expressed as a linear combination

$$v = x^1 e_1 + x^2 e_2 + \cdots + x^m e_m ,$$

where  $x^1, x^2, \dots, x^m \in R$  and the set  $\{e_1, e_2, \dots, e_m\}$  is a fixed subset of  $V$ , then  $V$  is called an  **$m$ -dimensional vector space**, and the set  $\{e_1, e_2, \dots, e_m\}$  is called a **basis** of  $V$ . If a basis of  $V$  consists of infinitely many elements, then  $V$  is called **infinite-dimensional**.

Once a basis  $\{e_1, e_2, \dots, e_m\}$  is chosen and fixed, each element  $v \in V$  corresponds **one-to-one** to an ordered array  $(x^1, x^2, \dots, x^m)$  ( $x^i \in R$ ) of numbers from the field. Thus, the vector space  $V$  can be identified with the space of ordered  $m$ -tuples  $(x^1, x^2, \dots, x^m)$ .

Since  $R$  is the field of real numbers, an array  $(x^1, x^2, \dots, x^m)$  is a point in  $R^m$ . Hence, a vector in an  $m$ -dimensional real vector space corresponds to an array  $(x^1, x^2, \dots, x^m)$ , establishing a one-to-one correspondence between the  $m$ -dimensional real vector space and  $R^m$ . Conversely, an array  $(x^1, x^2, \dots, x^m)$  or a point in  $R^m$  can be regarded as a vector in an  $m$ -dimensional real vector space.

**Definition 7.1.3** Two vector spaces  $V$  and  $W$  over the real field  $R$  are said to be **isomorphic** if there exists a bijective map  $f: V \rightarrow W$  such that for any  $\alpha, \beta \in V$  and any  $k \in R$ , the following properties hold:

- 1)  $f(\alpha + \beta) = f(\alpha) + f(\beta)$ ,
- 2)  $f(k\alpha) = kf(\alpha)$ .

Such a map  $f$  is called an **isomorphism**.

**Theorem 7.1.1** Two finite-dimensional vector spaces over the real field  $R$  are isomorphic **if and only if** they have the same dimension.

## §7.2 Dual Space

**Definition 7.2.1** A mapping  $f: V \rightarrow R$  is called a **linear functional** on the vector space  $V$  if for any  $v_1, v_2 \in V$  and any  $a^1, a^2 \in R$ ,

$$f(a^1 v_1 + a^2 v_2) = a^1 f(v_1) + a^2 f(v_2).$$

The set of all linear functionals on  $V$  is denoted by  $V^*$ . It can be verified that the addition of linear functionals and the multiplication of a linear functional by a scalar satisfy the axioms of a vector space. Therefore,  $V^*$  is itself a vector space, called the **dual space** of  $V$ .

## §7.3 Linear Mappings

**Definition 7.3.1** Let  $V$  and  $Z$  be vector spaces over  $R$ . A mapping  $f: V \rightarrow Z$  is called a **linear mapping** if for any two elements  $v_1, v_2 \in V$  and any scalars  $a^1, a^2 \in R$ ,

$$f(a^1 v_1 + a^2 v_2) = a^1 f(v_1) + a^2 f(v_2).$$

If  $Z=R$  (the real numbers), then the mapping  $f$  is called a **linear function**.

Each real number in  $R$  corresponds to a point on the real number line, and the real line is called a one-dimensional Euclidean space. Therefore, the field  $R$  itself can also be regarded as a **one-dimensional vector space** over  $R$ .

**Definition 7.3.2** Let  $V, W, Z$  be vector spaces over  $R$ . A mapping  $f: V \times W \rightarrow Z$  is called a **bilinear mapping** if it is linear in each variable separately; that is, for any  $v, v_1, v_2 \in V$ ,  $w, w_1, w_2 \in W$ , and any  $a^1, a^2 \in R$ ,

$$\begin{cases} f(a^1 v_1 + a^2 v_2, w) = a^1 f(v_1, w) + a^2 f(v_2, w), \\ f(v, a^1 w_1 + a^2 w_2) = a^1 f(v, w_1) + a^2 f(v, w_2), \end{cases}$$

If  $Z=R$ , then  $f$  is called a **bilinear function**.

**Example 7.3.1** Let  $v, v_1, v_2 \in R^3$ ,  $w, w_1, w_2 \in R^3$  be vectors. The **dot product** (scalar product)  $v \cdot w$  is a bilinear functional. This can be verified as follows:

$$\begin{cases} (a^1 v_1 + a^2 v_2) \cdot w = a^1 (v_1 \cdot w) + a^2 (v_2 \cdot w), \\ v \cdot (a^1 w_1 + a^2 w_2) = a^1 (v \cdot w_1) + a^2 (v \cdot w_2). \end{cases}$$

**Example 7.3.2** Let  $v, v_1, v_2 \in R^3$ ,  $w, w_1, w_2 \in R^3$  be vectors. The **cross product**  $v \times w$  is a bilinear mapping. Verification:

$$\begin{cases} (a^1 v_1 + a^2 v_2) \times w = a^1 (v_1 \times w) + a^2 (v_2 \times w), \\ v \times (a^1 w_1 + a^2 w_2) = a^1 (v \times w_1) + a^2 (v \times w_2). \end{cases}$$

**Definition 7.3.3** Let  $V_1, V_2, \dots, V_k$  and  $Z$  be vector spaces over  $R$ . A mapping

$$f: V_1 \times V_2 \times \dots \times V_k \rightarrow Z$$

is called a  **$k$ -multilinear mapping** if it is linear in each variable separately; that is, for any

$v_i, w_i \in V_i, 1 \leq i \leq k$ , and any scalars  $a^1, a^2 \in R$ ,

$$f(v_1, \dots, a^1 v_i + a^2 w_i, \dots, v_k) = a^1 f(v_1, \dots, v_i, \dots, v_k) + a^2 f(v_1, \dots, w_i, \dots, v_k).$$

If  $Z=R$ , then  $f$  is called a  **$k$ -multilinear function**.

The set of all  $k$ -multilinear mappings from  $V_1, V_2, \dots, V_k$  to  $Z$  is denoted by  $L(V_1, V_2, \dots, V_k; Z)$ . For  $f, g \in L(V_1, V_2, \dots, V_k; Z)$  and  $a \in R$ , define

$$(f+g)(v_1, v_2, \dots, v_k) = f(v_1, v_2, \dots, v_k) + g(v_1, v_2, \dots, v_k),$$

$$(af)(v_1, v_2, \dots, v_k) = a(f(v_1, v_2, \dots, v_k)).$$

Then  $f+g$  and  $af$  are also  $k$ -multilinear mappings, i.e.,  $f+g, af \in L(V_1, V_2, \dots, V_k; Z)$ . It can be verified that the set  $L(V_1, V_2, \dots, V_k; Z)$  equipped with these operations forms a vector space over  $R$ .

## §7.4 Tensors

At each point  $p$  of an  $m$ -dimensional smooth manifold  $M$ , there are mutually dual vector spaces: the tangent space  $T_p M$  and the cotangent space  $T_p^* M$ . Both are vector spaces, so we can also define a multilinear mapping at each point  $p$  of  $M$ ; such a multilinear mapping is called a **tensor**.

**Definition 7.4.1** A **tensor of type  $(r, s)$**  at a point  $p$  of a smooth manifold  $M$  is an  $(r+s)$ -multilinear mapping

$$\tau : \underbrace{T_p^* M \times \dots \times T_p^* M}_{r \text{ copies}} \times \underbrace{T_p M \times \dots \times T_p M}_{s \text{ copies}} \rightarrow R,$$

Here  $r$  is called the **contravariant order** and  $s$  the **covariant order** of  $\tau$ .

Each such mapping is precisely a tensor of type  $(r, s)$ . The set of all tensors of type  $(r, s)$  at point  $p$  is denoted by  $T_s^r(p)$ , or

$$L(\underbrace{T_p^* M, \dots, T_p^* M}_{r \text{ copies}}, \underbrace{T_p M, \dots, T_p M}_{s \text{ copies}}; R).$$

$T_s^r(p)$  is a vector space. In particular, tensors of type  $(r, 0)$  are called  **$r$ -th order contravariant tensors**, and tensors of type  $(0, s)$  are called  **$s$ -th order covariant tensors**. Elements of  $T_0^1(p)$  are type  $(1, 0)$  tensors, i.e.,  $T_0^1(p) = T_p M$ . Elements of  $T_1^0(p)$  are type  $(0, 1)$  tensors, i.e.,  $T_1^0(p) = T_p^* M$ . Elements of  $T_0^0(p)$  are just real numbers.

Choose a basis  $\{dx^1, \dots, dx^r\}$  for the vector space  $T_p^* M$  and a dual basis  $\left\{\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^s}\right\}$

for  $T_p M$ . Let  $v^{*1}, \dots, v^{*r} \in T_p^* M$  and  $v_1, \dots, v_s \in T_p M$ . Write

$$v^{*\lambda} = \alpha_i^\lambda dx^i, \quad 1 \leq \lambda \leq r,$$

$$v_\mu = \beta_\mu^i \frac{\partial}{\partial x^i}, \quad 1 \leq \mu \leq s,$$

$$\begin{aligned} f(v^{*1}, \dots, v^{*r}, v_1, \dots, v_s) &= \alpha_{i_1}^1 \dots \alpha_{i_r}^r \beta_1^{j_1} \dots \beta_s^{j_s} f(dx^{i_1}, \dots, dx^{i_r}, \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_s}}) \\ &= f^{i_1 \dots i_r}_{j_1 \dots j_s} \alpha_{i_1}^1 \dots \alpha_{i_r}^r \beta_1^{j_1} \dots \beta_s^{j_s}. \end{aligned} \quad (7.4.1)$$

Here

$$f^{i_1 \dots i_r}_{j_1 \dots j_s} = f(dx^{i_1}, \dots, dx^{i_r}, \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_s}}) \quad (7.4.2)$$

are the **components** of the tensor  $f$  with respect to the chosen bases.

From the basis representation of a tensor  $f$ , it follows that the dimension of the tensor space

$T_s^r(p)$  is  $m^{r+s}$ .

## §7.5 Tensor Product

**Definition 7.5.1** Let  $V$  and  $W$  be vector spaces,  $v \in V$ ,  $w \in W$ , and let  $V^*$  and  $W^*$  be the dual spaces of  $V$  and  $W$ , respectively. Let  $v^* \in V^*$ ,  $w^* \in W^*$ . Then the **tensor product**  $v^* \otimes w^*$  of  $v^*$  and  $w^*$  is a bilinear function on the product space  $V \times W$ , defined as

$$v^* \otimes w^*(v, w) = v^*(v)w^*(w) = \langle v, v^* \rangle \langle w, w^* \rangle. \quad (7.5.1)$$

**Definition 7.5.2** Let  $V^*$  and  $W^*$  be vector spaces, and let  $v^* \in V^*$ ,  $w^* \in W^*$ . The vector space formed by elements of the form  $v^* \otimes w^*$  is called the **tensor product** of  $V^*$  and  $W^*$ , denoted by  $V^* \otimes W^*$ .

The tensor product  $V^* \otimes W^*$  is essentially the direct product of the sets  $V^*$  and  $W^*$ .

**Definition 7.5.3** Let  $V$  and  $W$  be vector spaces, and let  $v \in V$ ,  $w \in W$ . The vector space formed by elements of the form  $v \otimes w$  is called the **tensor product** of  $V$  and  $W$ , denoted by  $V \otimes W$ .

At point  $p$  on a smooth manifold  $M$ , the tensor product can also be constructed from  $r$ -fold tensor product of the tangent space and the  $s$ -fold tensor product of the cotangent space, denoted as

$$\underbrace{T_p M \otimes \cdots \otimes T_p M}_{r \text{ copies}} \otimes \underbrace{T_p^* M \otimes \cdots \otimes T_p^* M}_{s \text{ copies}},$$

the dimension of the resulting vector space is

$$\underbrace{m \times \cdots \times m}_{r \text{ copies}} \times \underbrace{m \times \cdots \times m}_{s \text{ copies}} = m^{r+s}.$$

This vector space has the same dimension as the vector space of tensors of type  $(r, s)$  defined in Definition 7.4.1, and thus they are isomorphic, i.e.,

$$\underbrace{T_p M \otimes \cdots \otimes T_p M}_{r \text{ copies}} \otimes \underbrace{T_p^* M \otimes \cdots \otimes T_p^* M}_{s \text{ copies}} \cong L(\underbrace{T_p^* M, \dots, T_p^* M}_{r \text{ copies}}, \underbrace{T_p M, \dots, T_p M}_{s \text{ copies}}; R).$$

In a compatible local coordinate system  $(U; x^i)$  at point  $p$  on the smooth manifold  $M$ , each tangent space and cotangent space is equipped with a natural basis. Then a natural basis for the space  $T_s^r(p)$  can be obtained as follows:

$$\left. \frac{\partial}{\partial x^{i_1}} \right|_p \otimes \cdots \otimes \left. \frac{\partial}{\partial x^{i_r}} \right|_p \otimes dx^{j_1} \Big|_p \otimes \cdots \otimes dx^{j_s} \Big|_p, \\ 1 \leq i_1, \dots, i_r, j_1, \dots, j_s \leq m.$$

According to equation (7.5.1), for any  $v^{*1}, \dots, v^{*r} \in T_p^* M$  and  $v_1, \dots, v_s \in T_p M$ , we have (omitting the subscript  $p$ ):

$$\begin{aligned} & \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s} (v^{*1}, \dots, v^{*r}, v_1, \dots, v_s) \\ &= v^{*1} \left( \frac{\partial}{\partial x^{i_1}} \right) \cdots v^{*r} \left( \frac{\partial}{\partial x^{i_r}} \right) dx^{j_1}(v_1) \cdots dx^{j_s}(v_s) \\ &= \alpha_{i_1}^1 \cdots \alpha_{i_r}^r \beta_1^{j_1} \cdots \beta_s^{j_s}. \end{aligned}$$

Rewriting equation (7.4.1),

$$\begin{aligned} & f(v^{*1}, \dots, v^{*r}, v_1, \dots, v_s) \\ &= f^{i_1 \cdots i_r j_1 \cdots j_s} \alpha_{i_1}^1 \cdots \alpha_{i_r}^r \beta_1^{j_1} \cdots \beta_s^{j_s}, \end{aligned}$$

$$= f^{i_1 \cdots i_r}_{j_1 \cdots j_s} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s} (v^{*1}, \dots, v^{*r}, v_1, \dots, v_s).$$

Thus, a tensor  $f$  of type  $(r, s)$  can be expressed as

$$f = f^{i_1 \cdots i_r}_{j_1 \cdots j_s} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s},$$

where  $f^{i_1 \cdots i_r}_{j_1 \cdots j_s}$  are as shown in equation (7.4.2).

**Example 7.5.1** The tensor space  $T_1^2(p)$  of type  $(2, 1)$  at point  $p$  on a 2-dimensional manifold  $M$  is the vector space generated by taking the tensor product  $\otimes$  of two tangent spaces  $T_p M$  and one cotangent space  $T_p^* M$ :

$$T_p M \otimes T_p M \otimes T_p^* M.$$

Since  $m = 2, r = 2, s = 1$ , thus, the dimension of  $T_1^2(p)$  is  $2^{2+1} = 8$ , and its basis consists of eight basis vectors. Let  $(U; x^i)$  be a compatible local coordinate system at  $p$ . The natural basis for  $T_p M$  is  $\left\{ \frac{\partial}{\partial x^1} \Big|_p, \frac{\partial}{\partial x^2} \Big|_p \right\}$ , and the natural basis for  $T_p^* M$  is  $\{dx^1 \Big|_p, dx^2 \Big|_p\}$ . Then the natural

basis for  $T_1^2(p)$  consists of the following eight elements:

$$\begin{aligned} & \frac{\partial}{\partial x^1} \Big|_p \otimes \frac{\partial}{\partial x^1} \Big|_p \otimes dx^1 \Big|_p, & \frac{\partial}{\partial x^1} \Big|_p \otimes \frac{\partial}{\partial x^1} \Big|_p \otimes dx^2 \Big|_p, & \frac{\partial}{\partial x^1} \Big|_p \otimes \frac{\partial}{\partial x^2} \Big|_p \otimes dx^1 \Big|_p, \\ & \frac{\partial}{\partial x^1} \Big|_p \otimes \frac{\partial}{\partial x^2} \Big|_p \otimes dx^2 \Big|_p, & \frac{\partial}{\partial x^2} \Big|_p \otimes \frac{\partial}{\partial x^1} \Big|_p \otimes dx^1 \Big|_p, & \frac{\partial}{\partial x^2} \Big|_p \otimes \frac{\partial}{\partial x^1} \Big|_p \otimes dx^2 \Big|_p, \\ & \frac{\partial}{\partial x^2} \Big|_p \otimes \frac{\partial}{\partial x^2} \Big|_p \otimes dx^1 \Big|_p, & \frac{\partial}{\partial x^2} \Big|_p \otimes \frac{\partial}{\partial x^2} \Big|_p \otimes dx^2 \Big|_p. \end{aligned}$$

Any tensor  $t_1^2 \in T_1^2(p)$  of type  $(2, 1)$  can be expressed as

$$t_1^2 = t^{ij}_k \frac{\partial}{\partial x^i} \Big|_p \otimes \frac{\partial}{\partial x^j} \Big|_p \otimes dx^k \Big|_p,$$

where  $t^{ij}_k = t_1^2 \left( dx^i \Big|_p, dx^j \Big|_p, \frac{\partial}{\partial x^k} \Big|_p \right)$ .

## §7.6 Tensor Fields

Let

$$T_s^r(M) = \bigcup_{p \in M} T_s^r(p),$$

which is called the **bundle of tensors of type  $(r, s)$**  on the smooth manifold  $M$ . It is the collection of all smooth tensors of type  $(r, s)$  at every point of  $M$ .

**Definition 7.6.1** A **tensor field  $\tau$  of type  $(r, s)$**  on a smooth manifold  $M$  assigns to each point of  $M$  a tensor of type  $(r, s)$ ; that is, it is a map

$$\tau : M \rightarrow T_s^r(M),$$

such that for every point  $p \in M$ ,  $\tau(p) \in T_s^r(p)$  is a tensor of type  $(r, s)$ .

A tensor field of type  $(r, 0)$  on manifold  $M$  is called an  **$r$ -th order contravariant tensor field**, and a tensor field of type  $(0, s)$  is called an  **$s$ -th order covariant tensor field**.

**Definition 7.6.2** A tensor field  $\tau$  of type  $(r, s)$  is said to be **smooth** if for every point  $p \in M$ , there exists a compatible local coordinate system  $(U; x^i)$  around  $p$  such that the restriction of  $\tau$  to  $U$  can be expressed as

$$\tau|_U = \tau_{j_1 \dots j_s}^{i_1 \dots i_r} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \dots \otimes dx^{j_s},$$

and the components  $\tau_{j_1 \dots j_s}^{i_1 \dots i_r}$  are smooth functions on  $U$ .

A smooth tensor field of type  $(1,0)$  is a **smooth tangent vector field**. The set of all smooth tangent vector fields is denoted by

$$T_0^1(M) = \chi(M).$$

A smooth tensor field of type  $(0,1)$  is a **smooth covariant vector field of order 1** (i.e., a cotangent vector field), also called a **1-form**. The set of all 1-forms is denoted by

$$T_1^0(M) = A^1(M).$$

A smooth tensor field of type  $(0,0)$  is a **smooth function** on  $M$ . The set of all smooth functions is denoted by

$$T_0^0(M) = C^\infty(M).$$

In the set  $T_s^r(M)$  of all smooth tensor fields, each smooth tensor field can be multiplied by a number or a smooth function, and any two smooth tensor fields can be added or subjected to the tensor product operation pointwise.

**Example 7.6.1** Let  $f \in C^\infty(M)$ . Under a compatible local coordinate system  $(U; x^i)$  at point  $p$  on the manifold  $M$ ,

$$df|_U = \sum_{i=1}^m \frac{\partial f}{\partial x^i} dx^i,$$

where  $\frac{\partial f}{\partial x^i} \in C^\infty(U)$ . Thus, the differential  $df$  of  $f$  is a 1-form on  $M$ .

**Theorem 7.6.1** Let  $\tau$  be a smooth tensor field of type  $(r,s)$  on a smooth manifold  $M$ . Then  $\tau$  induces an  $(r+s)$ -multilinear map

$$\tilde{\tau} : \underbrace{A^1(M) \times \dots \times A^1(M)}_{r \text{ copies}} \times \underbrace{\chi(M) \times \dots \times \chi(M)}_{s \text{ copies}} \rightarrow C^\infty(M)$$

defined as follows: for any  $\alpha^1, \dots, \alpha^r \in A^1(M)$ ,  $X_1, \dots, X_s \in \chi(M)$  and  $p \in M$ ,

$$\begin{aligned} & \tilde{\tau}(\alpha^1, \dots, \alpha^r, X_1, \dots, X_s)(p) \\ &= \tau(p)(\alpha^1(p), \dots, \alpha^r(p), X_1(p), \dots, X_s(p)), \end{aligned} \quad (7.6.1)$$

and the map  $\tilde{\tau}$  is  $C^\infty(M)$ -linear in each argument, i.e.,

$$\begin{aligned} & \tilde{\tau}(\alpha^1, \dots, \alpha^\lambda + f\alpha'^\lambda, \dots, \alpha^r, X_1, \dots, X_s) \\ &= \tilde{\tau}(\alpha^1, \dots, \alpha^\lambda, \dots, \alpha^r, X_1, \dots, X_s) + f \cdot \tilde{\tau}(\alpha^1, \dots, \alpha'^\lambda, \dots, \alpha^r, X_1, \dots, X_s) \\ & \quad (\forall \alpha^\lambda, \alpha'^\lambda \in A^1(M), f \in C^\infty(M), 1 \leq \lambda \leq r); \\ & \tilde{\tau}(\alpha^1, \dots, \alpha^r, X_1, \dots, X_\mu + fX'_\mu, \dots, X_s) \\ &= \tilde{\tau}(\alpha^1, \dots, \alpha^r, X_1, \dots, X_\mu, \dots, X_s) + f \cdot \tilde{\tau}(\alpha^1, \dots, \alpha^r, X_1, \dots, X'_\mu, \dots, X_s) \\ & \quad (\forall X_\mu, X'_\mu \in \chi(M), f \in C^\infty(M), 1 \leq \mu \leq s). \end{aligned}$$

In particular,  $\tilde{\tau}$  is also  $R$ -linear.

Conversely, if any  $(r+s)$ -multilinear map

$$\tilde{\tau} : \underbrace{A^1(M) \times \dots \times A^1(M)}_{r \uparrow} \times \underbrace{\chi(M) \times \dots \times \chi(M)}_{s \uparrow} \rightarrow C^\infty(M),$$

is  $C^\infty(M)$ -linear in each argument, then there exists a unique smooth tensor field  $\tau$  of type  $(r,s)$  on  $M$  satisfying (7.6.1).

This is a very important theorem. It tells us that an  $(r+s)$ -multilinear map  $\tilde{\tau}$  on a smooth manifold  $M$  and the corresponding smooth tensor field  $\tilde{\tau}$  of type  $(r,s)$  are essentially the same. The linear map  $\tilde{\tau}$  or the smooth tensor field  $\tilde{\tau}$  can also be viewed as a linear operator. The

theorem further shows that  $C^\infty(M)$ –**linearity** is the characteristic property of smooth tensor fields, also known as the **tensor property**. Therefore, to determine whether a given multilinear map

$$\tilde{\tau} : \underbrace{A^1(M) \times \cdots \times A^1(M)}_{r \text{ copies}} \times \underbrace{\chi(M) \times \cdots \times \chi(M)}_{s \text{ copies}} \rightarrow C^\infty(M),$$

corresponds to a smooth tensor field, one simply needs to verify that it is  $C^\infty(M)$ –linearity in each argument.

## Chapter 8 Differential Forms

### §8.1 Permutations and Permutation Groups

To understand exterior products and exterior differentiation, it is necessary first to learn about permutation groups.

Suppose there are five symbols: 1, 2, 3, 4, 5, originally arranged in the order  $\{1,2,3,4,5\}$ . If we swap 1 with 2 and 4 with 5, we obtain the new arrangement  $\{2,1,3,5,4\}$ . Such a reordering operation is called a **permutation**.

Conversely, given a set of numbers  $\{2,1,3,5,4\}$ , suppose they are obtained by permuting  $\{1,2,3,4,5\}$ . How can we determine what permutation was applied? There is a method to solve this problem.

**Definition 8.1.1** In a permutation, if a preceding number is greater than a following number, they are called an **inversion pair**. The total number of inversion pairs in a permutation is called the **inversion number** of the permutation.

For example, in the permutation  $\{2,1,3,5,4\}$ , the inversion pairs are (2,1) and (5,4), so the inversion number is 2. In the permutation  $\{4,1,5,2,3\}$ , the inversion pairs are (4,1), (4,2), (4,3), (5,2) and (5,3), so the inversion number is 5.

**Definition 8.1.2** A permutation with an even inversion number is called an **even permutation**; a permutation with an odd inversion number is called an **odd permutation**.

For example,  $\{2,1,3,5,4\}$  is an even permutation;  $\{4,1,5,2,3\}$  is an odd permutation; the inversion number of  $\{1,2,3,4,5\}$  is zero, so it is an even permutation.

A permutation operation is denoted by the symbol  $\sigma$ . The **parity** of a permutation is denoted by  $\text{sgn } \sigma$ . If  $\text{sgn } \sigma = 1$ , it indicates an even permutation; if  $\text{sgn } \sigma = -1$ , it indicates an odd permutation, i.e.,

$$\text{sgn } \sigma = \begin{cases} 1, & \sigma \text{ is an even permutation,} \\ -1, & \sigma \text{ is an odd permutation.} \end{cases}$$

The relationship between  $\text{sgn } \sigma$  and the inversion number is

$$\text{sgn } \sigma = (-1)^{\text{inversion number of } \sigma}.$$

A permutation operation can be expressed in matrix form. For example, transforming  $\{1,2,3,4,5\}$  into  $\{4,1,5,2,3\}$  can be written as the matrix:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 2 & 3 \end{pmatrix}.$$

For  $\{1,2,3\}$ , a total of 6 permutations can be performed, as shown in Table 8.1.1.

**Table 8.1.1 Permutation Operations on  $\{1,2,3\}$**

Permutation Operation $\sigma$	Inversion Pairs	$\text{sgn } \sigma$
$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$	None	$\text{sgn } \sigma_1 = (-1)^0 = 1$
$\sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$	(2,1)	$\text{sgn } \sigma_2 = (-1)^1 = -1$
$\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$	(3,2), (3,1), (2,1)	$\text{sgn } \sigma_3 = (-1)^3 = -1$



$\sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$	$(3,2)$	$\text{sgn } \sigma_4 = (-1)^1 = -1$
$\sigma_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$	$(3,1), (3,2)$	$\text{sgn } \sigma_5 = (-1)^2 = 1$
$\sigma_6 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$	$(2,1), (3,1)$	$\text{sgn } \sigma_6 = (-1)^2 = 1$

These six permutation operations,  $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$ , and  $\sigma_6$ , form a group, called the **permutation group**, where the identity element is  $\sigma_1$ . Among these six permutations, the product of any two permutations equals another permutation. For example,

$$\sigma_2 \sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

Starting from the matrix on the right: in the first column of the right matrix, 1 is mapped to 3; then in the third column of the left matrix, 3 is mapped to 3. Combining both matrices, this maps 1 to 3:

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

In the second column of the right matrix, 2 is mapped to 2; then in the second column of the left matrix, 2 is mapped to 1. Combining both matrices, this maps 2 to 1. In the third column of the right matrix, 3 is mapped to 1; then in the first column of the left matrix, 1 is mapped to 2. Combining both matrices, this maps 3 to 2. Therefore,

$$\sigma_2 \sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \sigma_5.$$

The permutation group on the natural numbers  $\{1, 2, \dots, r\}$  is denoted by  $G_r$ . For example, the permutation group on  $\{1, 2, 3\}$  is denoted by  $G_3$ .

## §8.2 Symmetric and Antisymmetric Covariant Tensors

**Definition 8.2.1** Let  $\xi$  be an  $s$ -th order covariant tensor at point  $p$  on manifold  $M$ , i.e.,  $\xi(v_1, \dots, v_s) \in T_p M$  is an  $s$ -multilinear map

$$\xi : \underbrace{T_p M \times \dots \times T_p M}_{s \text{ copies}} \rightarrow R.$$

If swapping any two arguments leaves the value of  $\xi$  unchanged, i.e., for all permutations  $\sigma \in G_s$ ,

$$(\sigma \xi)(v_1, \dots, v_s) = \xi(v_{\sigma(1)}, \dots, v_{\sigma(s)}) = \xi(v_1, \dots, v_s),$$

then  $\xi$  is called a **symmetric  $s$ -th order covariant tensor**. If swapping any two arguments only changes the sign, i.e., for all permutations  $\sigma \in G_s$ ,

$$(\sigma \xi)(v_1, \dots, v_s) = \xi(v_{\sigma(1)}, \dots, v_{\sigma(s)}) = (\text{sgn } \sigma) \xi(v_1, \dots, v_s),$$

then  $\xi$  is called an **antisymmetric  $s$ -th order covariant tensor**.

**Example 8.2.1** The dot product of two vectors  $v_1, v_2$  is a symmetric tensor. Let

$$\xi(v_1, v_2) = v_1 \cdot v_2.$$

Denote by  $G_2$  the permutation group on the natural numbers  $\{1, 2\}$ . This group has only two elements:

$$\sigma_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Since

$$\xi(v_{\sigma_1(1)}, v_{\sigma_1(2)}) = \xi(v_1, v_2) = v_1 \cdot v_2,$$

$$\xi(v_{\sigma_2(1)}, v_{\sigma_2(2)}) = \xi(v_2, v_1) = v_2 \cdot v_1 = v_1 \cdot v_2 = \xi(v_1, v_2),$$

and thus  $\xi$  is a symmetric 2nd-order covariant tensor.

**Example 8.2.2** Let

$$v_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}, \quad v_3 = \begin{pmatrix} a_3 \\ b_3 \\ c_3 \end{pmatrix}$$

be vectors in  $R^3$ .

$$\xi(v_1, v_2, v_3) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

is a 3-linear map. Denote by  $G_3$  the permutation group on the natural numbers  $\{1, 2, 3\}$ ; the elements of  $G_3$  are given in Table 8.1.1. Since

$$\begin{aligned} \xi(v_{\sigma_1(1)}, v_{\sigma_1(2)}, v_{\sigma_1(3)}) &= \xi(v_1, v_2, v_3) = (\text{sgn } \sigma_1) \xi(v_1, v_2, v_3), \\ \xi(v_{\sigma_2(1)}, v_{\sigma_2(2)}, v_{\sigma_2(3)}) &= \xi(v_2, v_1, v_3) = \begin{vmatrix} a_2 & a_1 & a_3 \\ b_2 & b_1 & b_3 \\ c_2 & c_1 & c_3 \end{vmatrix} \\ &= -\xi(v_1, v_2, v_3) = (\text{sgn } \sigma_2) \xi(v_1, v_2, v_3), \\ \xi(v_{\sigma_3(1)}, v_{\sigma_3(2)}, v_{\sigma_3(3)}) &= \xi(v_3, v_2, v_1) = \begin{vmatrix} a_3 & a_2 & a_1 \\ b_3 & b_2 & b_1 \\ c_3 & c_2 & c_1 \end{vmatrix} \\ &= -\xi(v_1, v_2, v_3) = (\text{sgn } \sigma_3) \xi(v_1, v_2, v_3), \\ \xi(v_{\sigma_4(1)}, v_{\sigma_4(2)}, v_{\sigma_4(3)}) &= \xi(v_1, v_3, v_2) = \begin{vmatrix} a_1 & a_3 & a_2 \\ b_1 & b_3 & b_2 \\ c_1 & c_3 & c_2 \end{vmatrix} \\ &= -\xi(v_1, v_2, v_3) = (\text{sgn } \sigma_4) \xi(v_1, v_2, v_3), \\ \xi(v_{\sigma_5(1)}, v_{\sigma_5(2)}, v_{\sigma_5(3)}) &= \xi(v_3, v_1, v_2) = \begin{vmatrix} a_3 & a_1 & a_2 \\ b_3 & b_1 & b_2 \\ c_3 & c_1 & c_2 \end{vmatrix} \\ &= \xi(v_1, v_2, v_3) = (\text{sgn } \sigma_5) \xi(v_1, v_2, v_3), \\ \xi(v_{\sigma_6(1)}, v_{\sigma_6(2)}, v_{\sigma_6(3)}) &= \xi(v_2, v_3, v_1) = \begin{vmatrix} a_2 & a_3 & a_1 \\ b_2 & b_3 & b_1 \\ c_2 & c_3 & c_1 \end{vmatrix} \\ &= \xi(v_1, v_2, v_3) = (\text{sgn } \sigma_6) \xi(v_1, v_2, v_3), \end{aligned}$$

Thus,  $\xi$  is an antisymmetric 3rd-order covariant tensor.

**Theorem 8.2.1**  $\xi$  is a symmetric tensor if and only if its components are symmetric with respect to all indices;  $\xi$  is an antisymmetric tensor if and only if its components are antisymmetric with respect to all indices.

Let  $P^s(V^*)$  denote the set of all symmetric  $s$ -th order covariant tensors, and let  $\Lambda^s(V^*)$  denote the set of all antisymmetric  $s$ -th order covariant tensors.

### §8.3 Symmetrization and Antisymmetrization Operators

Given any  $s$ -multilinear function  $\xi(v_1, \dots, v_s)$ , we can generate a symmetric or antisymmetric  $s$ -multilinear function by using the symmetrization and antisymmetrization

operators.

**Definition 8.3.1** Let  $\xi(v_1, \dots, v_s)$  be an  $s$ -multilinear function. Define

$$S_s(\xi) = \sum_{\sigma \in G_s} \sigma \xi, \quad A_s(\xi) = \sum_{\sigma \in G_s} \text{sgn } \sigma \cdot \sigma \xi.$$

These are called the **symmetrization operator** and the **antisymmetrization operator** for  $s$ -th order covariant tensors, respectively.

**Example 8.3.1** Let  $\xi(v_1, v_2, v_3)$  be a 3-linear function. We use it to generate a symmetric 3-linear function. The expression  $S_3(\xi) = \sum_{\sigma \in G_3} \sigma \xi$  means applying all elements (i.e., permutations) of the permutation group  $G_3$  to the function  $\xi$  and summing all resulting functions. For this specific 3-linear function  $\xi$ , the permutation group is  $G_3$ .

$$\begin{aligned} S_3(\xi) &= \sum_{\sigma \in G_3} \sigma \xi = \sigma_1 \xi + \sigma_2 \xi + \sigma_3 \xi + \sigma_4 \xi + \sigma_5 \xi + \sigma_6 \xi \\ &= \xi(v_{\sigma_1(1)}, v_{\sigma_1(2)}, v_{\sigma_1(3)}) + \xi(v_{\sigma_2(1)}, v_{\sigma_2(2)}, v_{\sigma_2(3)}) + \xi(v_{\sigma_3(1)}, v_{\sigma_3(2)}, v_{\sigma_3(3)}) \\ &\quad + \xi(v_{\sigma_4(1)}, v_{\sigma_4(2)}, v_{\sigma_4(3)}) + \xi(v_{\sigma_5(1)}, v_{\sigma_5(2)}, v_{\sigma_5(3)}) + \xi(v_{\sigma_6(1)}, v_{\sigma_6(2)}, v_{\sigma_6(3)}) \\ &= \xi(v_1, v_2, v_3) + \xi(v_2, v_1, v_3) + \xi(v_3, v_2, v_1) \\ &\quad + \xi(v_1, v_3, v_2) + \xi(v_3, v_1, v_2) + \xi(v_2, v_3, v_1). \end{aligned}$$

It can be verified that  $S_3(\xi)$  is symmetric.

The expression  $A_3(\xi) = \sum_{\sigma \in G_3} \text{sgn } \sigma \cdot \sigma \xi$  means taking all permutations  $\sigma$  in  $G_3$ , multiplying by  $\text{sgn } \sigma$ , applying them to  $\xi$ , and summing all resulting functions.

$$\begin{aligned} A_3(\xi) &= \sum_{\sigma \in G_3} \text{sgn } \sigma \cdot \sigma \xi \\ &= \text{sgn } \sigma_1 \cdot \sigma_1 \xi + \text{sgn } \sigma_2 \cdot \sigma_2 \xi + \text{sgn } \sigma_3 \cdot \sigma_3 \xi + \text{sgn } \sigma_4 \cdot \sigma_4 \xi + \text{sgn } \sigma_5 \cdot \sigma_5 \xi + \text{sgn } \sigma_6 \cdot \sigma_6 \xi \\ &= (\text{sgn } \sigma_1) \xi(v_{\sigma_1(1)}, v_{\sigma_1(2)}, v_{\sigma_1(3)}) + (\text{sgn } \sigma_2) \xi(v_{\sigma_2(1)}, v_{\sigma_2(2)}, v_{\sigma_2(3)}) \\ &\quad + (\text{sgn } \sigma_3) \xi(v_{\sigma_3(1)}, v_{\sigma_3(2)}, v_{\sigma_3(3)}) + (\text{sgn } \sigma_4) \xi(v_{\sigma_4(1)}, v_{\sigma_4(2)}, v_{\sigma_4(3)}) \\ &\quad + (\text{sgn } \sigma_5) \xi(v_{\sigma_5(1)}, v_{\sigma_5(2)}, v_{\sigma_5(3)}) + (\text{sgn } \sigma_6) \xi(v_{\sigma_6(1)}, v_{\sigma_6(2)}, v_{\sigma_6(3)}) \\ &= (\text{sgn } \sigma_1) \xi(v_1, v_2, v_3) + (\text{sgn } \sigma_2) \xi(v_2, v_1, v_3) + (\text{sgn } \sigma_3) \xi(v_3, v_2, v_1) \\ &\quad + (\text{sgn } \sigma_4) \xi(v_1, v_3, v_2) + (\text{sgn } \sigma_5) \xi(v_3, v_1, v_2) + (\text{sgn } \sigma_6) \xi(v_2, v_3, v_1) \\ &= \xi(v_1, v_2, v_3) - \xi(v_2, v_1, v_3) - \xi(v_3, v_2, v_1) \\ &\quad - \xi(v_1, v_3, v_2) + \xi(v_3, v_1, v_2) + \xi(v_2, v_3, v_1). \end{aligned}$$

It can be verified that  $A_3(\xi)$  is antisymmetric.

## §8.4 Exterior Product

**Definition 8.4.1** An antisymmetric  $s$ -th order covariant tensor is also called an **exterior form of degree  $s$** , or simply an  **$s$ -form**.

**Definition 8.4.2** For any  $f \in \Lambda^k(V^*)$  and  $g \in \Lambda^s(V^*)$ , define

$$f \wedge g = \frac{1}{k!s!} A_{k+s}(f \otimes g),$$

where  $A_{k+s}$  is the antisymmetrization operator defined in Definition 8.3.1. Then  $f \wedge g$  is an exterior form of degree  $k+s$ , called the **exterior product** (or **wedge product**) of the exterior forms  $f$  and  $g$ .

**Example 8.4.1** Let  $f(v_1, v_2) \in \Lambda^2(V^*)$  and  $g(v_3) \in \Lambda^1(V^*)$ . We compute  $f \wedge g$ .

$$\begin{aligned}
f \wedge g &= \frac{1}{2!!!} A_{2+1}(f \otimes g) = \frac{1}{2} \sum_{\sigma \in G_3} \text{sgn } \sigma \cdot \sigma(f \otimes g) \\
&= \frac{1}{2} \sum_{\sigma \in G_3} \text{sgn } \sigma f(v_{\sigma(1)}, v_{\sigma(2)}) g(v_{\sigma(3)}) \\
&= \frac{1}{2} ((\text{sgn } \sigma_1) f(v_{\sigma_1(1)}, v_{\sigma_1(2)}) g(v_{\sigma_1(3)}) + (\text{sgn } \sigma_2) f(v_{\sigma_2(1)}, v_{\sigma_2(2)}) g(v_{\sigma_2(3)}) \\
&\quad + (\text{sgn } \sigma_3) f(v_{\sigma_3(1)}, v_{\sigma_3(2)}) g(v_{\sigma_3(3)}) + (\text{sgn } \sigma_4) f(v_{\sigma_4(1)}, v_{\sigma_4(2)}) g(v_{\sigma_4(3)}) \\
&\quad + (\text{sgn } \sigma_5) f(v_{\sigma_5(1)}, v_{\sigma_5(2)}) g(v_{\sigma_5(3)}) + (\text{sgn } \sigma_6) f(v_{\sigma_6(1)}, v_{\sigma_6(2)}) g(v_{\sigma_6(3)})) \\
&= \frac{1}{2} ((\text{sgn } \sigma_1) f(v_1, v_2) g(v_3) + (\text{sgn } \sigma_2) f(v_2, v_1) g(v_3) \\
&\quad + (\text{sgn } \sigma_3) f(v_3, v_2) g(v_1) + (\text{sgn } \sigma_4) f(v_1, v_3) g(v_2) \\
&\quad + (\text{sgn } \sigma_5) f(v_3, v_1) g(v_2) + (\text{sgn } \sigma_6) f(v_2, v_3) g(v_1)) \\
&= \frac{1}{2} (f(v_1, v_2) g(v_3) - f(v_2, v_1) g(v_3) - f(v_3, v_2) g(v_1) - f(v_1, v_3) g(v_2) \\
&\quad + f(v_3, v_1) g(v_2) + f(v_2, v_3) g(v_1)) \\
&= \frac{1}{2} (2f(v_1, v_2) g(v_3) - 2f(v_3, v_2) g(v_1) - 2f(v_1, v_3) g(v_2)) \\
&= f(v_1, v_2) g(v_3) - f(v_3, v_2) g(v_1) - f(v_1, v_3) g(v_2).
\end{aligned}$$

**Theorem 8.4.1** The exterior product satisfies the following algebraic properties:

Let  $\alpha, \alpha_1, \alpha_2 \in \Lambda^k(V^*)$ ,  $\beta, \beta_1, \beta_2 \in \Lambda^s(V^*)$ , and  $\gamma \in \Lambda^t(V^*)$ . Then:

- 1) Distributive Laws  $(\alpha_1 + \alpha_2) \wedge \beta = \alpha_1 \wedge \beta + \alpha_2 \wedge \beta$ ,  
 $\beta \wedge (\alpha_1 + \alpha_2) = \beta \wedge \alpha_1 + \beta \wedge \alpha_2$ ;
- 2) Associative Law  $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$ ;
- 3) Anticommutative Law  $\alpha \wedge \beta = (-1)^{k+s} \beta \wedge \alpha$ .

From the anticommutative law, if  $\alpha$  and  $\beta$  are both 1-forms,  $\alpha, \beta \in \Lambda^1(V^*)$ , then

$$\alpha \wedge \beta = -\beta \wedge \alpha, \quad \alpha \wedge \alpha = \beta \wedge \beta = 0.$$

Consequently, if a polynomial expression involving exterior products contains two identical 1-forms, that term must vanish.

The exterior product operation can also be extended to antisymmetric  $r$ -th order contravariant tensors.

## §8.5 Exterior Differentiation

### 1. Differential Forms

**Definition 8.5.1** A smooth antisymmetric covariant tensor field  $\omega$  of order  $r$  on an  $m$ -dimensional smooth manifold  $M$  is called an **exterior differential form of degree  $r$**  (or simply an  **$r$ -form**) on  $M$ .

An exterior differential form of degree  $r$  is sometimes abbreviated as an  **$r$ -form**.

The set of all  $r$ -forms on  $M$  is denoted by  $A^r(M)$ , which is a vector space. When  $r \leq m$ , its dimension is  $C_m^r = \frac{m!}{r!(m-r)!}$ .

A 1-form on  $M$  is exactly a smooth covariant tensor field of order one on  $M$ , i.e.,  $A^1(M) = T_1^0(M)$ . By convention, a 0-form on  $M$  is a smooth function on  $M$ , i.e.,  $A^0(M) = C^\infty(M)$ .

Let  $\omega$  be an  $r$ -form on a smooth manifold  $M$ , and let  $X_1, \dots, X_r$  be smooth tangent vector

fields on  $M$ . Then  $\omega$  defines a function on  $M$  given by

$$(\omega(X_1, \dots, X_r))(p) = \omega_p((X_1)_p, \dots, (X_r)_p). \quad (8.5.1)$$

$\omega$  is a multilinear function satisfying

$$\omega(X_1, \dots, hX_i, \dots, X_r) = h\omega(X_1, \dots, X_i, \dots, X_r), \quad (8.5.2)$$

where  $h$  is a function on  $M$ .

If a local coordinate system  $(U; x^i)$  is established around a point  $p$  on  $M$ , then a basis for  $A^r(M)$  is

$$\{dx^{i_1} \wedge \dots \wedge dx^{i_r}, 1 \leq i_1 < \dots < i_r \leq m\}.$$

Thus, restricted to the coordinate neighborhood  $U$ , the  $r$ -form  $\omega$  can be expressed as

$$\omega = a_{i_1 \dots i_r} dx^{i_1} \wedge \dots \wedge dx^{i_r},$$

where the coefficients  $a_{i_1 \dots i_r}$  are smooth functions on  $U$ , since  $\omega$  is a smooth tensor field.

Moreover, because  $\omega$  is antisymmetric, the coefficients  $a_{i_1 \dots i_r}$  are antisymmetric in their indices.

**Example 8.5.1** Consider  $R^3$  with Cartesian coordinates  $(x, y, z)$ .

0) For 0-forms,  $r = 0$ . The dimension of  $A^0(M)$  is equal to  $C_3^0 = \frac{3!}{0!(3-0)!} = 1$ . Let  $f$  be a smooth function on  $R^3$ , i.e.,  $f \in C^\infty(M)$ . Then  $f$  is a 0-form on  $R^3$ . A basis for  $A^0(M)$  can be taken as the constant function 1,  $f = 1 \cdot f$ .

1) For 1-forms  $\alpha$ ,  $r = 1$ . The dimension of  $A^1(M)$  is equal to  $C_3^1 = \frac{3!}{1!(3-1)!} = 3$ . A basis for  $A^1(M)$  is

$$\{dx^{i_1} \wedge \dots \wedge dx^{i_r}, 1 \leq i_1 < \dots < i_r \leq m\} = \{dx^{i_1}, 1 \leq i_1 \leq 3\} = \{dx, dy, dz\}.$$

Let  $A, B, C$  be smooth functions on  $R^3$ . Then

$$\alpha = A dx + B dy + C dz.$$

2) For 2-forms  $\beta$ ,  $r = 2$ . The dimension of  $A^2(M)$  is equal to  $C_3^2 = \frac{3!}{2!(3-2)!} = 3$ . A basis for  $A^2(M)$  is

$$\begin{aligned} & \{dx^{i_1} \wedge \dots \wedge dx^{i_r}, 1 \leq i_1 < \dots < i_r \leq m\} \\ & = \{dx^{i_1} \wedge dx^{i_2}, 1 \leq i_1 < i_2 \leq 3\} = \{dx \wedge dy, dy \wedge dz, dz \wedge dx\}. \end{aligned}$$

Let  $A, B, C$  be smooth functions on  $R^3$ . Then

$$\beta = A dx \wedge dy + B dy \wedge dz + C dz \wedge dx.$$

3) For 3-form  $\gamma$ ,  $r = 3$ . The dimension of  $A^3(M)$  is equal to  $C_3^3 = \frac{3!}{3!(3-3)!} = 1$ . A basis for  $A^3(M)$  is

$$\begin{aligned} & \{dx^{i_1} \wedge \dots \wedge dx^{i_r}, 1 \leq i_1 < \dots < i_r \leq m\} \\ & = \{dx^{i_1} \wedge dx^{i_2} \wedge dx^{i_3}, 1 \leq i_1 < i_2 < i_3 \leq 3\} = \{dx \wedge dy \wedge dz\}. \end{aligned}$$

Let  $A$  be a smooth function on  $R^3$ . Then

$$\gamma = A dx \wedge dy \wedge dz.$$

4) For 4-forms  $\theta$ ,  $r = 4$ . Since the degree exceeds the dimension of  $R^3$ ,  $r = 4 > 3 = m$ , we define  $\theta = 0$ .

We can take the direct sum of the four spaces above to form a new graded vector space  $A(M)$ :

$$A^0(M) + A^1(M) + A^2(M) + A^3(M) = A(M).$$

The dimension of  $A(M)$  is

$$1 + 3 + 3 + 1 = 8 = 2^3.$$

A basis for  $A(M)$  is

$$\{1, dx, dy, dz, dx \wedge dy, dy \wedge dz, dz \wedge dx, dx \wedge dy \wedge dz\}.$$

**Definition 8.5.2** Let  $A^r(M)$  denote the set of  $r$ -forms on an  $m$ -dimensional smooth manifold  $M$ . Define

$$A(M) = A^0(M) + A^1(M) + A^2(M) + \cdots + A^m(M).$$

Elements of  $A(M)$  are called **differential forms** (or **exterior differential forms**) on  $M$ .

The dimension of  $A(M)$  is

$$1 + C_m^1 + C_m^2 + \cdots + C_m^m = 2^m.$$

Every differential form  $\omega \in A(M)$  can be expressed as

$$\omega = \omega^0 + \omega^1 + \omega^2 + \cdots + \omega^i + \cdots + \omega^m,$$

where  $\omega^i \in A^i(M)$  is an  $i$ -form.

For elements in  $A(M)$ , operations such as addition, scalar multiplication, and the wedge product can be defined pointwise.

## 2. Wedge Product of Differential Forms

The wedge product for exterior forms can be extended to the space of differential forms  $A(M)$ . The wedge product operation can be defined in  $A(M)$  as follows. Let  $\omega_1, \omega_2 \in A(M)$ . For any point  $p \in M$  on the smooth manifold  $M$ , define

$$(\omega_1 \wedge \omega_2)(p) = \omega_1(p) \wedge \omega_2(p).$$

Here, the right-hand side denotes the wedge product of the differential forms  $\omega_1(p)$  and  $\omega_2(p)$  at the point  $p \in M$ , resulting in a new differential form  $\omega_1 \wedge \omega_2$ . The value of  $\omega_1 \wedge \omega_2$  at point  $p \in M$  is  $(\omega_1 \wedge \omega_2)(p)$ . The result  $\omega_1 \wedge \omega_2$  is again a differential form, i.e.,  $\omega_1 \wedge \omega_2 \in A(M)$ . The wedge product  $\wedge$  thus corresponds to the mapping

$$\wedge : A^r(M) \times A^s(M) \rightarrow A^{r+s}(M).$$

The degree of a differential form cannot exceed the dimension of the smooth manifold  $M$ . Therefore, we stipulate that when  $r + s > m$ ,  $A^{r+s}(M) = 0$ .

**Example 8.5.2** Consider  $R^3$  with Cartesian coordinates  $(x, y, z)$ . Let

$$\alpha = A dx + B dy + C dz, \quad \beta = D dx + E dy + F dz,$$

where  $A, B, C, D, E, F \in C^\infty(R^3)$ . We compute the wedge product  $\alpha \wedge \beta$  of these two 1-forms.

$$\begin{aligned} \alpha \wedge \beta &= (A dx + B dy + C dz) \wedge (D dx + E dy + F dz) \\ &= AD dx \wedge dx + BD dy \wedge dx + CD dz \wedge dx + AE dx \wedge dy + BE dy \wedge dy + CE dz \wedge dy \\ &\quad + AF dx \wedge dz + BF dy \wedge dz + CF dz \wedge dz \\ &= BD dy \wedge dx + CD dz \wedge dx + AE dx \wedge dy + CE dz \wedge dy + AF dx \wedge dz + BF dy \wedge dz \\ &= -BD dx \wedge dy + CD dz \wedge dx + AE dx \wedge dy - CE dy \wedge dz - AF dz \wedge dx + BF dy \wedge dz \\ &= (AE - BD) dx \wedge dy + (BF - CE) dy \wedge dz + (CD - AF) dz \wedge dx. \end{aligned}$$

## 3. Exterior Differentiation of Differential Forms

**Theorem 8.5.1** Let  $M$  be an  $m$ -dimensional smooth manifold. Then there exists a unique mapping  $d : A(M) \rightarrow A(M)$  such that  $d(A^r(M)) \subset A^{r+1}(M)$ , and it satisfies the following conditions:

1)  $d$  is **linear**, i.e., for any  $\omega_1, \omega_2 \in A(M), a \in R$ ,

$$d(\omega_1 + a\omega_2) = d\omega_1 + ad\omega_2;$$

2) If  $\omega_1$  is an  $r$ -form (i.e.,  $\omega_1 \in A^r(M)$ ) and  $\omega_2$  is any differential form, then

$$d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^r \omega_1 \wedge d\omega_2;$$

3) If  $f$  is a smooth function on  $M$  (i.e.,  $f \in A^0(M)$ ), then  $df$  is the differential of  $f$ .

4) For any differential form  $f \in A^0(M)$ ,  $dd(f) = 0$ .

The mapping  $d$  determined by the above conditions is called the **exterior derivative**.

**Example 8.5.3** Assume a Cartesian coordinate system  $(x, y, z)$  is established in  $R^3$ , and let  $A, B$ , and  $C$  be smooth functions on  $R^3$ .

1) If  $f$  is a smooth function on  $R^3$ , then  $df$  is the total differential of  $f$ , so

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

The coefficients of  $df$  form the vector  $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$ , which is the **gradient** of  $f$ ,  $\text{grad}f$ .

2) Let  $\omega = A dx + B dy + C dz$ . Then

$$\begin{aligned} d\omega &= dA \wedge dx + dB \wedge dy + dC \wedge dz \\ &= \left(\frac{\partial A}{\partial x} dx + \frac{\partial A}{\partial y} dy + \frac{\partial A}{\partial z} dz\right) \wedge dx + \left(\frac{\partial B}{\partial x} dx + \frac{\partial B}{\partial y} dy + \frac{\partial B}{\partial z} dz\right) \wedge dy \\ &\quad + \left(\frac{\partial C}{\partial x} dx + \frac{\partial C}{\partial y} dy + \frac{\partial C}{\partial z} dz\right) \wedge dz \\ &= \frac{\partial A}{\partial y} dy \wedge dx + \frac{\partial A}{\partial z} dz \wedge dx + \frac{\partial B}{\partial x} dx \wedge dy + \frac{\partial B}{\partial z} dz \wedge dy \\ &\quad + \frac{\partial C}{\partial x} dx \wedge dz + \frac{\partial C}{\partial y} dy \wedge dz \\ &= -\frac{\partial A}{\partial y} dx \wedge dy + \frac{\partial A}{\partial z} dz \wedge dx + \frac{\partial B}{\partial x} dx \wedge dy - \frac{\partial B}{\partial z} dy \wedge dz \\ &\quad - \frac{\partial C}{\partial x} dz \wedge dx + \frac{\partial C}{\partial y} dy \wedge dz \\ &= \left(\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}\right) dy \wedge dz + \left(\frac{\partial A}{\partial z} - \frac{\partial C}{\partial x}\right) dz \wedge dx + \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) dx \wedge dy. \end{aligned}$$

If  $A, B, C$  are the components of a vector field  $X$ , then the coefficients of  $d\omega$  form the vector

$$\left(\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}, \frac{\partial A}{\partial z} - \frac{\partial C}{\partial x}, \frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)$$

which is the **curl** of  $X$ ,  $\text{curl}X$ .

3) Let  $\omega = A dy \wedge dz + B dz \wedge dx + C dx \wedge dy$ . Then

$$\begin{aligned} d\omega &= dA \wedge dy \wedge dz + dB \wedge dz \wedge dx + dC \wedge dx \wedge dy \\ &= \left(\frac{\partial A}{\partial x} dx + \frac{\partial A}{\partial y} dy + \frac{\partial A}{\partial z} dz\right) \wedge dy \wedge dz + \left(\frac{\partial B}{\partial x} dx + \frac{\partial B}{\partial y} dy + \frac{\partial B}{\partial z} dz\right) \wedge dz \wedge dx \\ &\quad + \left(\frac{\partial C}{\partial x} dx + \frac{\partial C}{\partial y} dy + \frac{\partial C}{\partial z} dz\right) \wedge dx \wedge dy \\ &= \frac{\partial A}{\partial x} dx \wedge dy \wedge dz + \frac{\partial B}{\partial y} dy \wedge dz \wedge dx + \frac{\partial C}{\partial z} dz \wedge dx \wedge dy \\ &= \frac{\partial A}{\partial x} dx \wedge dy \wedge dz + \frac{\partial B}{\partial y} dx \wedge dy \wedge dz + \frac{\partial C}{\partial z} dx \wedge dy \wedge dz \\ &= \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right) dx \wedge dy \wedge dz \\ &= \text{div}X dx \wedge dy \wedge dz. \end{aligned}$$

The coefficient of  $d\omega$  is the divergence of the vector field  $X = (A, B, C)$ ,  $\text{div}X$ .

**Theorem 8.5.2 (Poincaré Lemma)**  $d^2 = dd = 0$ , that is, for any differential form  $\omega$ ,  $dd(\omega) = 0$ .

#### 4. Pullback of Differential Forms

As previously noted, a smooth map  $f: M \rightarrow N$  between smooth manifolds induces the tangent map  $f_*$ , which maps a tangent vector  $X_p$  at  $p \in M$  to a tangent vector  $f_*(X_p)$  at  $f(p) \in N$ . However,  $f_*$  does not necessarily carry smooth vector fields on  $M$  to smooth vector fields on  $N$ ; that is, in general,  $f_*X$  may not produce a smooth vector field on  $N$ . For differential forms, we have the following theorem.

**Theorem 8.5.3** Let  $f: M \rightarrow N$  be a smooth map between smooth manifolds. Then  $f$  induces a linear map between the spaces of differential forms:

$$f^*: A(N) \rightarrow A(M),$$

and this map  $f^*$  commutes with the wedge product, i.e., for any  $\omega, \eta \in A(N)$ ,

$$f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta.$$

Let  $(U; x^i)$  be a local coordinate system of the  $m$ -dimensional smooth manifold  $M$  at the point  $p$ , and let  $(V; y^i)$  be a local coordinate system of the  $n$ -dimensional smooth manifold  $N$  at the point  $f(p) \in N$  and  $f(U) \subset V$ . Then the local coordinate expression of the map  $f$  is

$$y^j = f^j(x^1, \dots, x^m) \in C^\infty(U), \quad 1 \leq j \leq n.$$

Let an  $r$ -th order exterior differential form  $\omega$  be expressed on the coordinate domain  $V$  as

$$\omega|_V = a_{i_1 \dots i_r} dy^{i_1} \wedge \dots \wedge dy^{i_r},$$

where  $a_{i_1 \dots i_r}$  are smooth functions on  $V$ . Then

$$f^*\omega|_U = \tilde{a}_{i_1 \dots i_r} dx^{i_1} \wedge \dots \wedge dx^{i_r},$$

and

$$\tilde{a}_{i_1 \dots i_r} = (a_{i_1 \dots i_r} \circ f) \frac{\partial f^{j_1}}{\partial x^{i_1}} \dots \frac{\partial f^{j_r}}{\partial x^{i_r}}.$$

**Theorem 8.5.4** Let  $f: M \rightarrow N$  be a smooth map between smooth manifolds. Then the induced map  $f^*: A(N) \rightarrow A(M)$  on differential forms commutes with the exterior derivative  $d$ , i.e.,

$$f^* \circ d = d \circ f^*: A(N) \rightarrow A(M).$$

Let  $S \subset M$  be an immersed submanifold of  $M$ , and let  $i: S \rightarrow M$  be the inclusion map,  $p \in S$ ,  $v \in T_p S$ . If  $\omega$  is a 1-form on  $M$ , then the differential form obtained by restricting  $\omega$  to the submanifold  $S$ , denoted by  $\omega|_S$ , is defined by

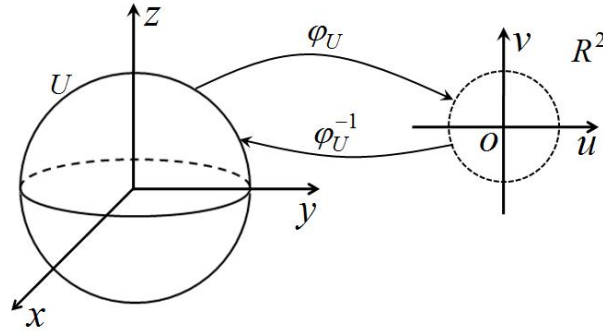
$$(\omega|_S)_p(v) = \omega_p(v).$$

**Theorem 8.5.5** Let  $i: S \rightarrow M$  be the inclusion map for the submanifold  $S$ , and let  $\omega$  be a 1-form on  $M$ . Then  $i^*\omega = \omega|_S$ .

The theorem states that the 1-form  $i^*\omega$  on the submanifold  $S$  obtained by pulling back a 1-form  $\omega$  from  $M$  to  $S$  is the same as the 1-form  $\omega|_S$  on  $S$  obtained by restricting the original 1-form  $\omega$  on  $M$  to  $S$ .

**Example 8.5.4** As shown in Figure 8.5.1, let  $S^2 = \{(x, y, z) \in R^3, x^2 + y^2 + z^2 = 1\}$  be the unit sphere in  $R^3$ . Let  $i: S^2 \rightarrow R^3$  be the inclusion map. Let  $dx, dy, dz$  be the standard 1-forms on  $R^3$ ,  $dx \wedge dy$  and  $dy \wedge dz$  be 2-forms on  $R^3$ , and  $dx \wedge dy \wedge dz$  be a 3-form on  $R^3$ . The forms obtained by the pull-back  $i^*$ , such as  $i^*dx, i^*dx \wedge dy, i^*dx \wedge dy \wedge dz$ , etc., are differential forms on  $S^2$ . We now express these exterior forms in local coordinates of  $S^2$ .





**Figure 8.5.1** Expressing the exterior forms on  $S^2$  using local coordinates on  $S^2$

Take a coordinate covering  $(U, \varphi_U)$  of  $S^2$ :

$$U = \{(x, y, z) \in S^2 : z > 0\}, \quad \varphi_U(x, y, z) = (x, y), \quad (x, y, z) \in U.$$

As shown in Figure 8.5.1, change the Cartesian coordinates  $(x, y)$  in  $R^2$  to  $(u, v)$ , i.e., set

$$u = x, \quad v = y,$$

then  $(u, v)$  become local coordinates on  $S^2$ . The inclusion map  $i: S^2 \rightarrow R^3$  becomes

$$i(u, v) = (u, v, \sqrt{1 - u^2 - v^2}).$$

The pull-back  $i^*$  induced by the inclusion map  $i$  maps 1-forms on  $R^3$  to 1-forms on  $S^2$ :

$$i^* dx = du, \quad i^* dy = dv, \quad i^* dz = d\sqrt{1 - u^2 - v^2} = -\frac{udu + vdv}{\sqrt{1 - u^2 - v^2}}.$$

According to Theorem 8.5.3, we compute the pull-back of the 2-forms on  $R^3$  to  $S^2$ :

$$i^*(dx \wedge dy) = i^* dx \wedge i^* dy = du \wedge dv,$$

$$i^*(dx \wedge dz) = i^* dx \wedge i^* dz = du \wedge \left( -\frac{udu + vdv}{\sqrt{1 - u^2 - v^2}} \right) = -\frac{vdu \wedge dv}{\sqrt{1 - u^2 - v^2}}.$$

Similarly, the pull-back of the 3-form on  $R^3$  to  $S^2$  is

$$\begin{aligned} i^*(dx \wedge dy \wedge dz) &= i^* dx \wedge i^* dy \wedge i^* dz \\ &= du \wedge dv \wedge \left( -\frac{udu + vdv}{\sqrt{1 - u^2 - v^2}} \right) = \frac{-du \wedge dv \wedge udu - du \wedge dv \wedge vdv}{\sqrt{1 - u^2 - v^2}} = 0. \end{aligned}$$

The 3-form vanishes on  $S^2$  because its degree (3) exceeds the dimension of  $S^2$  (which is 2).

Now consider the restriction of the 1-form  $dz$  on  $R^3$  to  $S^2$ , denoted by  $dz|_{S^2}$ . On  $S^2$ , the coordinate  $z$  satisfies  $x^2 + y^2 + z^2 = 1$ , i.e.,  $dz$  is not independent. For the upper hemisphere ( $z > 0$ ), we have  $z = \sqrt{1 - x^2 - y^2}$ . Substituting this into  $dz|_{S^2}$  gives

$$dz|_{S^2} = d(\sqrt{1 - x^2 - y^2})|_{S^2} = -\frac{1}{\sqrt{1 - x^2 - y^2}}(xdx + ydy)|_{S^2}.$$

Changing coordinates  $(x, y)$  to the local coordinates  $(u, v)$  on  $S^2$ , we obtain

$$dz|_{S^2} = -\frac{1}{\sqrt{1 - u^2 - v^2}}(udu + vdv)|_{S^2},$$

which coincides with  $i^* dz$ . Therefore, the 1-form  $i^* dz$  obtained by pulling back  $dz$  on  $R^3$  to the submanifold  $S^2$  is identical to the 1-form  $dz|_{S^2}$  obtained by restricting  $dz$  to  $S^2$ . The same holds for exterior forms of other degrees.

## 5. Closed Differential Forms and Exact Differential Forms

**Definition 8.5.3** Let  $\omega$  be an element of the space of  $r$ -forms  $A^r(M)$ . Define

$$Z^r(M, R) = \{\omega \mid \omega \in A^r(M), \text{ and } d\omega = 0\},$$

then  $Z^r(M, R)$  is called the **group of  $r$ -cocycles**, and its elements are called **closed  $r$ -forms**.

**Definition 8.5.4** Let  $\omega$  be an element of the space of  $r$ -forms  $A^r(M)$ . Define

$$B^r(M, R) = \{\omega \mid \omega \in A^r(M), \exists \beta \in A^{r-1}(M), \text{ such that } d\beta = \omega\},$$

then  $B^r(M, R)$  is called the **group of  $r$ -coboundaries**, and its elements are called **exact  $r$ -forms**.

Since  $d(d\beta) = d\omega = 0$ , it follows that  $B^r(M, R)$  is a subgroup of  $Z^r(M, R)$ . Hence, every exact  $r$ -form is closed. However, a closed  $r$ -form is not necessarily exact.

**Definition 8.5.5** The quotient group

$$H^r(M, R) = Z^r(M, R) / B^r(M, R)$$

is called the  **$r$ -th de Rham cohomology group** of the smooth manifold  $M$ .

**Theorem 8.5.6 (de Rham Theorem)** If the smooth manifold  $M$  is compact, then the de Rham cohomology group is isomorphic to the  $r$ -th singular cohomology group of  $M$ .

## §8.6 Interior Product

The operation that combines tensor (field) multiplication with contraction is called the **inner product** of tensors (fields). Here we introduce a specific multiplicative operation between a smooth tangent vector field  $X$  and an  $r$ -form  $\omega$  on a smooth manifold  $M$ —the **interior product**.

**Definition 8.6.1** Let  $X$  be a smooth tangent vector field on a smooth manifold  $M$  and let  $\omega \in A^r(M)$ . Define a linear operator

$$i_X : A^r(M) \rightarrow A^{r-1}(M)$$

satisfying the following conditions. This operator  $i_X$  is called the **interior multiplication** (or **interior product**) of  $X$  with  $\omega$ :

1) If  $r = 0$  and  $\omega = f \in A^0(M)$ , then  $i_X$  acts on  $A^0(M)$  as the zero map, i.e.,

$$i_X f = 0; \quad (8.6.1)$$

2) If  $r = 1$  and  $\omega \in A^1(M)$ , set

$$i_X \omega = \langle X, \omega \rangle; \quad (8.6.2)$$

3) If  $r > 1$ , then for any  $r-1$  smooth tangent vector fields  $Y_1, \dots, Y_{r-1}$ , set

$$i_X \omega(Y_1, \dots, Y_{r-1}) = \omega(X, Y_1, \dots, Y_{r-1}). \quad (8.6.3)$$

Sometimes we also write

$$X \lrcorner \omega = i_X \omega. \quad (8.6.4)$$

The notation  $X \lrcorner \omega$  is read as “ $X$  inserted into  $\omega$ .”

**Theorem 8.6.1** Let  $\alpha^1, \dots, \alpha^k$  all be 1-forms on the smooth manifold  $M$ , and let  $X$  be a smooth tangent vector field on  $M$ . Then

$$i_X(\alpha^1 \wedge \dots \wedge \alpha^k) = \sum_{i=1}^k (-1)^{i-1} \alpha^i(X) \alpha^1 \wedge \dots \wedge \hat{\alpha}^i \wedge \dots \wedge \alpha^k, \quad (8.6.5)$$

where the caret  $\hat{\phantom{x}}$  over  $\alpha^i$  indicates that  $\alpha^i$  is omitted from the wedge product.

**Example 8.6.1** Let  $\alpha^1 = dx$ ,  $\alpha^2 = dy$ ,  $\alpha^3 = dz$  be the standard 1-forms on  $R^3$ , and let  $X = \frac{\partial}{\partial z}$  be a smooth tangent vector field on  $R^3$ . Applying formula (8.6.5) we obtain

$$i_X(\alpha^1 \wedge \alpha^2 \wedge \alpha^3) = \sum_{i=1}^3 (-1)^{i-1} \alpha^i(X) \alpha^1 \wedge \dots \wedge \hat{\alpha}^i \wedge \dots \wedge \alpha^3$$

$$\begin{aligned}
&= (-1)^0 \alpha^1(X) \alpha^2 \wedge \alpha^3 + (-1)^1 \alpha^2(X) \alpha^1 \wedge \alpha^3 + (-1)^2 \alpha^3(X) \alpha^1 \wedge \alpha^2 \\
&= dx \left( \frac{\partial}{\partial z} \right) dy \wedge dz - dy \left( \frac{\partial}{\partial z} \right) dx \wedge dz + dz \left( \frac{\partial}{\partial z} \right) dx \wedge dy \\
&= dx \wedge dy, \\
&\therefore dx \left( \frac{\partial}{\partial z} \right) = \frac{\partial x}{\partial z} = 0, \quad dy \left( \frac{\partial}{\partial z} \right) = \frac{\partial y}{\partial z} = 0, \quad dz \left( \frac{\partial}{\partial z} \right) = \frac{\partial z}{\partial z} = 1.
\end{aligned}$$

**Theorem 8.6.2** Let  $X$  be a smooth tangent vector field on the smooth manifold  $M$ . Then

1)  $i_X \circ i_X = 0$  ;

2) If  $\omega \in A^k(M)$  and  $\eta \in A^l(M)$ , then

$$i_X(\omega \wedge \eta) = (i_X \omega) \wedge \eta + (-1)^k \omega \wedge (i_X \eta). \quad (8.6.6)$$

**Example 8.6.2** Let  $X = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$  be a smooth tangent vector field on  $R^2$ , and let

$\alpha = dx \wedge dy$  be a 2-form on  $R^2$ . We compute  $i_X \alpha$ .

First compute  $i_X dx$ , and  $i_X dy$ :

$$\begin{aligned}
i_X dx &= dx(X) = dx \left( x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) = x \frac{\partial x}{\partial x} + y \frac{\partial x}{\partial y} = x, \\
i_X dy &= dy(X) = dy \left( x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) = x \frac{\partial y}{\partial x} + y \frac{\partial y}{\partial y} = y.
\end{aligned}$$

Then, using (8.6.6),

$$i_X \alpha = i_X(dx \wedge dy) = (i_X dx) \wedge dy + (-1)^1 dx \wedge (i_X dy) = x dy - y dx.$$

One can also verify directly that

$$i_X \circ i_X(dx) = i_X(dx(X)) = i_X(x) = 0,$$

where the last equality uses the first rule in Definition 8.6.1.

**Example 8.6.3** We use formula (8.6.6) to compute  $i_V(dx \wedge dy \wedge dz)$ , where  $V = \frac{\partial}{\partial z}$ .

$$\begin{aligned}
&i_V(dx \wedge dy \wedge dz) \\
&= (i_V dx) \wedge dy \wedge dz + (-1)^1 dx \wedge (i_V(dy \wedge dz)) \\
&= (i_V dx) \wedge dy \wedge dz - dx \wedge ((i_V dy) \wedge dz + (-1)^1 dy \wedge (i_V dz)) \\
&= (dx(V)) \wedge dy \wedge dz - dx \wedge (dy(V)) \wedge dz + dx \wedge dy \wedge (dz(V)) \\
&= \frac{\partial x}{\partial z} \wedge dy \wedge dz - dx \wedge \frac{\partial y}{\partial z} \wedge dz + dx \wedge dy \wedge \frac{\partial z}{\partial z} \\
&= dx \wedge dy.
\end{aligned}$$

## Chapter 9 Orientation and Integration on Smooth Manifolds

### §9.1 Conditions for Orientability

#### 1. Orientation of a Vector Space

Orientation means fixing a sense of direction. For example, on the real number line we conventionally take the right side as the positive direction and the left side as the negative direction. In the plane, we usually define counterclockwise rotation as the positive direction and clockwise rotation as the opposite direction. In  $R^m$  ( $m \geq 3$ ), choosing an orientation is equivalent to choosing either a **right-hand rule** (right-handed screw rule or right-handed system) or a **left-hand rule** (left-handed screw rule or left-handed system). For instance, as shown in Figure 9.1.1, in  $R^3$  choosing an orientation means selecting either the right-hand rule or the left-hand rule.

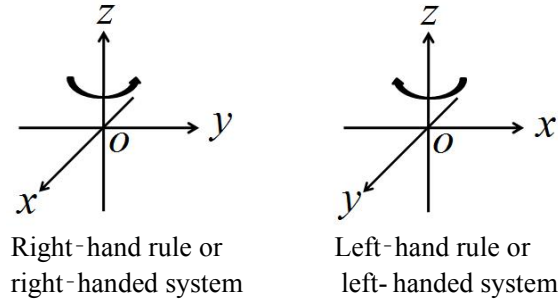


Figure 9.1.1 Right-hand and left-hand rules

Generally, let  $V$  be an  $m$ -dimensional vector space, and let  $\{e_i\}$  and  $\{\delta_i\}$  be two bases of  $V$ . Their transformation relation is

$$e_i = \sum_{j=1}^m a_i^j \delta_j,$$

Where  $\det(a_i^j) \neq 0$ . We stipulate that if  $\det(a_i^j) > 0$ , then the bases  $\{e_i\}$  and  $\{\delta_i\}$  are called **equivalent**, denoted by  $\{e_i\} \sim \{\delta_i\}$ . This relation  $\sim$  defines an equivalence relation among all bases of  $V$ . An equivalence class of bases of  $V$  under the relation  $\sim$  is called an **orientation** of  $V$ . A vector space has exactly two distinct orientations. Two bases belonging to the same equivalence class are said to have the **same orientation**. If the determinant of the transformation matrix relating  $\{e_i\}$  and  $\{\delta_i\}$  is positive,  $\det(a_i^j) > 0$ , then they have the same orientation. Two bases belonging to different equivalence classes are said to have **opposite orientations**. If the determinant of their transformation matrix is negative,  $\det(a_i^j) < 0$ , then they belong to different equivalence classes and have opposite orientations.

#### 2. Conditions for Orientability

The concept of orientation on a smooth manifold is similar to that in  $R^m$ , but due to the diversity of smooth manifold types, there are still differences between the orientation on a smooth manifold and that in  $R^m$ . First, it should be noted that not all smooth manifolds are orientable. Therefore, it is necessary to understand what kind of smooth manifolds can be oriented.

**Definition 9.1.1** Let  $M$  be an  $m$ -dimensional smooth manifold. If there exists a compatible collection of local coordinate systems  $\Sigma = \{(U; x^i), (V; y^i), (W; z^i), \dots\}$ ,  $1 \leq i \leq m$  on  $M$  satisfying the following two conditions:

1)  $\{U, V, W, \dots\}$  forms an open cover of  $M$ ,

2) For any two local coordinate systems  $(U; x^i)$  and  $(V; y^i)$ , when  $U \cap V \neq \emptyset$ , the Jacobian determinant of their coordinate transformation on  $U \cap V$  satisfies

$$\frac{\partial(x^1, \dots, x^m)}{\partial(y^1, \dots, y^m)} > 0, \quad (9.1.1)$$

then  $M$  is called an **orientable smooth manifold**.

Two local coordinate systems  $(U; x^i)$  and  $(V; y^i)$  satisfying (9.1.1) are said to be **orientation-compatible**. Since the definition requires that the two local coordinate systems be arbitrary, all local coordinate systems in the collection  $\Sigma$  are orientation-compatible. This indicates that an orientable smooth manifold is precisely a smooth manifold that possesses an open cover consisting of orientation-compatible admissible coordinate domains.

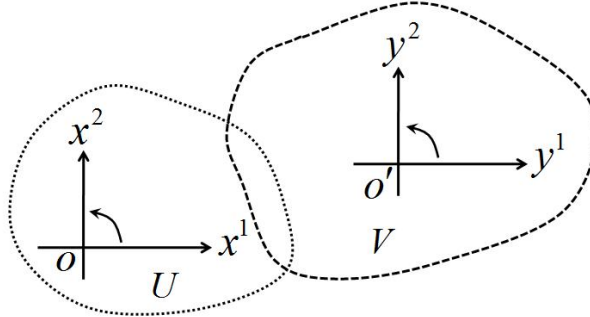
**Example 9.1.1** As shown in Figure 9.1.2, there are two open sets  $U, V$  on  $R^2$ . Let the coordinate system on the open set  $U$  around point  $o$  be  $x^1 o x^2$ , and the coordinate system on the open set  $V$  around point  $o'$  be  $y^1 o' y^2$ . Their transformation relation is

$$\begin{cases} x^1 = y^1 + 16, \\ x^2 = y^2 + 2. \end{cases}$$

The Jacobian determinant is

$$|J| = \begin{vmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^1}{\partial y^2} \\ \frac{\partial x^2}{\partial y^1} & \frac{\partial x^2}{\partial y^2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 > 0,$$

so these two open sets are **orientation-compatible** and share the same orientation. Hence,  $R^2$  is **orientable**.



**Figure 9.1.2** The open sets  $U$  and  $V$  are orientation-compatible

**Theorem 9.1.1** Assume that an  $m$ -dimensional smooth manifold  $M$  satisfies the second axiom of countability. Then  $M$  is orientable **if and only if** there exists an everywhere non-zero  $m$ -form on  $M$ .

We have already assumed that all smooth manifolds mentioned in this book satisfy the second axiom of countability.

**Theorem 9.1.2** If a smooth manifold can be covered by a single coordinate chart (i.e., a single open set), then the manifold is orientable.

Such a manifold has only one coordinate chart, hence only one coordinate system, and is therefore orientable—for example,  $R^m$  space.

**Theorem 9.1.3** Let  $M$  be an  $m$ -dimensional smooth manifold. Fix an orientation in the tangent space  $T_p M$  at a point  $p \in M$ . If, when propagating this orientation along any closed path based at  $p$ , the orientation of  $T_p M$  returns unchanged upon arriving back at  $p$ , then  $M$  is orientable.

**Theorem 9.1.4** Let  $M$  be an  $m$ -dimensional smooth manifold. Fix an orientation in the

tangent space  $T_p M$  at a point  $p \in M$ . If, when propagating this orientation along any closed path based at  $p$ , the orientation of  $T_p M$  becomes opposite to the original one upon returning to  $p$ , then  $M$  is non-orientable.

**Example 9.1.2** 1) The unit sphere  $S^m$  is orientable.

2) The projective space  $RP^m$  is orientable when  $m$  is odd, and non-orientable when  $m$  is even.

3) Every compact connected  $m$ -dimensional submanifold of  $R^{m+1}$  is orientable.

4) The Möbius strip and the Klein bottle are non-orientable.

## §9.2 Methods of Orientation

### 1. Orientation via Coordinate Systems

**Definition 9.2.1** Let  $M$  be an orientable  $m$ -dimensional smooth manifold. If

$$\Sigma = \{(U; x^i), (V; y^i), (W; z^i), \dots\}$$

is a compatible collection of local coordinate systems satisfying Definition 9.1.1 and is **maximal**—that is, for any compatible local coordinate system  $(\hat{U}; \hat{x}^i)$ , whenever it is orientation-compatible with every coordinate system belonging to  $\Sigma$ , it follows that  $(\hat{U}; \hat{x}^i) \in \Sigma$ —then  $\Sigma$  is called an **orientation** of  $M$ .

For a manifold  $M$  that has already been assigned an orientation, we denote by  $-M$  the same manifold equipped with the opposite orientation. If the collection of oriented local coordinate systems

$$\Sigma = \{(U; x^1, x^2, \dots, x^m), (V; y^1, y^2, \dots, y^m), (W; z^1, z^2, \dots, z^m), \dots\}$$

determines the orientation of  $M$ , then

$$\Sigma = \{(U; -x^1, x^2, \dots, x^m), (V; -y^1, y^2, \dots, y^m), (W; -z^1, z^2, \dots, z^m), \dots\}$$

determines the orientation of  $-M$ .

A smooth manifold together with a specified orientation is called an **oriented smooth manifold**.

**Example 9.2.1** Consider an orientable two-dimensional smooth manifold  $M$ . Suppose the orientation in the local coordinate system  $(U; x^i)$  is taken as the right-hand rule, similar to that shown in Figure 9.1.1. Now consider another local coordinate system  $(V; y^i)$ , with  $U \cap V \neq \emptyset$ . To make the orientations of these two coordinate systems the same, the Jacobian determinant of their coordinate transformation must be positive. Assume the transformation relation is

$$\begin{cases} x^1 = y^1 \cos y^2, \\ x^2 = y^1 \sin y^2, \end{cases}$$

with Jacobian determinant

$$\begin{vmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^1}{\partial y^2} \\ \frac{\partial x^2}{\partial y^1} & \frac{\partial x^2}{\partial y^2} \end{vmatrix} = \begin{vmatrix} \cos y^2 & -y^1 \sin y^2 \\ \sin y^2 & y^1 \cos y^2 \end{vmatrix} = y^1.$$

If  $y^1 > 0$ , then this transformation preserves the orientation consistency between the two local coordinate systems. If  $y^1 < 0$ , then the transformation causes the orientations of the two coordinate systems  $(U; x^i)$  and  $(V; y^i)$  to be opposite.

### 2. Orientation by an $m$ -Form

Assume that the smooth manifold  $M$  is orientable. According to Theorem 9.1.1, there exists

an everywhere non-zero  $m$ -form  $\omega$  on  $M$ . Let  $(U; x^i)$  be a local coordinate system on  $M$ . Since the space of  $m$ -forms is one-dimensional over  $U$ , we may write

$$\omega|_U = f \, dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m,$$

where  $f$  is a continuous function on  $U$  that is nowhere zero, and we assume  $f > 0$  on  $U$ .

Now suppose there is another  $m$ -form  $\omega'$  on  $M$ ,

$$\omega'|_U = g \, dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m,$$

where  $g$  is also a continuous function on  $M$  that is nowhere zero, and we assume  $g > 0$  on  $M$ . Since

$$\omega|_U = h\omega'|_U,$$

and the ratio  $h$  is a nowhere-zero continuous function on  $M$  with  $h > 0$ , the forms  $\omega$  and  $\omega'$  determine the **same orientation** of  $M$ .

However, consider another  $m$ -form  $\psi$  whose local expression uses a different ordering of the basis elements, for example

$$\psi|_U = f \, dx^2 \wedge dx^1 \wedge dx^3 \cdots \wedge dx^m.$$

We have

$$dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m = -dx^2 \wedge dx^1 \wedge dx^3 \cdots \wedge dx^m.$$

Thus

$$\psi|_U = -\omega|_U.$$

Therefore,  $\psi$  and  $\omega$  determine **opposite orientations** of  $M$ .

From the above discussion, it can be seen that if  $f = 1$ , then the basis of the  $m$ -form  $\omega$ ,

$$dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m$$

determines one orientation of the manifold  $M$ . If the basis of the  $m$ -form  $\omega$  is

$$dx^2 \wedge dx^1 \wedge dx^3 \cdots \wedge dx^m,$$

that determines the opposite orientation of manifold  $M$ .

The bases  $dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m$  and  $dx^2 \wedge dx^3 \wedge dx^1 \cdots \wedge dx^m$  naturally determine the same orientation of the manifold  $M$ , because they are identical:

$$\begin{aligned} dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m &= -dx^2 \wedge dx^1 \wedge dx^3 \cdots \wedge dx^m \\ &= dx^2 \wedge dx^3 \wedge dx^1 \cdots \wedge dx^m. \end{aligned}$$

For two bases expressed in different local coordinate systems, say  $dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m$  and  $dy^1 \wedge dy^2 \wedge dy^3 \cdots \wedge dy^m$  to determine the same orientation of the manifold  $M$ , they must satisfy the relation:

$$dx^1 \wedge dx^2 \wedge dx^3 \cdots \wedge dx^m = |J| \, dy^1 \wedge dy^2 \wedge dy^3 \cdots \wedge dy^m,$$

where  $|J|$  must be an everywhere positive function. Here,  $|J|$  is precisely the Jacobian determinant of the coordinate transformation between the two local coordinate systems. Therefore, it is required that

$$\frac{\partial(x^1, \dots, x^m)}{\partial(y^1, \dots, y^m)} > 0.$$

**Example 9.2.2** As shown in Figure 9.2.1, let the coordinate system on the open set  $U$  around point  $o$  in  $R^3$  be  $o - x^1 x^2 x^3$ , and the coordinate system on the open set  $V$  around point  $o'$  be  $o' - y^2 y^1 y^3$ . Their transformation relation is

$$\begin{cases} x^1 = y^2 + 16, \\ x^2 = y^1 + 2, \\ x^3 = y^3. \end{cases}$$

The Jacobian determinant is

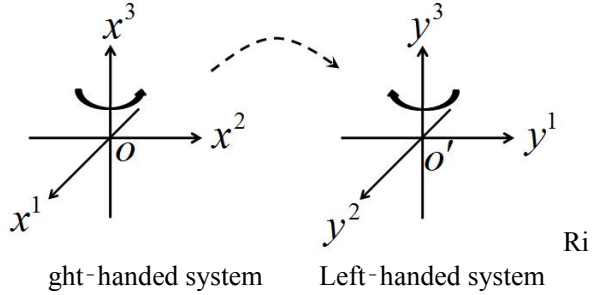
$$\begin{vmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^1}{\partial y^2} & \frac{\partial x^1}{\partial y^3} \\ \frac{\partial x^2}{\partial y^1} & \frac{\partial x^2}{\partial y^2} & \frac{\partial x^2}{\partial y^3} \\ \frac{\partial x^3}{\partial y^1} & \frac{\partial x^3}{\partial y^2} & \frac{\partial x^3}{\partial y^3} \end{vmatrix} = \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -1 < 0,$$

so the orientations of these two open sets are opposite; they belong to different orientations.

In Figure 9.2.1, the right-handed orientation corresponds to a counter-clockwise direction or the right-hand rule, which can be determined by the basis of the 3-form  $dx^1 \wedge dx^2 \wedge dx^3$ . The left-handed orientation corresponds to a clockwise direction or the left-hand rule, which can be determined by the basis of the 3-form  $dy^2 \wedge dy^1 \wedge dy^3$ . Clearly,

$$dx^1 \wedge dx^2 \wedge dx^3 = -dy^2 \wedge dy^1 \wedge dy^3,$$

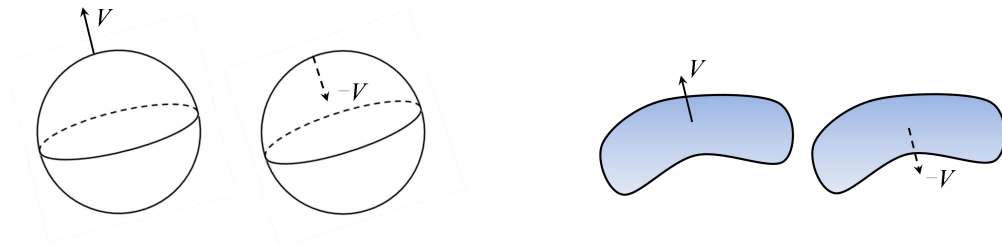
so these two orientations are opposite.



**Figure 9.2.1** Orientations in  $R^3$

### 3.Orientation of Hypersurfaces

**Theorem 9.2.1** Let  $M$  be an  $m$ -dimensional submanifold of  $R^{m+1}$ . If  $M$  admits a nowhere-zero normal vector field, then  $M$  is orientable.



**Figure 9.2.2** Determining the orientation of a surface by its normal vector

Hypersurfaces in  $R^{m+1}$ , such as the sphere  $S^k$  ( $k \leq m$ ) and the torus  $T^k$  ( $k \leq m$ ), all admit a nowhere-zero normal vector field and are therefore orientable. For a hypersurface, its orientation can be specified by its normal vector. As shown in Figure 9.2.2, if we take the outward normal vector  $V$  of the sphere  $S^2$  as its positive orientation, then the inward normal vector  $-V$  determines the opposite (negative) orientation of the sphere. Similarly, for the two surfaces on the right side of Figure 9.2.2, if we define the side toward which the normal vector points as the positive direction, then the side opposite to it corresponds to the negative orientation.



### §9.3 Types of Orientations

As shown in Figure 9.3.1, a manifold formed by the disjoint union of two orientable manifolds is **disconnected**, but still orientable. Such a manifold admits exactly **four** distinct orientations.



**Figure 9.3.1** A disconnected orientable manifold composed of two disjoint parts admits four different orientations

However, a **connected** orientable smooth manifold  $M$  has only **two** distinct orientations—for example, the plane  $R^2$ . This is because the space of  $m$ -forms on an  $m$ -dimensional smooth manifold  $M$  is one-dimensional. If two  $m$ -forms  $\omega$  and  $\omega'$  are both used to specify an orientation on  $M$ , then, since they are one-dimensional, there must exist a relation on  $M$ :

$$\omega' = f\omega,$$

where  $f$  is a continuous function that is nowhere zero on  $M$ . Because  $M$  is connected, the function  $f$  must have the same sign everywhere—either  $f > 0$  everywhere or  $f < 0$  everywhere. If  $f > 0$ , then  $\omega'$  determines the same orientation as  $\omega$ . If  $f < 0$ , then  $\omega'$  determines the orientation opposite to that of  $\omega$ , or equivalently, the same orientation as  $-\omega$ .

### §9.4 Preservation and Reversal of Orientation

**Definition 9.4.1** Let  $M$  and  $N$  be orientable smooth manifolds, and let  $f: M \rightarrow N$  be a smooth diffeomorphism. If  $f$  maps the orientation of  $M$  to that of  $N$ —i.e., for any point  $p \in M$  and any basis  $\{e_i\}$  of  $T_p M$  compatible with the orientation of  $M$ , the pushed-forward basis  $\{f_*(e_i)\}$  is always compatible with the orientation of  $T_{f(p)} N$ —then  $f$  is said to **preserve orientation** (or be **orientation-preserving**). Conversely, if the pushed-forward basis is opposite to the orientation of  $T_{f(p)} N$ , then  $f$  is said to **reverse orientation**.

**Theorem 9.4.1** A smooth diffeomorphism  $f: M \rightarrow N$  preserves orientation **if and only if** its Jacobian determinant is strictly positive in any pair of orientation-compatible local coordinate systems on  $M$  and  $N$ .

**Theorem 9.4.2** Let  $M$  and  $N$  be orientable smooth manifolds, and let  $f: M \rightarrow N$  be a local diffeomorphism. If  $N$  is already oriented, then  $M$  acquires a unique orientation called the **pullback orientation induced by  $f$** ; in this case  $f$  is said to **pull back orientation preserving**.

If the orientation of the  $m$ -dimensional manifold  $N$  is determined by an  $m$ -form  $\omega$ , then the pullback orientation on  $M$  is determined by the  $m$ -form  $f^*\omega$ .

**Theorem 9.4.3** Let  $M$  and  $N$  be already oriented smooth manifolds, and let  $f: M \rightarrow N$  be a local diffeomorphism. If  $M$  is connected, then  $f$  either preserves orientation everywhere or reverses orientation everywhere.

**Theorem 9.4.4** Let  $M$  be an oriented  $m$ -dimensional smooth manifold,  $S$  an immersed hypersurface in  $M$ , and  $V$  a vector field on  $S$  that is nowhere tangent to  $S$ . For a point  $p \in S$ , let  $(e_1, e_2, \dots, e_{m-1})$  be a basis of the tangent space  $T_p S$ . Then  $S$  has a unique orientation, and  $(e_1, e_2, \dots, e_{m-1})$  determines an orientation of  $S$  **if and only if** the orientation of  $M$  is determined by the ordered frame  $(V_p, e_1, e_2, \dots, e_{m-1})$ . Let  $f: S \rightarrow M$  be the inclusion map. If  $\omega$  is an  $m$ -form that determines the orientation of  $M$ , then the corresponding orientation of  $S$  is determined

by the  $(m-1)$ -form  $\sigma = f_S^*(V \lrcorner \omega)$ , where  $V \lrcorner$  denotes the interior product with  $V$  and

$$\sigma_p(e_1, e_2, \dots, e_{m-1}) = \omega_p(V_p, e_1, e_2, \dots, e_{m-1}) \neq 0. \quad (9.4.1)$$

**Theorem 9.4.5** Let  $M$  be an oriented  $m$ -dimensional smooth manifold (with or without boundary), and let  $S \subseteq M$  be an  $m$ -dimensional submanifold (with or without boundary). Let  $f : S \rightarrow M$  be the inclusion map. Then the orientation of  $M$  restricts to an orientation of  $S$ , and if  $\omega$  is an  $m$ -form that determines the orientation of  $M$ , then  $\sigma = f_S^* \omega$  is an  $m$ -form that determines the orientation of  $S$ .

**Example 9.4.1** Let the basis vectors of the coordinate axes  $x, y, z$  in  $R^3$  be denoted by  $e_1, e_2, e_3$ . The orientation of  $R^3$  determined by the ordered basis  $e_1, e_2, e_3$  is the right-handed orientation. The 3-form on  $R^3$ ,

$$\omega(e_1, e_2, e_3) = dx \wedge dy \wedge dz$$

together with  $\omega' = g dx \wedge dy \wedge dz$ ,  $g > 0$ , determines the same orientation on  $R^3$ , called the right-handed orientation.

Consider the disk  $S$  centered at the origin with radius 1 in the  $xy$ -plane  $R^2$ , which is a submanifold of  $R^3$ . The inclusion map is

$$f : S \rightarrow R^3.$$

The vector field  $V = \frac{\partial}{\partial z}$  is a normal vector field on  $S$ . Compute:

$$i_V \omega(e_1, e_2, e_3) = i_V (dx \wedge dy \wedge dz) = dx \wedge dy,$$

where the last step uses the result from Example 8.6.3. Therefore, we obtain

$$V \lrcorner \omega = i_V \omega(e_1, e_2, e_3) = \omega(V, e_1, e_2, e_3) = dx \wedge dy.$$

The form  $V \lrcorner \omega$  is a 2-form on  $R^3$ , but via the pullback  $f_S^*$  it is mapped to the disk  $S$ , giving

$$\begin{aligned} \sigma &= f_S^*(V \lrcorner \omega) = f_S^*(dx \wedge dy) = (f_S^* dx) \wedge (f_S^* dy) \\ &= (d(f_S^* x)) \wedge (d(f_S^* y)) = dx \wedge dy. \end{aligned}$$

This 2-form  $\sigma = dx \wedge dy$  on  $S$  determines the orientation of the disk  $S$ , which is consistent with the right-handed orientation of  $R^3$ .

## §9.5 Orientation of Product Manifolds

**Theorem 9.5.1** Let  $M$  and  $N$  be connected smooth manifolds.

- 1) If both  $M$  and  $N$  are orientable, then  $M \times N$  is also orientable.
- 2) If either  $M$  or  $N$  is non-orientable, then  $M \times N$  is also non-orientable.

**Theorem 9.5.2** Let  $M_1, \dots, M_k$  be orientable smooth manifolds. There exists a unique **product orientation** on the product manifold  $M_1 \times \dots \times M_k$ , determined as follows:

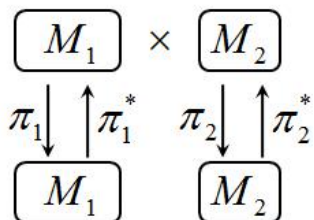
If for each  $i$  an  $m_i$ -form  $\omega_i$  gives the orientation of  $M_i$ , then the form

$$\pi_1^* \omega_1 \wedge \dots \wedge \pi_k^* \omega_k$$

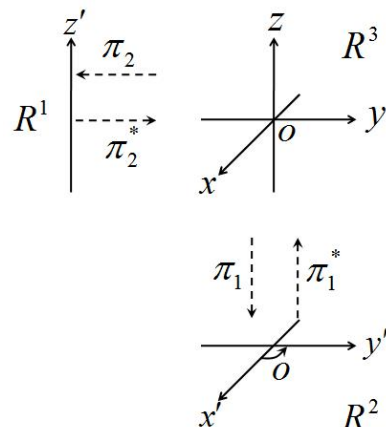
determines the product orientation, where  $\pi_i : M_1 \times \dots \times M_k \rightarrow M_i$  is the natural projection, and  $\pi_i^*$  denotes the pullback via  $\pi_i$ .

**Example 9.5.1** As shown in Figure 9.5.1, consider two orientable smooth manifolds  $M_1$  and  $M_2$ . Let the differential form  $\omega_1$  determine the orientation of  $M_1$ , and  $\omega_2$  determine the orientation of  $M_2$ . The pullback  $\pi_1^*$  via the natural projection  $\pi_1 : M_1 \times M_2 \rightarrow M_1$  maps the differential form  $\omega_1$  on  $M_1$  to the product manifold  $M_1 \times M_2$ , yielding  $\pi_1^* \omega_1$ . Similarly, the pullback  $\pi_2^*$  via the natural projection  $\pi_2 : M_1 \times M_2 \rightarrow M_2$  maps  $\omega_2$  on  $M_2$  to  $M_1 \times M_2$ ,

yielding  $\pi_2^* \omega_2$ . The exterior product  $\pi_1^* \omega_1 \wedge \pi_2^* \omega_2$  then determines the unique product orientation of the product manifold  $M_1 \times M_2$ .



**Figure 9.5.1** Orientation of the product manifold



**Figure 9.5.2** Product orientation of the product manifold  $R^3$  (Part 1)

**Example 9.5.2** As shown in Figure 9.5.2, the orientation of  $R^2$  is determined by  $dx' \wedge dy'$ , with counterclockwise as the positive orientation. The orientation of  $R^1$  is determined by  $dz'$ , with upward as the positive orientation.  $R^3$  is the product of  $R^2$  and  $R^1$ , i.e.,  $R^3 = R^2 \times R^1$ . The natural projections from the product manifold  $R^2 \times R^1$  to each factor are

$$\pi_1 : R^2 \times R^1 \rightarrow R^2, \quad \pi_2 : R^2 \times R^1 \rightarrow R^1,$$

that is, for any point  $(x, y, z) \in R^2 \times R^1$ ,

$$\pi_1(x, y, z) \rightarrow (x', y'), \quad \pi_2(x, y, z) \rightarrow z'.$$

The natural projection  $\pi_1$  expressed in terms of functions is

$$x' = x, \quad y' = y,$$

where on the left side  $(x', y') \in R^2$  and on the right side  $x, y$  is the coordinate on  $R^2 \times R^1$ . According to Theorem 8.5.3 and Theorem 8.5.4, we obtain

$$\pi_1^*(dx' \wedge dy') = \pi_1^*(dx') \wedge \pi_1^*(dy') = d(\pi_1^*(x')) \wedge d(\pi_1^*(y')) = dx \wedge dy,$$

so the mapping  $\pi_1^*$  pulls back the 1-form  $dx' \wedge dy'$  on  $R^2$  to  $dx \wedge dy$  on  $R^2 \times R^1$ .

The natural projection  $\pi_2$  expressed in terms of functions is

$$z' = z,$$

where on the left side  $z' \in R^1$  and on the right side  $z$  is the coordinate on  $R^2 \times R^1$ . According to Theorem 8.5.3 and Theorem 8.5.4, we obtain

$$\pi_2^*(dz') = d(\pi_2^*(z')) = dz,$$

so the mapping  $\pi_2^*$  pulls back the 1-form  $dz'$  on  $R^1$  to  $dz$  on  $R^2 \times R^1$ .

Since

$$\pi_1^*(dx' \wedge dy') \wedge \pi_2^*(dz') = dx \wedge dy \wedge dz,$$

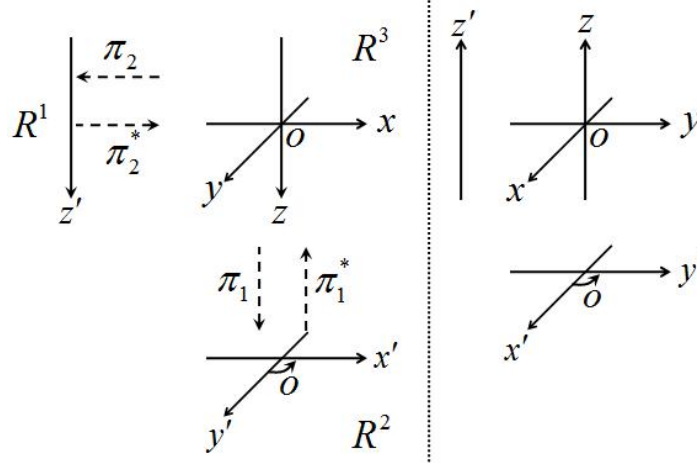
it follows that  $dx \wedge dy \wedge dz$  determines the unique product orientation on the product manifold  $R^3$ , which is precisely the right-hand rule.

**Example 9.5.3** As shown in Figure 9.5.3, the orientation of  $R^2$  is determined by  $dy' \wedge dx'$ , with clockwise as the positive orientation. The orientation of  $R^1$  is determined by  $-dz'$ , with downward as the positive orientation. The pullback of  $dy' \wedge dx'$  from  $R^2$  to  $R^3$  via  $\pi_1^*$  is  $\pi_1^*(dy' \wedge dx') = dy \wedge dx$ , and the pullback of  $-dz'$  from  $R^1$  to  $R^3$  via  $\pi_2^*$  is

$\pi_2^*(-dz') = -dz$ . Since

$$\pi_1^*(dy' \wedge dx') \wedge \pi_2^*(-dz') = dy \wedge dx \wedge (-dz) = -dy \wedge dx \wedge dz = dx \wedge dy \wedge dz,$$

it follows that  $dx \wedge dy \wedge dz$  determines the unique product orientation on the product manifold  $R^3$ , which also corresponds to the right-hand rule.



**Figure 9.5.3** Product orientation of the product manifold  $R^3$  (Part 2)

In both Example 9.5.2 and 9.5.3, the orientations determined by  $\pi_1^*(dy' \wedge dx') \wedge \pi_2^*(-dz')$  and  $\pi_1^*(dx' \wedge dy') \wedge \pi_2^*dz'$  on  $M$  are the same. This can be understood as follows: in transitioning from Figure 9.5.3 to Figure 9.5.2, the orientations of both  $R^2$  and  $R^1$  are reversed simultaneously—the orientation of  $R^2$  is changed from clockwise to counterclockwise, while the orientation of  $R^1$  is changed from downward to upward.

## §9.6 Integration of Differential Forms on Manifolds

**Definition 9.6.1** Let  $M$  be an oriented  $m$ -dimensional smooth manifold. At the point  $p \in M$ , establish a local coordinate system  $(U; u^i)$  consistent with the orientation. Let  $\omega$  be an  $m$ -form on  $U$ ,

$$\omega = f(u^1, \dots, u^m) du^1 \wedge \dots \wedge du^m,$$

then the integral of  $\omega$  over  $M$  is defined as the Riemann integral of  $\varphi$  over  $U$ , i.e.,

$$\int_M \omega = \int_U f(u^1, \dots, u^m) du^1 \wedge \dots \wedge du^m. \quad (9.6.1)$$

Here,  $\omega$  is an  $m$ -form and also a tensor field, which remains invariant under coordinate transformations. Therefore, the integral (9.6.1) is also invariant under coordinate transformations, meaning it is independent of the choice of coordinate system. This is illustrated with examples below.

**Example 9.6.1** Suppose the local coordinate system  $(U; u, v)$  on a two-dimensional manifold  $M$  is compatible with the orientation of  $M$ . Let  $\varphi = f(u, v) du \wedge dv$  be a 2-form on  $M$ , with the integral over  $U$  given by

$$\int_M \varphi = \int_U f(u, v) du \wedge dv.$$

Consider another local coordinate system  $(V; x, y)$  on  $M$  that is also orientation-compatible, satisfying

$$u = u(x, y), \quad v = v(x, y),$$

then

$$\begin{aligned}
du \wedge dv &= \left( \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy \right) \wedge \left( \frac{\partial v}{\partial x} dx + \frac{\partial v}{\partial y} dy \right) \\
&= \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} dx \wedge dy + \frac{\partial u}{\partial y} \frac{\partial v}{\partial x} dy \wedge dx \\
&= \left( \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \frac{\partial v}{\partial x} \right) dx \wedge dy \\
&= \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} dx \wedge dy = \frac{\partial(u, v)}{\partial(x, y)} dx \wedge dy,
\end{aligned}$$

where

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}.$$

Thus,

$$\int_M \varphi = \int_U f(u, v) du \wedge dv = \int_V f(u, v) \frac{\partial(u, v)}{\partial(x, y)} dx \wedge dy = \int_V g(x, y) dx \wedge dy, \quad (9.6.2)$$

where

$$f(u, v) \frac{\partial(u, v)}{\partial(x, y)} = g(x, y).$$

If  $f(u, v) = 1$ , then equation (9.6.2) becomes

$$\int_M \varphi = \int_U du \wedge dv = \int_V \frac{\partial(u, v)}{\partial(x, y)} dx \wedge dy. \quad (9.6.3)$$

Here,  $\varphi = du \wedge dv$  is referred to as the **area form** on  $M$ . If the wedge product notation is omitted,  $\varphi$  can be written as  $\varphi = dudv$ , which aligns with the area element in classical calculus.

**Example 9.6.2** Suppose the local coordinate system  $(U; x, y, z)$  on the three-dimensional manifold  $S^3$  is compatible with the orientation of  $S^3$ . Let  $\varphi = dx \wedge dy \wedge dz$  be a 3-form on  $S^3$ , with the integral over  $U$  given by

$$\int_{S^3} \varphi = \int_U dx \wedge dy \wedge dz.$$

Consider another orientation-compatible local coordinate system on  $S^3$ : the spherical coordinate system  $(V; r, \theta, \varphi)$ , where

$$x = r \sin \theta \cos \varphi, \quad y = r \sin \theta \sin \varphi, \quad z = r \cos \theta,$$

then

$$\begin{aligned}
\varphi = dx \wedge dy \wedge dz &= \begin{vmatrix} \sin \theta \cos \varphi & r \cos \theta \cos \varphi & -r \sin \theta \sin \varphi \\ \sin \theta \sin \varphi & r \cos \theta \sin \varphi & r \sin \theta \cos \varphi \\ \cos \theta & -r \sin \theta & 0 \end{vmatrix} dr \wedge d\theta \wedge d\varphi \\
&= r^2 \sin \theta dr \wedge d\theta \wedge d\varphi.
\end{aligned} \quad (9.6.4)$$

Therefore,

$$\int_{S^3} \varphi = \int_U dx \wedge dy \wedge dz = \int_V r^2 \sin \theta dr \wedge d\theta \wedge d\varphi. \quad (9.6.5)$$

Here,  $\varphi = dx \wedge dy \wedge dz$  is referred to as the **volume form** on  $S^3$  in Cartesian coordinates. In spherical coordinates, the volume form becomes  $\varphi = r^2 \sin \theta dr \wedge d\theta \wedge d\varphi$ . Similarly, if the wedge product notation is omitted,  $\varphi$  can be written as  $\varphi = dx dy dz$  or  $\varphi = r^2 \sin \theta dr d\theta d\varphi$ , which

aligns with the volume element in classical calculus.

## §9.7 Degree of a Map and Brouwer Degree

**Definition 9.7.1** Let  $M$  and  $N$  be  $C^r$  manifolds,  $r \geq 1$ , and let  $f: M \rightarrow N$  be a  $C^1$  map. If at a point  $p \in M$ ,  $\text{rank}(f) < \dim N$ , then the point  $p \in M$  is called a **critical point** of the map  $f$ . The set of all critical points of  $f$  is denoted by  $C_f$ .

If at a point  $p \in M$ ,  $\text{rank}(f) = \dim N$ , then the point  $p \in M$  is called a **regular point** of the map  $f$ .

**Definition 9.7.2** Let  $M$  and  $N$  be smooth manifolds, and let  $f: M \rightarrow N$  be a smooth map, with  $q \in N$ . If  $f^{-1}(q) \cap C_f \neq \emptyset$ , then the point  $q$  is called a **critical value** of  $f$ . If  $f^{-1}(q) \cap C_f = \emptyset$ , then the point  $q$  is called a **regular value** of  $f$ .

Denote the set of critical values of  $f$  as  $f(C_f)$ . The set of regular values of  $f$  is denoted by  $N - f(C_f)$ , or simply by  $N \setminus f(C_f)$ . When  $f^{-1}(q) = \emptyset$ , the point  $q \in N$  is also considered a regular value of  $f$ .

**Theorem 9.7.1** Let  $M$  and  $N$  be  $m$ -dimensional oriented  $C^1$  manifolds, with  $M$  compact. Let  $f: M \rightarrow N$  be a  $C^1$  map, and let  $q \in N$  be a regular value of  $f$ . Then the preimage  $f^{-1}(q)$  is a finite set (which may be empty).

**Definition 9.7.3** Let  $M$  and  $N$  be compact, oriented smooth manifolds of the same dimension, and let  $f: M \rightarrow N$  be a  $C^1$  map. Let  $p \in M$  be a regular point of  $f$ , with  $q = f(p)$ . If the induced linear map  $f_*: T_p M \rightarrow T_q N$  on tangent spaces is orientation-preserving, then  $p$  is said to be of **positive type**, denoted by  $\deg_p f = 1$ . If  $f_*$  reverses orientation, then  $p$  is said to be of **negative type**, denoted by  $\deg_p f = -1$ . We call  $\deg_p f$  the **degree** of the map  $f$  at the point  $p \in M$ .

**Definition 9.7.4** Assume  $q \in N$  is an arbitrary regular value of the map  $f$ . Define the **Brouwer degree** of  $f$  with respect to the regular value  $q \in N$  as

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f ; \quad (9.7.1)$$

if  $f^{-1}(q) = \emptyset$ , then set  $\deg(f, q) = 0$ .

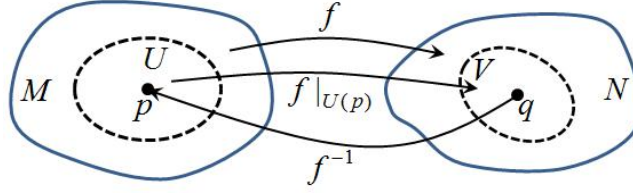
Sometimes the Brouwer degree is also referred to as the **winding number**.

The Brouwer degree essentially counts the number of preimages, taking into account the orientation-determined sign. According to Theorem 9.7.1, the set  $f^{-1}(q)$  is finite, meaning it has finitely many elements. Therefore, the right-hand side of equation (9.7.1) is a finite sum, and  $\deg(f, q)$  is an integer.

**Theorem 9.7.2** Let  $M$  and  $N$  be  $m$ -dimensional oriented  $C^2$  manifolds, with  $M$  compact and  $N$  connected. Let  $f: M \rightarrow N$  be a  $C^2$  map. Then  $\deg(f, q)$  does not depend on the choice of the regular value  $q$  of  $f$ .

Let us understand the geometric meaning of  $\deg(f, q)$ . Suppose  $f^{-1}(q)$  contains  $n$  points of positive type and  $m$  points of negative type; then  $\deg(f, q) = n - m$ . According to the inverse function theorem, there exists an open neighborhood  $V \subset N$  of  $q$  and, for each  $p \in f^{-1}(q)$ , an open neighborhood  $U(p) \subset M$  of  $p$  such that  $f|_{U(p)}: U(p) \rightarrow V$  is a  $C^1$  diffeomorphism. This diffeomorphism preserves orientation if  $p$  is of positive type and reverses orientation if  $p$  is of negative type. Thus,  $\deg(f, q)$  represents the algebraic sum of the covering multiplicity of  $V$

under the map  $f$  via the open sets  $U(p)$ . See Figure 9.7.1.



**Figure 9.7.1** The open set  $V$  is covered by the open sets  $U(p)$  with multiplicity  $\deg(f, q)$

**Example 9.7.1** As shown in Figure 9.7.2, the mapping  $f: S^1 \rightarrow R^1$  is defined by projecting the unit circle  $S^1$  onto the real line  $R^1 = \{(u, 0) \mid u \in R\}$  with respect to the fixed point  $(0, 2)$ . Let the coordinates of a point on the unit circle be  $(x, y) = (\cos \theta, \sin \theta)$ , where  $0 \leq \theta < 2\pi$ . Then the mapping  $f$  can be expressed as  $u = f(x, y)$ . Set

$$t = \frac{\overrightarrow{p_1 C}}{\overrightarrow{p_1 A}} = \frac{\overrightarrow{OC} - \overrightarrow{Op_1}}{\overrightarrow{OA} - \overrightarrow{Op_1}} = \frac{(u, 0) - (x, y)}{(0, 2) - (x, y)},$$

then

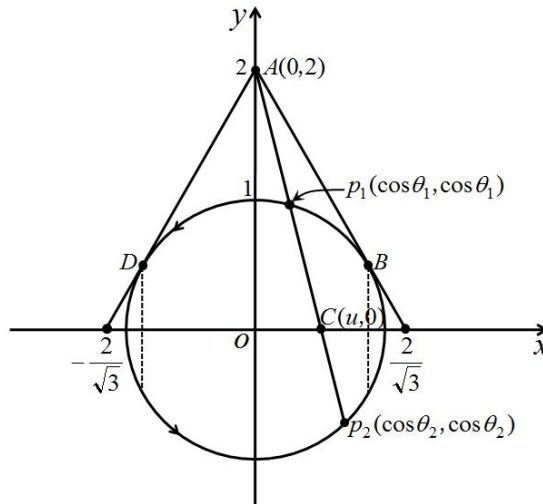
$$\begin{aligned} t(0, 2) - t(x, y) &= (u, 0) - (x, y), \\ t(0, 2) + (1 - t)(x, y) &= (u, 0), \\ ((1 - t)x, (1 - t)y + 2t) &= (u, 0), \\ u &= (1 - t)x, (1 - t)y + 2t = 0. \end{aligned}$$

Hence,

$$t = \frac{y}{y - 2}, \quad u = (1 - t)x = \left(1 - \frac{y}{y - 2}\right)x = \frac{2x}{2 - y} = \frac{2 \cos \theta}{2 - \sin \theta}.$$

The Jacobian matrix of  $f$  is

$$J = \left( \frac{du}{d\theta} \right) = \left( \frac{2(1 - 2 \sin \theta)}{(2 - \sin \theta)^2} \right).$$



**Figure 9.7.2** Mapping from the unit circle  $S^1$  to the real axis  $R^1$

When  $\theta = \frac{\pi}{6}$  or  $\theta = \frac{5\pi}{6}$ , the partial derivatives  $\frac{du}{d\theta} = 0$ , making the rank of the Jacobian matrix  $J$  becomes 0, which is less than the dimension 1 of  $R^1$ . Therefore, the points

$B(\cos \frac{\pi}{6}, \sin \frac{\pi}{6}) = B(\frac{\sqrt{3}}{2}, \frac{1}{2})$  and  $D(\cos \frac{5\pi}{6}, \sin \frac{5\pi}{6}) = D(-\frac{\sqrt{3}}{2}, \frac{1}{2})$  on  $S^1$  are critical points.

Denote the set of critical points of  $f$  as

$$C_f = \left\{ B(\frac{\sqrt{3}}{2}, \frac{1}{2}), D(-\frac{\sqrt{3}}{2}, \frac{1}{2}) \right\}.$$

The critical value corresponding to the critical point  $B(\frac{\sqrt{3}}{2}, \frac{1}{2})$  is

$$f(B) = f(\frac{\sqrt{3}}{2}, \frac{1}{2}) = \frac{2 \times \frac{\sqrt{3}}{2}}{2 - \frac{1}{2}} = \frac{2}{\sqrt{3}}.$$

The critical value corresponding to the critical point  $D(-\frac{\sqrt{3}}{2}, \frac{1}{2})$  is

$$f(D) = f(-\frac{\sqrt{3}}{2}, \frac{1}{2}) = \frac{2 \times (-\frac{\sqrt{3}}{2})}{2 - \frac{1}{2}} = -\frac{2}{\sqrt{3}}.$$

Denote the set of critical values of the mapping  $f$  as

$$f(C_f) = \left\{ \frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}} \right\},$$

then the set of regular values of  $f$  is

$$R^1 - f(C_f) = R^1 - \left\{ \frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}} \right\}.$$

The orientation on  $R^1$  is taken as the positive direction of the  $x$ -axis, and the orientation on  $S^1$  is taken as counterclockwise.

The open arc  $Bp_1D$  (excluding points  $B$  and  $D$ ) in the upper half-circle projects to  $R^1$  with an orientation opposite to that of  $R^1$  (orientation-reversing), so

$$\deg_p f = -1, \quad \frac{\pi}{6} < \theta < \frac{5\pi}{6}.$$

The open arc  $Dp_2B$  projects to  $R^1$  with the same orientation as  $R^1$  (orientation-preserving), so

$$\deg_p f = 1, \quad \frac{5\pi}{6} < \theta < 2\pi + \frac{\pi}{6}.$$

That is,

$$\begin{cases} \deg_p f = -1, & p = (\cos \theta, \sin \theta), \quad \frac{\pi}{6} < \theta < \frac{5\pi}{6}; \\ \deg_p f = 1, & p = (\cos \theta, \sin \theta), \quad \frac{5\pi}{6} < \theta < 2\pi + \frac{\pi}{6}. \end{cases}$$

The Brouwer degree of  $f$  with respect to the regular value  $q \in R^1 - f(C_f)$  is

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = -1 + 1 = 0.$$

As illustrated in Figure 9.7.2, although the preimage of point  $C$  consists of two points  $p_1$  and  $p_2$ , point  $p_1$  lies on the open arc  $Bp_1D$ , which projects to  $R^1$  with reversed orientation, so  $\deg_p f = -1$ . Point  $p_2$  lies on the open arc  $Dp_2B$ , which projects to  $R^1$  with preserved orientation, so  $\deg_p f = 1$ . Therefore, the total degree sums to zero:  $\deg(f, q) = 0$ .

**Example 9.7.2** As shown in Figure 9.7.3, let the coordinates of a point on the unit circle



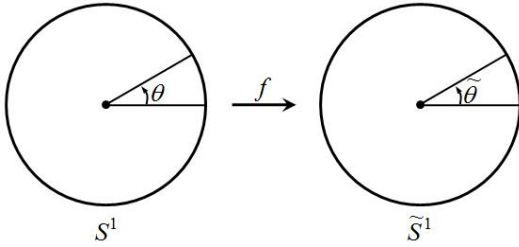
$S^1$  be  $(x, y) = (\cos \theta, \sin \theta)$ ,  $0 \leq \theta < 2\pi$ , and the coordinates of a point on another unit circle  $\tilde{S}^1$  be  $(\tilde{x}, \tilde{y}) = (\cos \tilde{\theta}, \sin \tilde{\theta})$ . The mapping  $f: S^1 \rightarrow \tilde{S}^1$  sends the unit circle  $S^1$  to the unit circle  $\tilde{S}^1$ . The mapping  $f$  is defined by

$$f: e^{i\theta} \rightarrow e^{i\tilde{\theta}}, \quad \theta \mapsto \tilde{\theta} = n\theta,$$

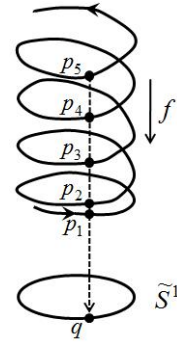
where  $n$  is an integer. Assume the orientation of  $S^1$  is counterclockwise, which is taken as the positive orientation.

If  $n=1$ , then  $\theta \mapsto \tilde{\theta} = \theta$ ,  $0 \leq \tilde{\theta} < 2\pi$ , and the mapping  $f$  preserves the positive orientation (i.e., the orientations of  $\tilde{S}^1$  and  $S^1$  coincide) and covers the target circle exactly once (the mapping  $f$  rotates the unit circle  $S^1$ , then after one full rotation the domain circle  $S^1$  is mapped onto the target circle  $\tilde{S}^1$ ). Hence,  $\deg_p f = 1$ ,  $0 \leq \theta < 2\pi$ , so

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = \deg_p f = 1.$$



**Figure 9.7.3**  $S^1$  covers  $\tilde{S}^1$   $n$  times



**Figure 9.7.4**  $S^1$  rotating 5 times becomes  $\tilde{S}^1$

If  $n=5$ , then  $\theta \mapsto \tilde{\theta} = 5\theta$ ,  $0 \leq \tilde{\theta} < 10\pi$ , and  $f$  preserves the positive orientation while covering the target circle five times. In other words, the map  $f$  winds the domain circle  $S^1$  five times counterclockwise onto the target circle  $\tilde{S}^1$ , as shown in Figure 9.7.4. The preimage of a point  $q$  consists of five points:  $p_1, p_2, p_3, p_4$  and  $p_5$ . Since the map is orientation-preserving, each point contributes a degree of +1, so

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = 1 + 1 + 1 + 1 + 1 = 5.$$

If  $n=k$ , then  $\theta \mapsto \tilde{\theta} = k\theta$ ,  $0 \leq \tilde{\theta} < 2k\pi$ , and  $f$  preserves the positive orientation while covering the target circle  $k$  times. As the map is orientation-preserving, each preimage point contributes +1,  $\deg_p f = 1$ ,  $0 \leq \theta < 2\pi$ , giving

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = \underbrace{1 + \cdots + 1}_{k \text{ copies}} = k.$$

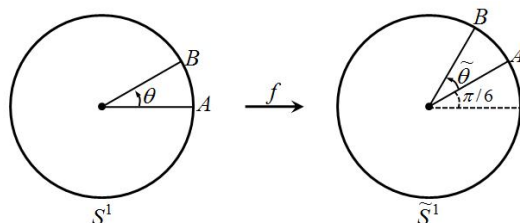
If  $n=-1$ , then  $\theta \mapsto \tilde{\theta} = -\theta$ ,  $-2\pi < \tilde{\theta} \leq 0$ , and the map  $f$  covers the target circle once but reverses the orientation (clockwise). Thus, the map is orientation-reversing, so each preimage point contributes -1,  $\deg_p f = -1$ ,  $0 \leq \theta < 2\pi$ , and

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = \deg_p f = -1.$$

As shown in Figure 9.7.5, If  $\theta \mapsto \tilde{\theta} = \theta + \pi/6$ , then  $\pi/6 \leq \tilde{\theta} < 2\pi + \pi/6$ , and starting from the initial point  $\theta = \pi/6$ , the mapping  $f$  rotates the unit circle  $S^1$ , then after one full rotation the domain circle  $S^1$  is mapped onto the target circle  $\tilde{S}^1$ , and the orientations coincide. Similarly, If  $\theta \mapsto \tilde{\theta} = \theta + \pi/3$ , then  $\pi/3 \leq \tilde{\theta} < 2\pi + \pi/3$ , and starting from the initial point  $\theta = \pi/3$ , the

mapping  $f$  rotates the unit circle  $S^1$ , then after one full rotation the domain circle  $S^1$  is mapped onto the target circle  $\tilde{S}^1$ , and the orientations coincide, that is the mapping also preserves orientation after one full rotation. In both cases, the preimage of a regular value consists of one point with degree  $+1$ ,  $\deg_p f = 1$ ,  $0 \leq \theta < 2\pi$ , so

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = \deg_p f = 1.$$



**Figure 9.7.5** After one rotation starting from  $\theta = \pi/6$ ,  $S^1$  becomes  $\tilde{S}^1$

If the range of  $\theta$  is changed to  $0 \leq \theta \leq 2\pi + \pi/6$  and the mapping  $f$  is defined as  $\theta \mapsto \tilde{\theta} = \theta$ , then the map  $f$  rotates the domain circle  $S^1$  by an angle  $\theta = 2\pi + \pi/6$  and maps it onto the target circle  $\tilde{S}^1$  while preserving the orientation. Hence, the mapping  $f$  is orientation-preserving, and therefore,

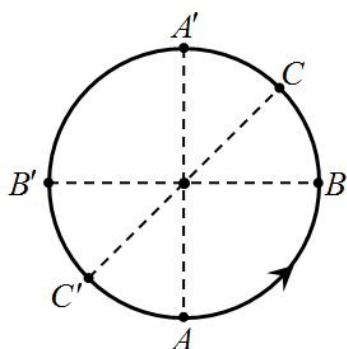
$$\begin{cases} \deg_p f = 2, & p = (\cos \theta, \sin \theta), 0 \leq \theta \leq \frac{\pi}{6}; \\ \deg_p f = 1, & p = (\cos \theta, \sin \theta), \frac{\pi}{6} < \theta < 2\pi. \end{cases}$$

**Definition 9.7.5** Assume  $S^m$  is the unit sphere in  $R^{m+1}$ . The **antipodal map** of  $S^m$  is the central symmetry about the origin, denoted by  $f: S^m \rightarrow S^m$ , i.e.,  $f(x) = -x$ ,  $\forall x \in S^m$ .

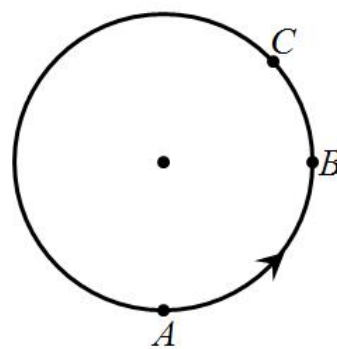
**Theorem 9.7.3** Let  $f: S^m \rightarrow S^m$  be the antipodal map. Then

$$\deg(f) = (-1)^{m+1}.$$

**Example 9.7.3** The antipodal map  $f$  on the unit circle  $S^1$  is illustrated in Figure 9.7.6. Point  $A$  maps to point  $A'$ , point  $A'$  maps to point  $A$ ; point  $B$  maps to point  $B'$ , point  $B'$  maps to point  $B$ ; point  $C$  maps to point  $C'$ , point  $C'$  maps to point  $C$ ; and so on. This antipodal map  $f$  corresponds to rotating the unit circle  $S^1$  by half a turn starting from point  $A$ . Thus, the right semicircle covers the left semicircle, and the left semicircle covers the right semicircle. Consequently, the mapping degree of this antipodal map  $f$  is 1, and its Brouwer degree is also 1.



**Figure 9.7.6** The antipodal map on  $S^1$



**Figure 9.7.7** The identity map on  $S^1$

**Theorem 9.7.4** The Brouwer degree of the identity map  $id: M \rightarrow M$  is equal to 1, i.e.,  $\deg(id) = 1$ .

**Example 9.7.4** If a unit circle  $S^1$  rotates once around itself, as shown in Figure 9.7.7, then

point  $A$  returns to its original position after one full rotation. Similarly, point  $B$  returns to its original position after one full rotation, and so on. Therefore, rotating the unit circle  $S^1$  once is equivalent to the identity map, meaning the unit circle covers itself exactly once. Thus, the mapping degree of the rotation map is 1, and its Brouwer degree is also 1.

**Theorem 9.7.5** If the map  $f : M \rightarrow N$  is a  $C^1$  diffeomorphism, then

$$\deg(f, q) = \sum_{p \in f^{-1}(q)} \deg_p f = \deg_p f = \begin{cases} +1, & \text{if } f \text{ is orientation - preserving} \\ -1, & \text{if } f \text{ is orientation - reversing,} \end{cases}$$

where  $f(p) = q$ .

## Chapter 10 Affine Connections

### §10.1 Affine Connections

The purpose of introducing an affine connection is to differentiate or take derivatives of tensor fields on a manifold. We are already familiar with the differentiation of vectors in Euclidean space, so let us first review how vectors are differentiated in Euclidean space, and then extend the method to general manifolds.

As shown in Figure 10.1.1, in Euclidean space, the derivative of a vector  $A(t)$  is computed by parallel transporting the vector  $A(t + \Delta t)$  along a straight line to point  $p$ , so that the starting point of  $A(t + \Delta t)$  coincides with that of  $A(t)$ . Since the magnitude and direction of the vector do not change during this parallel transport, the vector  $A'(t + \Delta t)$  at point  $p$  equals the vector  $A(t + \Delta t)$  at point  $q$ , i.e.,

$$A'(t + \Delta t) = A(t + \Delta t).$$

Thus, comparing the magnitude and direction of  $A'(t + \Delta t)$  and  $A(t)$  at point  $p$  is equivalent to comparing the vectors  $A(t)$  at point  $p$  and  $A(t + \Delta t)$  at points  $q$ . Therefore, the difference between  $A(t + \Delta t)$  and  $A(t)$  can be computed at point  $p$ , and the derivative of  $A(t)$  is given by

$$\frac{dA(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{A(t + \Delta t) - A(t)}{\Delta t}.$$

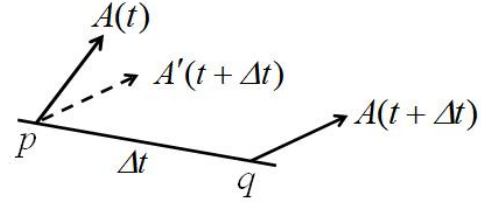
This method can be generalized to a general manifold  $M$ . As shown in Figure 10.1.2, let  $X$  be the tangent vector field along a curve  $\gamma$ , i.e.,  $X = \frac{d}{dt}$ . At point  $p$ , take a tangent vector  $Y$  in the tangent space  $T_p M$ , and then parallel transport it along the curve  $\gamma$ . This yields a tangent vector field generated by  $Y$ , still denoted as  $Y$ . Since  $Y$  is parallel transported, we say that  $Y$  does not change along  $\gamma$ ; that is, the rate of change of  $Y$  is zero. Denote the rate of change of  $Y$  along  $\gamma$  as  $D_X Y$ ; then

$$D_X Y = 0 \Leftrightarrow \text{when } Y \text{ is parallel transported along } \gamma.$$

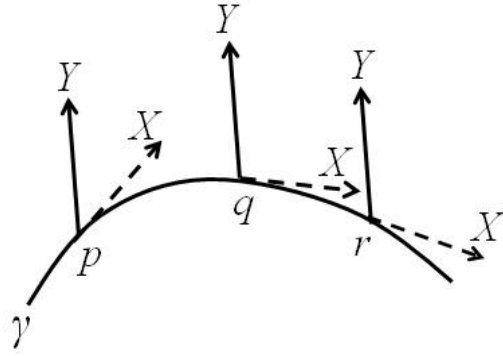
If  $Y$  is not parallel transported along  $\gamma$ , then  $D_X Y \neq 0$ . The question then is: how do we compute  $D_X Y$  when it is nonzero?

The method used in Euclidean space can serve as a reference. However, it is important to note that there are differences between the computational approach in Euclidean space and that on a general manifold:

- 1) In Euclidean space, the result  $\frac{dA(t)}{dt}$  obtained from the computation is the same at every point. On a general manifold, the derivative of a vector field generally has different values at different points.
- 2) In Euclidean space, algebraic operations such as addition, subtraction, and multiplication



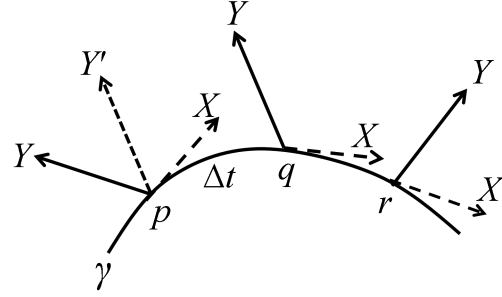
**Figure 10.1.1** Computing the derivative of a vector  $A(t)$  in Euclidean space



**Figure 10.1.2** Parallel transport of a tangent vector  $Y$  along a curve  $\gamma$

can be performed on vectors at any two distinct points. On a general manifold, however, algebraic operations on vectors at different points are not directly possible; such operations can only be carried out within the tangent space at the same point.

As shown in Figure 10.1.3, we compute  $(D_X Y)_p$  at point  $p$ . Although the tangent vector  $Y$  is not parallel transported along the curve  $\gamma$ , we can parallel transport the tangent vector  $Y_q$  from point  $q$  to point  $p$ , obtaining a new vector field  $Y'$ . Since  $Y'$  is obtained by parallel transporting  $Y_q$ , we have  $D_X Y' = 0$  along  $\gamma$ . Within the tangent space  $T_p M$  at point  $p$ , we can compare  $Y_p$  and



**Figure 10.1.3** Computing  $(D_X Y)_p$  at point  $p$

$Y'_p$ . Because at point  $q$ ,  $Y'_q = Y_q$ , comparing  $Y_p$  and  $Y'_p$  in  $T_p M$  is equivalent to comparing  $Y_p$  at point  $p$  and  $Y_q$  at point  $q$ . Thus, we overcome the problem that algebraic operations cannot be directly performed on vectors at different points on a general manifold. In this way, we can compute the rate of change of the tangent vector  $Y$  along the curve  $\gamma$  within the tangent space  $T_p M$ . Assuming the parameter value at point  $q$  is greater than that at point  $p$  by  $\Delta t$ , we obtain

$$(D_X Y)_p = \lim_{\Delta t \rightarrow 0} \frac{Y'_p - Y_p}{\Delta t}. \quad (10.1.1)$$

Here,  $D$  is called an **affine connection**. From equation (10.1.1), it can be seen that  $D_X$  is a differential operator and should therefore obey the Leibniz rule:

$$D_X(Y + W) = D_X Y + D_X W,$$

$$D_X(fY) = fD_X Y + YD_X f = fD_X Y + Y \frac{df}{dt}.$$

In the last term above, we stipulate that  $D_X$  acting on a function  $f$  equals the directional derivative  $\frac{df}{dt}$ , i.e.,

$$D_X f = \frac{df}{dt}.$$

At the same point, the operators  $D_X$  and  $D_W$  along different vectors should be additive, meaning

$$(D_X Y)_p + (D_W Y)_p = (D_{X+W} Y)_p.$$

If the parameter of the tangent vector  $X$  is changed from  $t$  to  $\mu$ , with  $t = t(\mu)$ , then

$$\frac{d}{dt} = \frac{d\mu}{dt} \frac{d}{d\mu},$$

and

$$dt = \frac{dt}{d\mu} d\mu.$$

Let  $g = \frac{dt}{d\mu}$ , we obtain

$$\frac{d}{dt} = \frac{1}{g} \frac{d}{d\mu},$$

and

$$\Delta t = g \Delta \mu$$

or

$$\Delta\mu = \frac{1}{g} \Delta t.$$

Hence, when the parameter of the tangent vector  $X = \frac{d}{dt}$  is changed from  $t$  to  $\mu$ , the tangent vector  $X$  should be replaced by  $X' = \frac{d}{d\mu} = g \frac{d}{dt} = gX$ , and  $\Delta t$  in equation (10.1.1) should be replaced by  $\Delta\mu$ , i.e.,

$$(D_{gX}Y)_p = \lim_{\Delta\mu \rightarrow 0} \frac{Y'_p - Y_p}{\Delta\mu} = \lim_{\Delta t \rightarrow 0} \frac{Y'_p - Y_p}{\frac{1}{g} \Delta t} = g \lim_{\Delta t \rightarrow 0} \frac{Y'_p - Y_p}{\Delta t}. \quad (10.1.2)$$

Thus, we obtain

$$D_{gX}Y = gD_XY.$$

Synthesizing the above discussion, we can arrive at the following general conclusion regarding affine connections.

**Theorem 10.1.1** Let  $M$  be a smooth manifold, and let  $\chi(M)$  denote the set of all smooth tangent vector fields on  $M$ . An **affine connection** (or simply a **connection**)  $D$  on  $M$  is a mapping

$$D: \chi(M) \times \chi(M) \rightarrow \chi(M).$$

$D$  satisfies the following conditions:

Let  $X, Y, Z \in \chi(M)$ , and let  $f, g \in C^\infty(M)$ . Then

- 1)  $D_{fX+gY}Z = fD_XZ + gD_YZ$ ;
- 2)  $D_X(fY + gZ) = fD_XY + gD_XZ + (Xf)Y + (Xg)Z$ .

$D_XY$  is called the **covariant derivative** (or **covariant differential**) of the vector field  $Y$  along  $X$ .

A manifold equipped with a given affine connection is called an **affine connection space**.

Let  $(U; x^i)$  be a local coordinate system on a smooth manifold  $M$ , and define

$$D \frac{\partial}{\partial x^k} = \Gamma_{ik}^j \frac{\partial}{\partial x^j},$$

where  $\Gamma_{ik}^j$  are smooth functions on  $U$ , called the **connection coefficients** of  $D$  with respect to the local coordinates  $x^i$ .

Suppose  $(W; w^i)$  is another coordinate system on  $M$ , and in this coordinate system, the connection coefficients are  $\Gamma_{ik}^j$ . Then

$$\Gamma_{ik}^j = \frac{\partial w^j}{\partial x^q} \frac{\partial x^p}{\partial w^i} \frac{\partial x^r}{\partial w^k} \Gamma_{p r}^q + \frac{\partial^2 x^p}{\partial w^i \partial w^k} \frac{\partial w^j}{\partial x^p}. \quad (10.1.3)$$

This is the transformation formula for the connection coefficients  $\Gamma_{ik}^j$ . This formula shows that  $\Gamma_{ik}^j$  is not a tensor field on  $M$ . Therefore, it is possible to find a coordinate system such that the connection coefficients vanish at a given point  $p \in M$ .

The purpose of introducing a connection  $D$  on a smooth manifold  $M$  is to differentiate tensor fields on the manifold. For any  $X, Y \in \chi(M)$ ,  $X|_U = X^i \frac{\partial}{\partial x^i}$ ,  $Y|_U = Y^j \frac{\partial}{\partial x^j}$ , it follows from the definition of the connection that

$$(D_Y X)|_U = D_{Y^j \frac{\partial}{\partial x^j}} \left( X^i \frac{\partial}{\partial x^i} \right) = Y^j \left( \frac{\partial X^i}{\partial x^j} \frac{\partial}{\partial x^i} + X^i \Gamma_{ij}^k \frac{\partial}{\partial x^k} \right) = Y^j \left( \frac{\partial X^i}{\partial x^j} + X^k \Gamma_{kj}^i \right) \frac{\partial}{\partial x^i}.$$

Let

$$X^i{}_{;j} = \frac{\partial X^i}{\partial x^j} + X^k \Gamma_{kj}^i. \quad (10.1.4)$$

In the local coordinate system  $(U; x^i)$ , the 1-form (cotangent vector field)  $\alpha \in A^1(M)$  is locally expressed as  $\alpha|_U = a_i dx^i$ . Then

$$(D_X \alpha)|_U = X^i \left( \frac{\partial a_j}{\partial x^i} - a_k \Gamma_{ji}^k \right) dx^j.$$

Let

$$a_{j; i} = \frac{\partial a_j}{\partial x^i} - a_k \Gamma_{ji}^k. \quad (10.1.5)$$

Equations (10.1.4) and (10.1.5) correspond respectively to the classical formulas for covariant differentiation of contravariant and covariant vectors in tensor theory. We adopt the convention: a semicolon “;” denotes covariant differentiation, while a comma “,” denotes ordinary partial differentiation.

Using the connection  $D$ , we can take the covariant derivative of any smooth  $(r,s)$ -type tensor field  $A$  on  $M$ , obtaining a tensor field  $DA$  of type  $(r,s+1)$ .

The covariant derivative of a scalar field (i.e., a smooth function) on a smooth manifold  $M$  is defined to be its ordinary derivative.

Below are the covariant derivative formulas for several common types of tensor fields.

The formulas for the covariant derivatives of the second-order tensor fields  $A_{ij}$ ,  $A^{ij}$ , and  $A_j^i$  with respect to  $x^k$  are as follows:

$$A_{ij; k} = A_{ij, k} - A_{pj} \Gamma_{ik}^p - A_{ip} \Gamma_{jk}^p, \quad (10.1.6)$$

$$A^{ij; k} = A^{ij, k} + A^{pj} \Gamma_{pk}^i + A^{ip} \Gamma_{pk}^j, \quad (10.1.7)$$

$$A_j^i{}_{; k} = A_j^i{}_{, k} + A_j^p \Gamma_{pk}^i - A_p^i \Gamma_{jk}^p. \quad (10.1.8)$$

For a third-order mixed tensor field  $A_{rs}^i$  of type  $(1,2)$ , the covariant derivative formula with respect to  $x^k$  is

$$A_{rs}^i{}_{; k} = A_{rs}^i{}_{, k} + A_{rs}^p \Gamma_{pk}^i - A_{ps}^i \Gamma_{rk}^p - A_{rp}^i \Gamma_{sk}^p. \quad (10.1.9)$$

For a third-order mixed tensor field  $A_r^{ij}$  of type  $(2,1)$ , the covariant derivative formula with respect to  $x^k$  is

$$A_r^{ij}{}_{; k} = A_r^{ij}{}_{, k} + A_r^{pj} \Gamma_{pk}^i + A_r^{ip} \Gamma_{pk}^j - A_p^{ij} \Gamma_{rk}^p. \quad (10.1.10)$$

For a fourth-order mixed tensor field  $A_{rs}^{ij}$  of type  $(2,2)$ , the covariant derivative formula with respect to  $x^k$  is

$$A_{rs}^{ij}{}_{; k} = A_{rs}^{ij}{}_{, k} + A_{rs}^{pj} \Gamma_{pk}^i + A_{rs}^{ip} \Gamma_{pk}^j - A_{ps}^{ij} \Gamma_{rk}^p - A_{rp}^{ij} \Gamma_{sk}^p. \quad (10.1.11)$$

By examining these formulas, we can summarize the general rule for taking the covariant derivative  $;x^k$  of a tensor field of arbitrary type:

1) The covariant derivative of an  $m$ -th order tensor field consists of  $m+1$  terms.

2) The first term is the ordinary derivative of the tensor field components with respect to  $x^k$ .

3) For each contravariant index (e.g.,  $i$ ), the corresponding term is constructed as follows:

(a) The term is positive.

(b) Apart from  $i$ , all other indices of the tensor field component remain unchanged. The index  $i$  is moved to become the first upper index of the Christoffel symbol, and fill the vacated position with an index (e.g.,  $p$ ) that matches the lower index of the Christoffel symbol, serving as a summation index. The second lower index of the Christoffel symbol is  $k$ .

4) For each covariant index (e.g.,  $r$ ), the corresponding term is constructed as follows:

(a) The term is negative.

(b) Apart from  $r$ , all other indices of the tensor field component remain unchanged. The index  $r$  is moved to become the first lower index of the Christoffel symbol, and fill the vacated position with an index (e.g.,  $p$ ) that matches the upper index of the Christoffel symbol, serving as a

summation index. The second lower index of the Christoffel symbol is  $k$ .

## §10.2 Geodesics

In the previous section, we discussed that if a tangent vector field  $Y$  undergoes parallel translation along a curve  $\gamma$  whose tangent vector is  $X$ , then

$$D_X Y = 0. \quad (10.2.1)$$

Next, we write the explicit expression for equation (10.2.1).

In a local coordinate system  $(U; x^i)$ , let  $X = \frac{d}{d\lambda} = X^i \frac{\partial}{\partial x^i}$  and  $Y = Y^j \frac{\partial}{\partial x^j}$ . Then

$$\begin{aligned} D_X Y &= D_{X^i \frac{\partial}{\partial x^i}} \left( Y^j \frac{\partial}{\partial x^j} \right) = X^i D_{\frac{\partial}{\partial x^i}} \left( Y^j \frac{\partial}{\partial x^j} \right) \\ &= X^i \left( D_{\frac{\partial}{\partial x^i}} Y^j \right) \frac{\partial}{\partial x^j} + X^i Y^j D_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}. \end{aligned}$$

Since  $Y^j$  is a function, the covariant derivative of a function is simply its ordinary derivative. Thus,

$$X^i \left( D_{\frac{\partial}{\partial x^i}} Y^j \right) = D_{X^i \frac{\partial}{\partial x^i}} Y^j = \frac{dY^j}{d\lambda},$$

and we obtain

$$D_X Y = \frac{dY^j}{d\lambda} \frac{\partial}{\partial x^j} + X^i Y^j D_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \frac{dY^j}{d\lambda} \frac{\partial}{\partial x^j} + X^i Y^j \Gamma_{ji}^k \frac{\partial}{\partial x^k}.$$

Swapping the indices  $j$  and  $k$  in the last term of the above equation yields

$$D_X Y = \frac{dY^j}{d\lambda} \frac{\partial}{\partial x^j} + X^i Y^k \Gamma_{ki}^j \frac{\partial}{\partial x^j} = \left( \frac{dY^j}{d\lambda} + X^i Y^k \Gamma_{ki}^j \right) \frac{\partial}{\partial x^j}. \quad (10.2.2)$$

Finally, the equation  $D_X Y = 0$  becomes

$$\frac{dY^j}{d\lambda} + X^i Y^k \Gamma_{ki}^j = 0. \quad (10.2.3)$$

This is a system of first-order linear ordinary differential equations.

From the system of equations (10.2.3), we can obtain the increments of the components  $A^j(p)$  when the tangent vector  $A$  is parallel transported along the curve  $\gamma$  from point  $p$  to point  $q$ :

$$\delta A^j(p) = A^j(p \rightarrow q) - A^j(p) = -X^i(p) A^k(p) \Gamma_{ki}^j(p). \quad (10.2.4)$$

In a local coordinate system  $(U; x^i)$ , let the parametric equation of the curve  $\gamma$  be  $\gamma: x^i = x^i(\lambda)$ .

Then  $\frac{dx^i}{d\lambda} = X^i$ , and equation (10.2.4) can be written as

$$\delta A^j(p) = A^j(p \rightarrow q) - A^j(p) = -\Gamma_{ki}^j(p) A^k(p) dx^i.$$

By swapping indices, we obtain:

$$\delta A^i(p) = A^i(p \rightarrow q) - A^i(p) = -\Gamma_{jk}^i(p) A^j(p) dx^k. \quad (10.2.5)$$

If the tangent vector  $X$  of the curve  $\gamma$  is parallel along  $\gamma$  itself, then  $\gamma$  is called a **geodesic**. The geodesic equation is

$$D_X X = 0,$$

which implies:

$$\frac{dX^j}{d\lambda} + X^i X^k \Gamma_{ki}^j = 0.$$



Since  $\frac{dx^i}{d\lambda} = X^i$  in the local coordinate system  $(U; x^i)$ , the above equation can be written as

$$\frac{d^2 x^j}{d\lambda^2} + \frac{dx^k}{d\lambda} \frac{dx^i}{d\lambda} \Gamma_{ki}^j = 0.$$

By interchanging the indices  $i$  and  $j$ , we obtain

$$\frac{d^2 x^i}{d\lambda^2} + \frac{dx^k}{d\lambda} \frac{dx^j}{d\lambda} \Gamma_{kj}^i = 0. \quad (10.2.6)$$

This is a system of second-order ordinary differential equations. Therefore, given a point on the manifold  $M$  and a tangent vector at that point, there exists a unique geodesic passing through the point and tangent to the given vector.

**Example 10.2.1** 1) In Euclidean space  $R^m$ , the geodesics are straight lines or segments of straight lines.

2) On the sphere  $S^m$ , the geodesics are great circles or arcs of great circles.

Let a tangent vector field  $Y = Y(\lambda)$  be defined along a curve  $\gamma$  on a smooth manifold  $M$ . In a local coordinate system  $(U; x^i)$ , let the parametric equation of the curve be  $x^i(\gamma(\lambda)) = x^i(\lambda)$ , and let the tangent vector field be

$$Y = Y^j \frac{\partial}{\partial x^j}.$$

Then, according to formula (10.2.2), the covariant derivative of the tangent vector field  $Y = Y(\lambda)$  along the curve  $\gamma = \gamma(\lambda)$  can be defined as

$$\frac{DY(\lambda)}{d\lambda} = D_{\gamma'(\lambda)} Y = \left( \frac{dY^j(\lambda)}{d\lambda} + Y^k \frac{dx^i(\lambda)}{d\lambda} \Gamma_{ki}^j(\gamma(\lambda)) \right) \frac{\partial}{\partial x^j} \Big|_{\gamma(\lambda)}. \quad (10.2.7)$$

## §10.3 Curvature and Torsion

### 1. Curvature Tensor

For any two smooth tangent vector fields  $X$  and  $Y$  on a smooth manifold  $M$ , there is a **curvature operator**  $R(X, Y)$  that maps tangent vector fields on  $M$  to tangent vector fields. It can be expressed as

$$R(X, Y) = D_X D_Y - D_Y D_X - D_{[X, Y]}. \quad (10.3.1)$$

If  $R(X, Y)$  acts on a tangent vector field  $Z$ , it is expressed as

$$R(X, Y)Z = D_X D_Y Z - D_Y D_X Z - D_{[X, Y]} Z. \quad (10.3.2)$$

Let the local coordinate representations of the tangent vector fields  $X$ ,  $Y$ , and  $Z$  be

$$X = X^i \frac{\partial}{\partial x^i}, \quad Y = Y^i \frac{\partial}{\partial x^i}, \quad Z = Z^i \frac{\partial}{\partial x^i},$$

then

$$R(X, Y)Z = R_{i\ kl}^j Z^i X^k Y^l \frac{\partial}{\partial x^j}, \quad (10.3.3)$$

where

$$R_{i\ kl}^j = \left\langle R \left( \frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l} \right) \frac{\partial}{\partial x^i}, dx^j \right\rangle, \quad (10.3.4)$$

or

$$R_{i\ kl}^j = \frac{\partial \Gamma_{i\ l}^j}{\partial x^k} - \frac{\partial \Gamma_{i\ k}^j}{\partial x^l} + \Gamma_{i\ l}^h \Gamma_{h\ k}^j - \Gamma_{i\ k}^h \Gamma_{h\ l}^j. \quad (10.3.5)$$

$R_{i\ kl}^j$  is a tensor of type (1,3), called the **Riemann curvature tensor**.

It should be noted that some literature defines  $R_{ikl}^j$  with a sign opposite to that of equation (10.3.5).

Let us analyze the geometric meaning of the Riemann curvature tensor. As shown in Figure 10.3.1, consider a closed path  $o \rightarrow p \rightarrow s \rightarrow q \rightarrow o$  in the local coordinate system  $(U; x^i)$  of a smooth manifold  $M$ . We compute the change in a tangent vector  $A^i(o)$  after parallel transporting it once around this closed path.

Let the tangent vector at point  $o$  be  $A^i(o)$ . After parallel transporting  $A^i(o)$  from point  $o$  to point  $p$ , denote the resulting tangent vector as  $A^i(p)$ . From equation (10.2.5), we obtain

$$A^i(p) = A^i(o) - \Gamma_{jk}^i(o) A^j(o) dx^k. \quad (10.3.6)$$

Let the tangent vector after parallel transporting  $A^i(p)$  from point  $p$  to point  $s$  be denoted as  $A^i(s)$ . Then

$$A^i(s) = A^i(p) - \Gamma_{lh}^i(p) A^l(p) \delta x^h. \quad (10.3.7)$$

$\Gamma_{lh}^i(p)$  can be approximated as

$$\Gamma_{lh}^i(p) = \Gamma_{lh}^i(o) + \Gamma_{lmh}^i(o) dx^m. \quad (10.3.8)$$

Substituting equations (10.3.6) and (10.3.8) into (10.3.7) yields

$$A^i(s) = A^i(o) - \Gamma_{jk}^i(o) A^j(o) dx^k - [\Gamma_{lh}^i(o) + \Gamma_{lmh}^i(o) dx^m] [A^l(o) - \Gamma_{jk}^l(o) A^j(o) dx^k] \delta x^h.$$

Expanding and rearranging the above expression, and omitting the letter  $o$ , we obtain

$$A^i(s) = A^i - \Gamma_{jk}^i A^j dx^k - \Gamma_{lh}^i A^l \delta x^h - \Gamma_{lh, m}^i A^l dx^m \delta x^h + \Gamma_{lh}^i \Gamma_{jk}^l A^j dx^k \delta x^h + O[(dx^k)^3],$$

where  $O[(dx^k)^3]$  is a function of  $(dx^k)^3$ . Therefore, the increment of the tangent vector  $A^i(o)$  when transported along the path  $o \rightarrow p \rightarrow s$  to point  $s$  is

$$\begin{aligned} \delta A^i(o \rightarrow p \rightarrow s) &= A^i(s) - A^i(o) = -\Gamma_{jk}^i A^j dx^k - \Gamma_{lh}^i A^l \delta x^h \\ &\quad - \Gamma_{lh, m}^i A^l dx^m \delta x^h + \Gamma_{lh}^i \Gamma_{jk}^l A^j dx^k \delta x^h + O[(dx^k)^3]. \end{aligned} \quad (10.3.9)$$

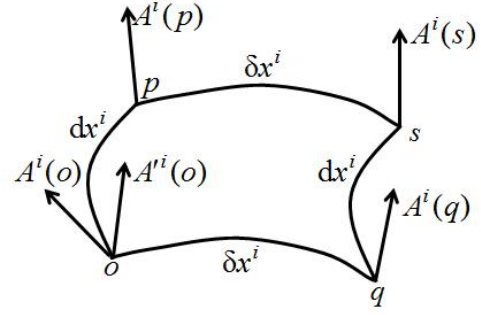
If we exchange  $dx$  and  $\delta x$  in equation (10.3.9), we obtain the increment of the tangent vector  $A^i(o)$  when transported along the path  $o \rightarrow q \rightarrow s$  to point  $s$ :

$$\begin{aligned} \delta A^i(o \rightarrow q \rightarrow s) &= -\Gamma_{jk}^i A^j \delta x^k - \Gamma_{lh}^i A^l dx^h \\ &\quad - \Gamma_{lh, m}^i A^l \delta x^m dx^h + \Gamma_{lh}^i \Gamma_{jk}^l A^j \delta x^k dx^h + O[(\delta x^k)^3]. \end{aligned}$$

The increment of the tangent vector  $A^i(o)$  when parallel transported once around the closed path  $o \rightarrow p \rightarrow s \rightarrow q \rightarrow o$  back to point  $o$  equals the increment along the path  $o \rightarrow p \rightarrow s$  plus the increment along the path  $s \rightarrow q \rightarrow o$ . That is,

$$\delta A^i(o \rightarrow p \rightarrow s \rightarrow q \rightarrow o) = \delta A^i(o \rightarrow p \rightarrow s) + \delta A^i(s \rightarrow q \rightarrow o).$$

It can be verified that, neglecting higher-order infinitesimals, the increment when parallel transporting from point  $s$  to point  $o$  is equal in magnitude but opposite in sign to the increment when transporting from point  $o$  to point  $s$ :



**Figure 10.3.1** The tangent vector  $A^i(o)$  changes after being parallel transported once around the closed path

$$\delta A^i(s \rightarrow q \rightarrow o) = -\delta A^i(o \rightarrow q \rightarrow s),$$

Thus, the total increment of the tangent vector  $A^i(o)$  when parallel transported once around the closed path back to point  $o$  is

$$\begin{aligned} \delta A^i &= \delta A^i(o \rightarrow p \rightarrow s) + \delta A^i(s \rightarrow q \rightarrow o) \\ &= \delta A^i(o \rightarrow p \rightarrow s) - \delta A^i(o \rightarrow q \rightarrow s) \\ &= -\Gamma_{j k}^i A^j dx^k - \Gamma_{l h}^i A^l \delta x^h - \Gamma_{l h, m}^i A^l dx^m \delta x^h + \Gamma_{l h}^i \Gamma_{j k}^l A^j dx^k \delta x^h \\ &\quad + \Gamma_{j k}^i A^j \delta x^k + \Gamma_{l h}^i A^l dx^h + \Gamma_{l h, m}^i A^l \delta x^m dx^h - \Gamma_{l h}^i \Gamma_{j k}^l A^j \delta x^k dx^h \\ &= -\Gamma_{l h, m}^i A^l dx^m \delta x^h + \Gamma_{l h, m}^i A^l dx^h \delta x^m \\ &\quad - \Gamma_{l h}^i \Gamma_{j k}^l A^j dx^h \delta x^k + \Gamma_{l h}^i \Gamma_{j k}^l A^j dx^k \delta x^h. \end{aligned}$$

In the first term, replace  $m$  with  $k$ ,  $l$  with  $j$ ; in the second term, replace  $h$  with  $k$ ,  $m$  with  $h$ ,  $l$  with  $j$ ; and in the third term, replace  $h$  with  $k$ ,  $k$  with  $h$ . This yields

$$\begin{aligned} \delta A^i &= -\Gamma_{j h, k}^i A^j dx^k \delta x^h + \Gamma_{j k, h}^i A^j dx^k \delta x^h \\ &\quad - \Gamma_{l k}^i \Gamma_{j h}^l A^j dx^k \delta x^h + \Gamma_{l h}^i \Gamma_{j k}^l A^j dx^k \delta x^h \\ &= -[\Gamma_{j h, k}^i - \Gamma_{j k, h}^i + \Gamma_{l k}^i \Gamma_{j h}^l - \Gamma_{l h}^i \Gamma_{j k}^l] A^j dx^k \delta x^h \\ &= -R_{j kh}^i A^j dx^k \delta x^h, \end{aligned}$$

i.e.,

$$\delta A^i = -R_{j kh}^i A^j dx^k \delta x^h. \quad (10.3.10)$$

As shown in Figure 10.3.1, if  $\delta A^i \neq 0$ , then after the tangent vector  $A^i(o)$  is parallel transported once around the closed path  $o \rightarrow p \rightarrow s \rightarrow q \rightarrow o$  back to point  $o$ , it becomes  $A'^i(o)$ , and  $A'^i(o)$  does not coincide with  $A^i(o)$ ; they are two different tangent vectors. This phenomenon can only occur if the region enclosed by the path is not flat.

In general, if  $\delta A^i \neq 0$ , it indicates that the smooth manifold is curved in the region  $U$  covered by the local coordinate system  $(U; x^i)$ . If  $\delta A^i = 0$ , it indicates that the manifold is flat in the region  $U$ . Since  $A^j$ ,  $dx^k$  and  $\delta x^h$  are nonzero, when  $\delta A^i \neq 0$ ,  $R_{j kh}^i \neq 0$ ; and when  $\delta A^i = 0$ ,  $R_{j kh}^i = 0$ . Therefore,  $R_{j kh}^i$  captures the geometric property of whether the smooth manifold is locally curved.

## 2. Torsion Tensor

Although the connection coefficients  $\Gamma_{i k}^j$  do not transform like tensor components, if we define

$$T_{i k}^j = \Gamma_{k i}^j - \Gamma_{i k}^j, \quad (10.3.11)$$

then from the transformation formula (10.1.3) for the connection coefficients  $\Gamma_{i k}^j$ , we obtain

$$T_{i k}^{'j} = \frac{\partial w^j}{\partial x^q} \frac{\partial x^p}{\partial w^i} \frac{\partial x^r}{\partial w^k} T_{p r}^q. \quad (10.3.12)$$

From equation (10.3.12), it can be seen that  $T_{i k}^{'j}$  satisfies the transformation rule for components of a tensor of type (1,2). Therefore,

$$T = T_{i k}^j \frac{\partial}{\partial x^j} \otimes dx^i \otimes dx^k \quad (10.3.13)$$

is a tensor field of type (1,2), called the **torsion tensor** of the affine connection  $D$ .

Let  $X$  and  $Y$  be any two smooth tangent vector fields on the smooth manifold  $M$ . Then the

torsion tensor field  $T(X, Y)$  acts as a tangent vector field on  $M$ , and

$$T(X, Y) = D_X Y - D_Y X - [X, Y]. \quad (10.3.14)$$

**Definition 10.3.1** If the torsion tensor of an affine connection  $D$  is zero, the connection is said to be **torsion-free**.

From equation (10.3.11), it follows that a torsion-free connection is symmetric, i.e.,

$$\Gamma_{ik}^j = \Gamma_{ki}^j. \quad (10.3.15)$$

Torsion-free connections always exist, and they possess the following favorable property as stated in the theorem below.

**Theorem 10.3.1** Let  $D$  be a torsion-free affine connection on an  $m$ -dimensional smooth manifold  $M$ . Then, at any point  $p$  on  $M$ , there exists a local coordinate system  $(U; x^i)$  such that the corresponding connection coefficients  $\Gamma_{ik}^j$  vanish at that point, i.e.,

$$\Gamma_{ik}^j(p) = 0, \quad 1 \leq i, j, k \leq m.$$

**Theorem 10.3.2** Let  $D$  be a torsion-free affine connection on a smooth manifold  $M$ . Then the following Bianchi identity holds:

$$R_{ikl}^j + R_{ilh}^j + R_{ihk}^j = 0. \quad (10.3.16)$$

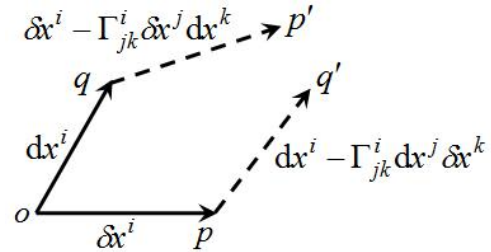
Finally, let us understand the geometric meaning of the torsion tensor. As shown in Figure 10.3.2, on a smooth manifold  $M$ , when the tangent vector  $dx^i$  is translated a small distance  $\delta x^i$  to the right, it becomes  $dx^i - \Gamma_{jk}^i dx^j \delta x^k$ , point  $o$  moves to point  $p$ , and point  $q$  moves to point  $q'$ . When the tangent vector  $\delta x^i$  is translated a small distance  $dx^i$  upward, it becomes  $\delta x^i - \Gamma_{jk}^i \delta x^j dx^k$ , point  $o$  moves to point  $q$ , and point  $p$  moves to point  $p'$ . The difference between the positions of points  $p'$  and  $q'$  is

$$\begin{aligned} \Delta &= [dx^i + (\delta x^i - \Gamma_{jk}^i \delta x^j dx^k)] - [\delta x^i + (dx^i - \Gamma_{jk}^i dx^j \delta x^k)] \\ &= (\Gamma_{jk}^i - \Gamma_{kj}^i) dx^j \delta x^k = T_{kj}^i dx^j \delta x^k. \end{aligned} \quad (10.3.17)$$

According to equation (10.3.17), if the torsion tensor of the smooth manifold  $M$  satisfies  $T_{kj}^i = 0$ , then  $\Delta = 0$ , meaning points  $p'$  and  $q'$  coincide. Consequently, the four tangent vectors in Figure 10.3.2 form a closed quadrilateral, and in this case, the smooth manifold  $M$  is said to be **torsion-free**.

Conversely, if  $T_{kj}^i \neq 0$ , the four tangent vectors in Figure 10.3.2 cannot form a closed quadrilateral, and the smooth manifold  $M$  is

said to have **torsion** or be **twisted**. The torsion tensor  $T_{kj}^i$  measures the degree of this twisting. In Figure 10.3.1, the quadrilateral is drawn as closed, which implicitly assumes that the torsion of the smooth manifold  $M$  is zero.



**Figure 10.3.2** If torsion is nonzero, a closed quadrilateral cannot be formed

# Chapter 11 Riemannian Manifolds

## §11.1 Metric Tensor

### 1. Metric Tensor and Riemannian Metric

Let  $M$  be an  $m$ -dimensional smooth manifold. We define on  $M$  a **symmetric second-order covariant tensor field**  $g$ . If  $(U; x^i)$  is a local coordinate system on  $M$ , the tensor field  $g$  can be expressed on  $U$  as

$$g = g_{ij} dx^i \otimes dx^j, \quad (11.1.1)$$

where  $g_{ij} = g_{ji}$  are smooth functions on  $U$ . At each point  $p \in M$ ,  $g$  gives a bilinear function (or linear mapping) on the tangent space  $T_p(M)$ :

$$g : T_p M \times T_p M \rightarrow R.$$

Let  $X = X^i \frac{\partial}{\partial x^i}$ ,  $Y = Y^i \frac{\partial}{\partial x^i}$ , and define

$$g(X, Y) = g_{ij} X^i Y^j. \quad (11.1.2)$$

If  $X^i = 1$ ,  $Y^i = 1$ , then

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right). \quad (11.1.3)$$

The smooth functions  $g_{ij}$  form a matrix called the **metric matrix**  $(g_{ij})$ :

$$(g_{ij}) = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix}.$$

**Definition 11.1.1** If there exists a tangent vector  $X \in T_p M$  such that

$$g(X, Y) = 0$$

for all  $Y \in T_p M$ , implies that  $X = 0$ , then the tensor  $g$  is said to be **non-degenerate** at point  $p$ .

According to equation (11.1.2), writing  $g(X, Y) = 0$  in matrix form yields

$$g(X, Y) = (Y^1, Y^2, \dots, Y^m) \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix} \begin{pmatrix} X^1 \\ X^2 \\ \vdots \\ X^m \end{pmatrix} = 0. \quad (11.1.4)$$

For  $g(X, Y) = 0$  to hold for all  $Y \in T_p M$ , it is necessary that

$$\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix} \begin{pmatrix} X^1 \\ X^2 \\ \vdots \\ X^m \end{pmatrix} = 0.$$

Requiring that this system of equations has only the trivial solution  $X = 0$  implies that the determinant must satisfy

$$\det(g_{ij}(p)) \neq 0.$$

Therefore,  $g$  is non-degenerate at point  $p$  if and only if  $\det(g_{ij}(p)) \neq 0$ .

**Definition 11.1.2** If for any  $X \in T_p M$ , we have

$$g(X, X) \geq 0,$$

and equality holds only when  $X = 0$ , then the tensor  $g$  is said to be **positive definite** at point  $p$ .

Replacing  $Y$  with  $X$  in equation (11.1.4), we can express  $g(X, X)$  in matrix form. Since the smooth functions  $g_{ij}$  take numerical values at point  $p$ ,  $g(X, X)$  at point  $p$  is a quadratic form. From linear algebra, it is known that  $g$  is positive definite if and only if the matrix  $(g_{ij})$  is positive definite, and thus  $\det(g_{ij}(p)) > 0$ . Therefore, a positive definite tensor  $g$  is necessarily non-degenerate.

**Definition 11.1.3** If an everywhere non-degenerate, smooth, symmetric, second-order covariant tensor field  $g$  is given on an  $m$ -dimensional smooth manifold  $M$ , then  $M$  is called a **generalized Riemannian manifold**, denoted as  $(M, g)$ . The tensor  $g$  is called the **metric tensor** or **fundamental tensor** of the generalized Riemannian manifold  $M$ .

If  $g$  is positive definite, then  $M$  is called a **Riemannian manifold**, also denoted as  $(M, g)$ .

When the context is clear, generalized Riemannian manifolds or Riemannian manifolds may be simply denoted as  $M$ .

Similar to defining an inner product in a Euclidean vector space, equation (11.1.2) defines an inner product on the tangent space  $T_p M$  at each point  $p$  of the smooth manifold  $M$ . Specifically, for  $X, Y \in T_p M$ , we define

$$X \cdot Y = g(X, Y) = g_{ij}(p) X^i Y^j. \quad (11.1.5)$$

When  $X$  and  $Y$  are smooth tangent vector fields, the inner product  $X \cdot Y$  is a smooth function on  $M$ . For a generalized Riemannian manifold  $(M, g)$ ,  $g$  is not necessarily positive definite, but for a Riemannian manifold,  $g$  is positive definite.

If  $g$  is positive definite, we can compute the length of tangent vectors, the angle between two tangent vectors at the same point, as well as arc lengths, areas, volumes, etc. The length of a tangent vector  $X$  is

$$|X| = \sqrt{g_{ij} X^i X^j}. \quad (11.1.6)$$

The angle between two tangent vectors  $X$  and  $Y$  at the same point is

$$\cos \angle(X, Y) = \frac{X \cdot Y}{|X| \cdot |Y|}. \quad (11.1.7)$$

If  $X$  and  $Y$  are infinitesimal tangent vectors, let  $X^i = dx^i$ ,  $Y^i = dx^i$ . From equation (11.1.5), we obtain

$$ds^2 = g_{ij} dx^i dx^j. \quad (11.1.8)$$

This expression  $ds^2$  is called the **metric form** or **Riemannian metric**. It is a quadratic differential form independent of the choice of local coordinates  $x^i$ , meaning  $ds^2$  is invariant under coordinate transformations.

If the metric tensor  $g$  is positive definite, then  $ds^2 > 0$ , and  $ds$  represents the length of an infinitesimal tangent vector, known as the **arc length element**. With the arc length element, we can compute the length of a curve. In a local coordinate system  $(U; x^i)$ , let  $\gamma: x^i = x^i(t)$ ,  $a \leq t \leq b$  be a smooth parameterized curve on  $M$ . The arc length of  $\gamma$  is defined as

$$s = \int_a^b \frac{ds}{dt} dt = \int_a^b \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt. \quad (11.1.9)$$

If the metric tensor  $g$  is not positive definite,  $ds^2$  does not have a definite sign. In such cases, further analysis is required depending on whether  $ds^2$  is positive, negative, or zero.

**Theorem 11.1.1** On an  $m$ -dimensional smooth manifold  $M$ , there always exists a Riemannian metric.

On an  $m$ -dimensional smooth manifold  $M$ , a positive definite Riemannian metric always

exists, but a non-positive definite Riemannian metric may not necessarily exist. The existence of a positive definite Riemannian metric on  $M$  implies that there must exist a positive definite metric tensor field on  $M$ .

For a generalized Riemannian manifold  $(M, g)$ ,  $g$  is a symmetric second-order covariant tensor field. At each point  $p \in M$ ,  $g$  is a symmetric second-order covariant tensor. Therefore, when the local coordinate system changes from  $(U; x^i)$  to  $(V; y^i)$ , the transformation formula for the components of the metric tensor  $g$  is

$$g'_{ij} = g_{kl} \frac{\partial x^k}{\partial y^i} \frac{\partial x^l}{\partial y^j}. \quad (11.1.10)$$

Since the matrix  $(g_{ij})$  is non-degenerate,  $\det(g_{ij}(p)) \neq 0$ , and thus the inverse matrix exists.

We denote the elements of its inverse matrix as  $g^{ij}$ , so that

$$\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix} \begin{pmatrix} g^{11} & g^{12} & \cdots & g^{1m} \\ g^{21} & g^{22} & \cdots & g^{2m} \\ \cdots & \cdots & \cdots & \cdots \\ g^{m1} & g^{m2} & \cdots & g^{mm} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix},$$

i.e.,

$$g^{ik} g_{kj} = g_{jk} g^{ki} = \delta_j^i. \quad (11.1.11)$$

Here,  $(g^{ij})$  is a symmetric second-order contravariant tensor. The coordinate transformation formula for  $g^{ij}$  is

$$g'^{ij} = g^{kl} \frac{\partial y^i}{\partial x^k} \frac{\partial y^j}{\partial x^l}. \quad (11.1.12)$$

## 2. Levi-Civita Connection

**Definition 11.1.4** Let  $D$  be an affine connection on an  $m$ -dimensional generalized Riemannian manifold  $(M, g)$ . If

$$Dg = 0, \quad (11.1.13)$$

then  $D$  is called an **compatible connection** of the generalized Riemannian manifold  $(M, g)$ .

For an compatible connection  $D$ , the covariant derivative of the metric tensor  $g_{ij}$  vanishes, i.e.,

$$g_{ij; k} = 0. \quad (11.1.14)$$

The geometric meaning of an compatible connection  $D$  is that the metric properties remain invariant under parallel transport, i.e.,

$$Dds^2 = 0. \quad (11.1.15)$$

This parallel transport preserves the inner product of two vector fields. Therefore, on a Riemannian manifold, the lengths and angles of tangent vectors remain unchanged during parallel transport.

**Theorem 11.1.2 (Fundamental Theorem of Riemannian Geometry)** On an  $m$ -dimensional generalized Riemannian manifold  $(M, g)$ , there exists a unique torsion-free compatible connection. This connection is called the **Levi-Civita connection** or **Riemannian connection** of the generalized Riemannian manifold  $(M, g)$ .

The torsion-free compatible connection is uniquely determined by the metric tensor. In a local coordinate system  $(U; x^i)$ , it is given by

$$\Gamma_{ikj} = \frac{1}{2} \left( \frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} \right), \quad (11.1.16)$$

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl} \left( \frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \right). \quad (11.1.17)$$

Here,  $\Gamma_{ikj}$  and  $\Gamma_{ij}^k$  are called the **first kind** and **second kind** of **Christoffel symbols**, respectively.

### 3.A Simplified Method for Calculating Christoffel Symbols

Assume that when  $i \neq j$ , we have  $g_{ij} = 0$  and  $g^{ij} = 0$ . Then, the following simplified methods can be used to calculate  $\Gamma_{ij}^k$ :

1) If all three indices in  $\Gamma_{ij}^k$  are the same, then

$$\Gamma_{ii}^i = \frac{1}{2} g^{ii} \frac{\partial g_{ii}}{\partial x^i},$$

where the index  $i$  is not summed.

2) If in  $\Gamma_{ij}^k$ , we have  $k = i$ , then

$$\Gamma_{ij}^i = \frac{1}{2} g^{ii} \frac{\partial g_{ij}}{\partial x^j},$$

where the index  $i$  is not summed.

3) If in  $\Gamma_{ij}^k$ , the two lower indices are the same, i.e.,  $i = j$ , then

$$\Gamma_{ii}^k = -\frac{1}{2} g^{kk} \frac{\partial g_{ii}}{\partial x^k},$$

where the index  $i, k$  is not summed.

4) If all three indices in  $\Gamma_{ij}^k$  are distinct, i.e.,  $i \neq j \neq k$ , then

$$\Gamma_{ij}^k = 0.$$

### 4.Koszul Formula

Let  $D$  be the Levi-Civita connection on an  $m$ -dimensional Riemannian manifold  $M$ . Then, for any tangent vector fields  $X, Y, Z \in \chi(M)$ , the following holds:

$$2 \langle D_X Y, Z \rangle = \langle X \langle Y, Z \rangle + Y \langle Z, X \rangle - Z \langle X, Y \rangle + \langle [X, Y], Z \rangle + \langle [Z, X], Y \rangle - \langle [Y, Z], X \rangle. \quad (11.1.18)$$

This equation is known as the **Koszul formula**. Here,  $\langle \cdot, \cdot \rangle$  denotes the inner product symbol,  $D_X Y$  is a tangent vector field on  $M$ , and  $\langle D_X Y, Z \rangle$  denotes the inner product of the two tangent vector fields  $D_X Y$  and  $Z$ . Equation (11.1.18) implies equation (11.1.17).

If an orthonormal frame field  $X_1, \dots, X_m$  is given on an open subset  $U$  of the Riemannian manifold  $M$ , i.e., everywhere satisfying

$$\langle X_i, X_j \rangle = 0, \quad i \neq j,$$

then from equation (11.1.18), we obtain

$$2 \langle D_{X_i} X_j, X_k \rangle = \langle [X_i, X_j], X_k \rangle + \langle [X_k, X_i], X_j \rangle - \langle [X_j, X_k], X_i \rangle, \quad 1 \leq i, j, k \leq m. \quad (11.1.19)$$

### 5.Exponential Map

**Definition 11.1.5** Let  $M$  be an  $m$ -dimensional Riemannian manifold. If a parameterized curve  $\gamma$  is a geodesic with respect to the Levi-Civita connection, then  $\gamma$  is called a **geodesic** of the Riemannian manifold  $M$ .

In a general Riemannian manifold, geodesics are only locally the shortest paths.

**Theorem 11.1.3** For any point  $p_0 \in M$  of a Riemannian manifold  $M$ , there exists a neighborhood  $U$  of  $p_0$  and a positive number  $\varepsilon > 0$  such that for each point  $p \in U$  and each tangent vector  $X \in T_p M$  with  $|X| < \varepsilon$ , there exists a unique geodesic

$$\gamma_X : (-2, 2) \rightarrow M,$$



satisfying the following conditions:

$$\gamma_X(0) = p, \quad \gamma'_X(0) = X. \quad (11.1.20)$$

Let  $X \in T_p M$  and suppose there exists a geodesic  $\gamma: [0, 1] \rightarrow M$  such that

$$\gamma(0) = p, \quad \gamma'(0) = X. \quad (11.1.21)$$

Then denote the point on the geodesic at  $\gamma(1)$  as  $\exp_p(X)$ , i.e.,

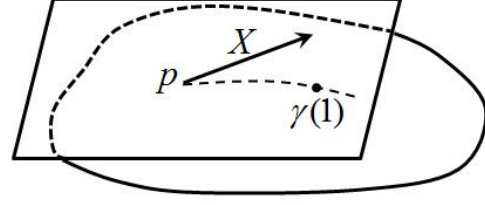
$$\exp_p(X) = \gamma(1). \quad (11.1.22)$$

Equation (11.1.22) defines a mapping from a neighborhood  $B \subset T_p M$  of the origin in  $T_p M$  to  $M$ :

$$\exp_p : B \rightarrow M. \quad (11.1.23)$$

**Definition 11.1.6** The mapping defined by equation (11.1.23) is called the **exponential map** with respect to point  $p$ .

Its geometric meaning is as follows: As shown in Figure 11.1.1, the exponential map  $\exp_p$  maps the tangent vector  $X \in T_p M$  to the point on the geodesic with initial conditions  $\gamma(0) = p$  and  $\gamma'(0) = X$ , such that the arc length from point  $p$  to this point equals  $|X|$ . In other words, the arc length along the geodesic  $\gamma$  from point  $p$  to point  $\gamma(1) = \exp_p(X)$  equals



**Figure 11.1.1** Exponential map

$|X|$ . Alternatively, the image of the tangent vector  $X \in T_p M$  under the exponential map  $\exp_p$  is the geodesic segment between point  $p$  and point  $\gamma(1) = \exp_p(X)$ , and the arc length between these two points equals the length  $|X|$  of the tangent vector  $X$ .

**Theorem 11.1.4** Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold. Then for any point  $p \in M$ , there exists an open neighborhood  $U$  of the origin (zero vector) in the tangent space  $T_p M$ , such that the exponential map  $\exp_p$  is a diffeomorphism from  $U$  onto an open subset  $V = \exp_p(U)$  of the Riemannian manifold  $M$ .

According to Theorem 11.1.3, when the length  $|X|$  of the tangent vector  $X$  is sufficiently small, the exponential map  $\exp_p(X)$  is well-defined and meaningful. However, when  $|X|$  is large, the exponential map  $\exp_p(X)$  may not necessarily be meaningful.

**Theorem 11.1.5** Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold,  $p \in M$ , and suppose that for a tangent vector  $X \in T_p M$ ,  $\exp_p(X)$  is meaningful. Then for every parameter  $t$ ,  $|t| \leq 1$ , the exponential map  $\exp_p(tX)$  is meaningful, and  $\gamma(t) = \exp_p(tX)$  is a geodesic satisfying the following conditions:

$$\gamma(0) = p, \quad \gamma'(0) = X. \quad (11.1.24)$$

## §11.2 Curvature

### 1. Curvature Tensor

Using the metric tensor, we can raise or lower the indices of a tensor. Let

$$X = X^i \frac{\partial}{\partial x^i}, \quad a = X_i dx^i.$$

Then we obtain

$$X_i = g_{ij} X^j, \quad X^j = g^{ij} X_i.$$

For  $R_{ikl}^h$ , we have

$$R_{ijkl} = R_{ikl}^h g_{hj},$$

where

$$R_{ijkl} = \frac{\partial \Gamma_{ijl}}{\partial x^k} - \frac{\partial \Gamma_{ijk}}{\partial x^l} + \Gamma_{ik}^h \Gamma_{jhl} - \Gamma_{il}^h \Gamma_{jhk}. \quad (11.2.1)$$

$R_{ijkl}$  is a **fourth-order covariant tensor**, called the **curvature tensor** of the generalized Riemannian manifold  $M$ .

**Note:** In this book, when lowering the upper index  $h$  of  $R_{ikl}^h$  to a lower index,  $h$  is placed in the second position among the lower indices.

**Theorem 11.2.1** The curvature tensor  $R_{ijkl}$  of a generalized Riemannian manifold  $(M, g)$  satisfies the following relations:

- 1) Antisymmetry  $R_{ijkl} = -R_{jikl} = -R_{ijlk}$ ;
- 2)  $R_{ijkl} + R_{iklj} + R_{iljk} = 0$ ;
- 3) Symmetry  $R_{ijkl} = R_{klij}$ ;
- 4) Bianchi identity

$$R_{ijkl;h} + R_{ijlh;k} + R_{ijhk;l} = 0.$$

## 2. Sectional Curvature

The curvature tensor  $R$  of an  $m$ -dimensional Riemannian manifold  $M$  is a tensor that captures the degree of "curvature" of the Riemannian manifold.  $R$  characterizes the intrinsic nature of the Riemannian manifold, being a fourth-order covariant tensor and also a linear function defined on the space of fourth-order contravariant tensors:

$$R : T_p(M) \times T_p(M) \times T_p(M) \times T_p(M) \rightarrow R.$$

In a local coordinate system  $(U; x^i)$ , the curvature tensor  $R$  can be expressed as

$$R = R_{ijkl} dx^i \otimes dx^j \otimes dx^k \otimes dx^l.$$

Let

$$X = X^i \frac{\partial}{\partial x^i}, \quad Y = Y^i \frac{\partial}{\partial x^i}, \quad Z = Z^i \frac{\partial}{\partial x^i}, \quad W = W^i \frac{\partial}{\partial x^i},$$

then

$$R(X, Y, Z, W) = R_{ijkl} X^i Y^j Z^k W^l, \quad (11.2.2)$$

where

$$R_{ijkl} = R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l}\right). \quad (11.2.3)$$

From equation (10.3.3), we have

$$R(Z, W)X = R_{i \quad kl}^j X^i Z^k W^l \frac{\partial}{\partial x^j}.$$

Since the connection  $D$  is the Levi-Civita connection, the inner product " $\bullet$ " of tangent vectors can be defined, thus yielding

$$R(X, Y, Z, W) = (R(Z, W)X) \cdot Y. \quad (11.2.4)$$

Using the metric tensor  $g$  of  $M$ , we can define the following quadrilinear function:

$$g(X, Y, Z, W) = g(X, Z)g(Y, W) - g(X, W)g(Y, Z).$$

Let  $X, Y \in T_p M$ , then

$$\begin{aligned} g(X, Y, X, Y) &= g(X, X)g(Y, Y) - g(X, Y)g(X, Y) \\ &= |X|^2 \cdot |Y|^2 - (X \cdot Y)^2 = |X|^2 \cdot |Y|^2 - |X|^2 \cdot |Y|^2 \cos^2 \angle(X, Y) \\ &= |X|^2 \cdot |Y|^2 \sin^2 \angle(X, Y). \end{aligned}$$

Therefore, for a Riemannian manifold,  $g(X, Y, X, Y)$  is exactly equal to the square of the area of

the parallelogram spanned by the tangent vectors  $X$  and  $Y$ . When  $X$  and  $Y$  are linearly independent (hence neither collinear nor parallel),  $g(X, Y, X, Y) \neq 0$ .

If  $X'$  and  $Y'$  are any two non-collinear (linearly independent) tangent vectors in the two-dimensional subspace  $E$  of the tangent space at point  $p \in M$ , then

$$X' = aX + bY, \quad Y' = cX + dY,$$

where  $ad - bc \neq 0$ . It can be verified that the following holds:

$$g(X', Y', X', Y') = (ad - bc)^2 g(X, Y, X, Y).$$

From the symmetry and antisymmetry properties of the curvature tensor  $R(X, Y, Z, W)$ , we obtain

$$R(X', Y', X', Y') = (ad - bc)^2 R(X, Y, X, Y).$$

Therefore,

$$K(E) = -\frac{R(X', Y', X', Y')}{g(X', Y', X', Y')} = -\frac{R(X, Y, X, Y)}{g(X, Y, X, Y)},$$

which shows that the quantity  $K(E)$  is independent of the choice of basis for the two-dimensional subspace  $E$ .

**Definition 11.2.1** Let  $E$  be a two-dimensional subspace of the tangent space  $T_p M$  of a Riemannian manifold  $M$  at point  $p \in M$ , and let  $X$  and  $Y$  be any two linearly independent tangent vectors in  $E$ . Then

$$K(E) = K(X, Y) = -\frac{R(X, Y, X, Y)}{g(X, Y, X, Y)}$$

is a function of the two-dimensional subspace  $E$ , independent of the choice of  $X$  and  $Y$  in  $E$ . It is called the **sectional curvature** of the Riemannian manifold  $M$  at point  $p$  for the subspace  $(p, E)$ .

**Example 11.2.1** The tangent space of a surface in three-dimensional Euclidean space is two-dimensional. Therefore, the surface has only one sectional curvature. Let two linearly independent tangent vectors in the tangent plane at point  $p$  on the surface be  $X = \frac{\partial}{\partial x}$  and  $Y = \frac{\partial}{\partial y}$ . Then, from equation (11.2.2), we have

$$R(X, Y, X, Y) = R_{1212}.$$

From equation (11.1.2), we obtain

$$g(X, X) = g_{11}, \quad g(Y, Y) = g_{22}, \quad g(X, Y) = g(Y, X) = g_{12},$$

$$g(X, Y, X, Y) = g(X, X)g(Y, Y) - g(X, Y)g(Y, X) = g_{11}g_{22} - g_{12}^2.$$

Therefore, the sectional curvature of the surface at point  $p$  is

$$K = -\frac{R_{1212}}{g_{11}g_{22} - g_{12}^2}.$$

This sectional curvature is precisely the **total curvature** or **Gaussian curvature** of the surface at point  $p$ .

**Definition 11.2.2** If the sectional curvature  $K(E)$  of a Riemannian manifold  $M$  at point  $p$  is independent of the choice of the two-dimensional subspace  $E$  and is constant, denoted as  $K(p)$ , then  $M$  is said to be **isotropic** at point  $p$ .

In local coordinates, the condition for a Riemannian manifold  $M$  to be isotropic at point  $p$  is

$$R_{ijkl}(p) = -K(p)(g_{ik}g_{jl} - g_{il}g_{jk})(p). \quad (11.2.5)$$

**Definition 11.2.3** If a Riemannian manifold  $M$  is isotropic everywhere and the sectional curvature  $K(p)$  is a constant function on  $M$ , then  $M$  is called a **space of constant curvature** (or a **constant curvature manifold**).

**Theorem 11.2.2 (F. Schur's Theorem)** If an  $m$ -dimensional Riemannian manifold  $M$  is connected, isotropic everywhere, and  $m \geq 3$ , then  $M$  is a space of constant curvature.

Both  $R^3$  and  $S^3$  are examples of spaces of constant curvature.

### 3. Ricci Curvature

**(1) The last pair of indices of the curvature tensor  $R^j_{i\ kl}$  are antisymmetric**

Since  $R_{jikl} = -R_{jilk}$ , it follows that

$$R^j_{i\ kl} = g^{aj} R_{iakl} = -g^{aj} R_{ialk},$$

i.e.,

$$R^j_{i\ kl} = -R^j_{i\ lk}.$$

**(2) Ricci Tensor**

The curvature tensor  $R^j_{i\ kl}$  is a mixed tensor of type (1,3). It can be contracted in two ways:

**1) Contraction with respect to the first and second indices  $j$  and  $i$ :**

$$A_{kl} = R^i_{i\ kl} = g^{ia} R_{iakl} = -g^{ia} R_{aikl} = -R^a_{a\ kl} = -R^i_{i\ kl},$$

so

$$A_{kl} = R^i_{i\ kl} = 0.$$

Hence, the tensor obtained by contracting the first and second indices is zero.

**2) Contraction with respect to the first and third indices  $j$  and  $k$ :**

$$B_{il} = R^j_{i\ jl} = g^{aj} R_{iajl},$$

which yields a nonzero tensor.

**3) Contraction with respect to the first and fourth indices  $j$  and  $l$ :**

$$C_{ik} = R^j_{i\ kj} = g^{aj} R_{iakj} = -g^{aj} R_{iajk} = -g^{aj} R_{iajl}.$$

Due to the antisymmetry of the last pair of indices of the curvature tensor  $R^j_{i\ kl}$ , the tensor  $C_{ik}$  obtained by contracting the first and fourth indices differs from  $B_{il}$  only by a sign:

$$C_{ik} = -B_{il}.$$

Therefore, among the possible contractions, only one independent nonzero tensor exists. This tensor is called the **Ricci tensor**, denoted by  $R_{kl}$ :

$$R_{kl} = R^j_{k\ jl} = g^{ij} R_{kijl}.$$

$R_{kl}$  is a symmetric tensor:

$$R_{kl} = R_{lk},$$

because

$$R_{kl} = R^i_{k\ il} = g^{ij} R_{kijl} = g^{ij} R_{ilkj} = g^{ij} R_{lijk} = R^j_{l\ jk} = R_{lk}.$$

The Ricci tensor is essentially a symmetric second-order covariant tensor field. It defines a linear map from  $\chi(M) \times \chi(M)$  to  $C^\infty(M)$  on a Riemannian manifold  $(M, g)$ , denoted as  $Ric(X, Y)$ , i.e.,

$$Ric(X, Y): \chi(M) \times \chi(M) \rightarrow C^\infty(M).$$

**Definition 11.2.4** Let  $(M, g)$  be a Riemannian manifold,  $p \in M$ , and  $X \in T_p M$  with  $X \neq 0$ . Then

$$Ric(X) = \frac{Ric(X, X)}{g(X, X)} = Ric\left(\frac{X}{|X|}, \frac{X}{|X|}\right) \quad (11.2.6)$$

is a function of the tangent direction  $X$ , called the **Ricci curvature** of the Riemannian manifold  $(M, g)$  at point  $p$  along the tangent direction  $X$ .

If an orthonormal basis  $\{e_i\}$  is chosen in the tangent space  $T_p M$  at any point  $p$  of an  $m$ -dimensional Riemannian manifold  $(M, g)$ , then according to quadratic form theory, the Ricci curvature tensor can be diagonalized into the following standard form:

$$Ric = \sum_{i=1}^m \kappa_i \omega^i \otimes \omega^i, \quad (11.2.7)$$

where  $\kappa_i = Ric(e_i)$ , and  $\{\omega^i\}$  is the dual basis to  $\{e_i\}$  in  $T_p^*M$ .

**Theorem 11.2.3** Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold, and let  $X$  be a nonzero tangent vector at point  $p \in M$ . If  $\{e_i\}$  is an orthonormal basis of  $T_p M$  such that

$$e_m = \frac{X}{|X|}, \text{ then}$$

$$Ric(X) = \sum_{i=1}^{m-1} K(e_i, e_m), \quad (11.2.8)$$

where  $K(e_i, e_m)$  denotes the sectional curvature of the two-dimensional subspace spanned by  $e_i$  and  $e_m$ .

#### 4. Scalar Curvature

**Definition 11.2.5** Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold. For any orthonormal basis  $\{e_i\}$  of the tangent space  $T_p M$  at point  $p \in M$ , the numerical value

$$R = \sum_{i=1}^m Ric(e_i)$$

is called the **scalar curvature** of the Riemannian manifold  $(M, g)$  at point  $p$ .

From equation (11.2.8), it follows that

$$R = \sum_{i,j} K(e_i, e_j) = \sum_{i,j} R(e_i, e_j, e_j, e_i). \quad (11.2.9)$$

In a local coordinate system  $(U; x^i)$ , the scalar curvature is obtained by contracting the Ricci tensor, i.e.,

$$R = g^{ij} R_{ij} = g^{ij} R_{ikj}^k = g^{ij} g^{kl} R_{ilkj}.$$

#### 5. Einstein Tensor

Starting from the Bianchi identity:

$$R_{i\ kl; a}^j + R_{i\ la; k}^j + R_{i\ ak; l}^j = 0,$$

interchanging the indices  $l$  and  $a$  in the second term, we obtain

$$R_{i\ kl; a}^j - R_{i\ al; k}^j + R_{i\ ak; l}^j = 0.$$

Contracting the indices  $j$  and  $a$ , we obtain

$$R_{i\ kl; j}^j - R_{i\ l; k} + R_{i\ k; l} = 0.$$

Multiplying the equation by  $g^{li}$  and using  $g^{li}_{; \alpha} = 0$ , we get

$$(g^{li} R_{i\ kl}^j)_{; j} - (g^{li} R_{i\ l})_{; k} + (g^{li} R_{i\ k})_{; l} = 0,$$

$$R_{k; j}^j - R_{; k} + R_{k; l}^l = 0,$$

i.e.,

$$R_{k; l}^l - \frac{1}{2} R_{; k} = 0.$$

Multiplying by  $g^{ik}$ , we obtain

$$(g^{ik} R_{k; l}^l)_{; l} - \frac{1}{2} (g^{ik} R)_{; k} = 0,$$

i.e.,

$$R^{il};_{\quad l} - \frac{1}{2}(g^{il}R);_{\quad l} = 0.$$

Relabeling the indices yields

$$(R^{kl} - \frac{1}{2}g^{kl}R);_{\quad l} = 0.$$

Now, define:

$$G^{kl} = R^{kl} - \frac{1}{2}g^{kl}R,$$

then

$$G^{kl};_{\quad l} = 0.$$

The tensor  $G^{kl}$  is called the **Einstein tensor**.

## 6.Einstein Manifolds

**Definition 11.2.6** Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold. If there exists a constant  $\lambda$  such that  $R_{jl} = \lambda g_{jl}$ , then  $(M, g)$  is called an **Einstein manifold**.

If  $(M, g)$  is an  $m$ -dimensional space of constant curvature, then from equation (11.2.5), we have

$$R_{ijkl} = -K(g_{ik}g_{jl} - g_{il}g_{jk}). \quad (11.2.10)$$

Multiplying both sides of equation (11.2.10) by  $g^{ik}$ , we obtain

$$\begin{aligned} R_{jl} &= g^{ik}R_{ijkl} = -K(g^{ik}g_{ik}g_{jl} - g^{ik}g_{il}g_{jk}) \\ &= -K(mg_{jl} - \delta_l^k g_{jk}) \quad (\because g^{ik}g_{ik} = m) \\ &= -K(mg_{jl} - g_{jl}) \\ &= -K(m-1)g_{jl}, \end{aligned}$$

i.e.,

$$R_{jl} = -K(m-1)g_{jl}. \quad (11.2.11)$$

Therefore, an  $m$ -dimensional space of constant curvature is an Einstein manifold.

Multiplying both sides of equation (11.2.11) by  $g^{jl}$ , we get

$$R = g^{jl}R_{jl} = -K(m-1)g^{jl}g_{jl},$$

i.e.,

$$R = -K(m-1)m. \quad (11.2.12)$$

Thus, the scalar curvature  $R$  of an  $m$ -dimensional space of constant curvature is a constant. The scalar curvature  $R$  of an Einstein manifold is also a constant.

**Theorem 11.2.4** A constant curvature manifold is an Einstein manifold. However, if  $(M, g)$  is a connected Einstein manifold and its dimension  $m$  is 3, then  $(M, g)$  is a Riemannian manifold of constant curvature.

**Theorem 11.2.5 (F. Schur Theorem)** Let  $(M, g)$  be a Riemannian manifold of dimension  $m \geq 3$  that satisfies one of the following two conditions:

1) For all two-dimensional subspaces  $E \subset T_p M$  at point  $p \in M$ , the sectional curvature  $K(E) = f(p)$  depends only on  $p$ .

2) For all  $X \in T_p M$  at point  $p \in M$ , the Ricci curvature satisfies  $Ric(X) = (m-1) \cdot f(p) \cdot X$ .

In both cases,  $f$  must be constant. In other words, the metric  $g$  has constant curvature, or  $(M, g)$  is an Einstein manifold.

This is another version of the F. Schur Theorem. Condition 1) implies condition 2). For a detailed proof of the theorem, refer to the literature *Riemannian Geometry*<sup>1</sup>.

## 7. Locally Euclidean Spaces

**Definition 11.2.7** Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold. If at every point  $p \in M$ , there exists a local coordinate system  $(U; x^i)$  around  $p \in M$  such that the components of the Riemannian metric  $g$  are

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \delta_{ij},$$

then  $(M, g)$  is called a **locally Euclidean space**.

**Theorem 11.2.6** A Riemannian manifold  $(M, g)$  is locally Euclidean if and only if its curvature tensor is identically zero.

**Definition 11.2.8** A Riemannian manifold whose Riemann curvature tensor is identically zero is called a **flat Riemannian manifold**, and the corresponding Riemannian metric is called a **flat Riemannian metric**.

A locally Euclidean space is precisely a flat Riemannian manifold.

## §11.3 Gauss–Bonnet Theorem

**Theorem 11.3.1 (Gauss–Bonnet Theorem)** Let  $M$  be a compact, oriented two-dimensional Riemannian manifold,  $K$  be the Gaussian curvature on  $M$ , and  $d\sigma$  be the oriented area element on  $M$ . Then

$$\frac{1}{2\pi} \int_M K d\sigma = \chi, \quad (11.3.1)$$

where  $\chi$  is the Euler characteristic of the manifold  $M$ .

For the torus  $T^2$ ,  $\chi = 0$ . For the sphere  $S^2$ ,  $\chi = 2$ .

The two-dimensional Riemannian manifold  $M$  in Theorem 11.3.1 is independent of Euclidean space—that is,  $M$  has the property described in the theorem without being embedded in Euclidean space. Such properties are called **intrinsic properties** of  $M$ .

**Theorem 11.3.2 (Gauss–Bonnet Theorem)** Let  $D$  be a compact region with boundary in an oriented two-dimensional Riemannian manifold  $M$ , where the boundary  $C$  is a piecewise smooth curve with the orientation induced from  $M$ . Then

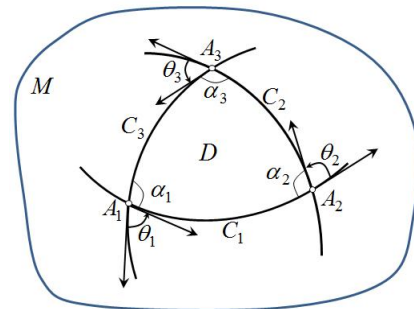
$$\int_C k_g ds + \int_D K d\sigma + \sum_i (\pi - \alpha_i) = 2\pi\chi, \quad (11.3.2)$$

where  $k_g$  is the geodesic curvature of  $C$ ,  $\pi - \alpha_i$  are the exterior angles at the vertices, and  $\chi$  is the Euler characteristic of  $D$ .

If the curve  $C$  consists of three geodesics  $C_1, C_2, C_3$ , then  $C$  is called a **geodesic triangle** on the surface  $M$ . As shown in Figure 11.3.1, the three geodesics  $C_1, C_2, C_3$  form a geodesic triangle  $\Delta A_1 A_2 A_3$ . Let its three interior angles be  $\alpha_1, \alpha_2, \alpha_3$ ; then the exterior angles are

$$\theta_i = \pi - \alpha_i \quad (i=1,2,3).$$

Since  $C_1, C_2, C_3$  are geodesics, the geodesic curvature  $k_g = 0$ . Thus, equation (11.3.2) becomes



**Figure 11.3.1** Geodesic triangle

<sup>1</sup> Peter Petersen. *Riemannian Geometry. Second Edition*[M]. Beijing: Science Press, 2007:39.

$$\int_D K d\sigma + \sum_i (\pi - \alpha_i) = 2\pi\chi. \quad (11.3.3)$$

The region  $D$  enclosed by the geodesics  $C_1, C_2, C_3$  is homeomorphic to a disk, and the Euler characteristic of a disk is  $\chi = 1$ . Therefore, the Euler characteristic of the surface region  $D$  is  $\chi = 1$ . Substituting this into equation (11.3.3) gives

$$\int_D K d\sigma + \sum_i (\pi - \alpha_i) = 2\pi. \quad (11.3.4)$$

Thus, the sum of the three interior angles of any geodesic triangle on the surface  $M$  is

$$\sum_i \alpha_i = \pi + \int_D K d\sigma. \quad (11.3.5)$$

If  $M$  is a plane, the Gaussian curvature  $K = 0$ , so the sum of the three interior angles of a planar triangle equals  $\pi$ . If  $M$  is the two-dimensional sphere  $S^2$ , the Gaussian curvature  $K$  is a positive constant. Hence,

$$\sum_i \alpha_i = \pi + K\sigma, \quad (11.3.6)$$

where  $\sigma$  denotes the area of the geodesic triangle. Therefore, the sum of the three interior angles of a geodesic triangle on the sphere  $S^2$  is greater than  $\pi$ .

## §11.4 Isometries and Conformal Transformations

### 1. Isometries

**Definition 11.4.1** Let  $(M, g)$  and  $(N, g')$  be Riemannian manifolds, and let  $f : (M, g) \rightarrow (N, g')$  be a smooth map.

$$f : (M, g) \rightarrow (N, g').$$

If for any  $p \in M$  and any tangent vectors  $X, Y \in T_p M$ , the following holds:

$$g(X, Y) = g'(f_* X, f_* Y),$$

then  $f$  is called an **isometric map** from the Riemannian manifold  $(M, g)$  to  $(N, g')$ .

**Definition 11.4.2** If  $f : (M, g) \rightarrow (N, g')$  is a local smooth diffeomorphism between smooth manifolds and satisfies  $g = f^* g'$ , then  $f$  is called a **local isometry** or **local isometric map**.

**Definition 11.4.3** If  $f : (M, g) \rightarrow (N, g')$  is a smooth diffeomorphism between smooth manifolds and satisfies  $g = f^* g'$ , then  $f$  is called an **isometry**. In this case, the Riemannian manifolds  $(M, g)$  and  $(N, g')$  are said to be **isometric** to each other.

**Definition 11.4.4** If  $f$  is an isometric map from a Riemannian manifold  $(M, g)$  to itself, then  $f$  is called an **isometry transformation** of  $(M, g)$ . The set of all isometry transformations of  $(M, g)$  forms a group, called the **isometry transformation group**.

**Theorem 11.4.1** Local isometry between Riemannian manifolds is **not** an equivalence relation, but isometry between Riemannian manifolds **is** an equivalence relation.

**Theorem 11.4.2** If two Riemannian manifolds are locally isometric, then their curvature tensors and sectional curvatures remain invariant under the local isometry.

### 2. Conformal Transformations

**Definition 11.4.5** Let  $f$  be a smooth diffeomorphism from a Riemannian manifold  $(M, g)$  to itself. If there exists a positive smooth function  $\lambda \in C^\infty(M)$  such that  $f^* g = \lambda^2 g$ , then the mapping  $f$  is called a **conformal transformation** (or **conformal map**) from the Riemannian manifold  $M$  to itself. Properties (or quantities) of the Riemannian manifold  $(M, g)$  that remain



invariant under all conformal transformations to itself are called **conformally invariant properties**(or **conformal invariants**).

Under a conformal transformation, the angle between two tangent vectors at a point remains unchanged.

**Definition 11.4.6** A Riemannian manifold  $(M, g)$  is said to be **locally conformally flat** if for every point  $p \in M$ , there exists an open neighborhood  $U$  containing  $p$  and a flat metric tensor  $\tilde{g}$  on  $U$ (i.e., with vanishing curvature tensor), such that  $g$  and  $\tilde{g}$  are conformal on  $U$ .

**Theorem 11.4.3** A Riemannian manifold of constant curvature is locally conformally flat. An  $m$ -dimensional ( $m \geq 3$ ) locally conformally flat Einstein manifold is a Riemannian manifold of constant curvature.

**Theorem 11.4.4** On every compact smooth Riemannian manifold  $(M, g)$  of dimension  $m(\geq 3)$ , there exists a positive smooth function  $\lambda \in C^\infty(M)$  such that the conformal metric  $\tilde{g} = \lambda^2 g$  has constant scalar curvature.

## §11.5 Riemannian Symmetric Spaces

**Definition 11.5.1** At a point  $p \in M$  of an  $m$ -dimensional Riemannian manifold  $(M, g)$ , if there exists a mapping  $f_p : M \rightarrow M$  satisfying the following conditions, then  $f_p$  is called a **central symmetry**(or **symmetry**) of  $M$  about the point  $p$ :

1)  $f_p$  is an isometry from the Riemannian manifold  $(M, g)$  to itself, i.e.,  $f_p$  is a diffeomorphism and satisfies  $f_p^*(g) = g$ .

2)  $f_p$  is an involution, i.e.,  $f_p \circ f_p = id$  (the identity map).

3)  $p$  is an isolated fixed point of  $f_p$ , i.e.,  $f_p(p) = p$ , and in a neighborhood  $U$  of  $p$ ,  $f_p$  has no other fixed points besides  $p$ .

If  $M$  has a central symmetry about the point  $p \in M$ , then the Riemannian manifold  $(M, g)$  is said to be **symmetric** about  $p$ .

**Definition 11.5.2** If a Riemannian manifold  $(M, g)$  is symmetric about every point of  $M$ , then  $(M, g)$  is called a **Riemannian symmetric space**.

Euclidean space  $R^m$  and the  $m$ -dimensional unit sphere  $S^m$  are Riemannian symmetric spaces. In fact, spaces of constant curvature have a central symmetry at every point and are therefore Riemannian symmetric spaces.

**Theorem 11.5.1** A Riemannian manifold is a Riemannian symmetric space if and only if its curvature tensor satisfies  $R_{ikl}^j{}_{;h} = 0$ .

## §11.6 Divergence of Tangent Vector Fields

Let  $X$  be a smooth tangent vector field on a Riemannian manifold  $(M, g)$ . In a local coordinate system  $(U; x^i)$ ,

$$X = X^i \frac{\partial}{\partial x^i}.$$

According to equation (10.1.4), the covariant derivative of  $X$  is computed as

$$X^i{}_{;j} = \frac{\partial X^i}{\partial x^j} + X^k \Gamma_{kj}^i.$$

Taking the contraction (trace) of  $X^i{}_{;j}$  yields

$$\operatorname{div} X = X^i{}_{;i} = \frac{\partial X^i}{\partial x^i} + X^k \Gamma_{ki}^i. \quad (11.6.1)$$

$\operatorname{div} X$  is called the **divergence** of the smooth tangent vector field  $X$ .

On an oriented  $m$ -dimensional Riemannian manifold  $(M, g)$ , a volume can be defined. Let  $\det(g_{ij})$  denote the determinant of the metric matrix of  $M$ . Suppose  $\Omega$  is an  $m$ -form on  $M$ . In a local coordinate system  $(U; x^i)$ ,

$$\Omega|_U = \sqrt{\det(g_{ij})} dx^1 \wedge \cdots \wedge dx^m, \quad (11.6.2)$$

then  $\Omega$  is called the **volume form** on  $M$ .

If  $M$  is a surface, then  $m = 2$ , and  $\Omega$  is the **area form** on  $M$ :

$$\Omega|_U = \sqrt{g_{11}g_{22} - g_{12}^2} dx^1 \wedge dx^2, \quad (11.6.3)$$

where  $\det(g_{ij}) = g_{11}g_{22} - g_{12}^2$ . The area of a region  $U$  is

$$\int_U \Omega|_U = \int_U \sqrt{g_{11}g_{22} - g_{12}^2} dx^1 \wedge dx^2. \quad (11.6.4)$$

The integral of a smooth function  $f$  over  $M$  can be defined as the integral of the  $m$ -form  $f\Omega$  over  $M$ , i.e.,  $\int_M f\Omega$ .

**Example 11.6.1** The metric form of three-dimensional Euclidean space  $R^3$  in spherical coordinates  $(U; r, \theta, \varphi)$  can be expressed as

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2,$$

where

$$g_{11} = 1, \quad g_{22} = r^2, \quad g_{33} = r^2 \sin^2 \theta.$$

Since

$$\det(g_{ij}) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{vmatrix} = r^4 \sin^2 \theta,$$

equation (11.6.2) becomes

$$\Omega|_U = \sqrt{r^4 \sin^2 \theta} dr \wedge d\theta \wedge d\varphi = r^2 \sin \theta dr \wedge d\theta \wedge d\varphi. \quad (11.6.5)$$

Regarding divergence, the following integral theorems hold.

**Theorem 11.6.1** Let  $(M, g)$  be an  $m$ -dimensional, oriented, compact Riemannian manifold without boundary. Then for any smooth tangent vector field  $X$  on  $M$ ,

$$\int_M (\operatorname{div} X) \Omega = 0.$$

**Theorem 11.6.2 (Divergence Theorem)** Let  $M$  be an  $m$ -dimensional, oriented, compact Riemannian manifold with boundary, and let  $n$  be the outward-pointing unit normal vector field on the boundary  $\partial M$ . Then for any smooth tangent vector field  $X$  on  $M$ , the following integral formula holds:

$$\int_M (\operatorname{div} X) dV_M = \int_{\partial M} g \langle n, X \rangle dV_{\partial M},$$

where  $\partial M$  carries the orientation induced from  $M$ , and  $dV_{\partial M}$  is the volume element on  $\partial M$ .

**Theorem 11.6.3** Let  $M$  be an  $m$ -dimensional Riemannian manifold. Then the scalar curvature  $R(p)$  at point  $p \in M$  can be expressed as

$$R(p) = \frac{1}{V(B^m)} \int_{S^{m-1}} \operatorname{Ric}_p(X) dV_{S^{m-1}}, \quad (11.6.6)$$

where  $dV_{S^{m-1}}$  is the volume element of the unit sphere  $S^{m-1}$  in the tangent space  $T_p M$ , and  $V(B^m)$  is the volume of the unit ball  $B^m$  enclosed by the unit sphere  $S^{m-1}$ .

**Theorem 11.6.4<sup>2</sup>** Let  $M$  be an  $m$ -dimensional Riemannian manifold. In  $T_p M$ , choose an orthonormal basis  $\{e_i\}$  such that  $R_{ij} = \text{Ric}(e_i, e_j) = \lambda_i \delta_{ij}$ . Take any tangent vector  $u = u^i e_i$  in  $T_p M$  and let  $X(u) = \sum_i \lambda_i u^i e_i$ ; then  $X$  is a tangent vector field on  $T_p M$ . Furthermore, let  $v$  be an outward unit normal vector on  $S^{m-1} \subset T_p M$ . Then the scalar curvature  $R(p)$  at point  $p \in M$  can be expressed as

$$R(p) = \frac{1}{V(B^m)} \int_{B^m} \text{div}(X) dV_{B^m}, \quad (11.6.7)$$

or

$$R(p) = \frac{1}{V(B^m)} \int_{S^{m-1}} \langle X(v), v \rangle dV_{S^{m-1}}, \quad (11.6.8)$$

where  $V(B^m)$  is the volume of the unit ball  $B^m$  enclosed by the unit sphere  $S^{m-1}$  in  $T_p M$ ,  $dV_{B^m}$  is the volume element of  $B^m$ , and  $\langle X(v), v \rangle$  is the inner product of  $X(v)$  and  $v$ .

Since

$$R(p) = \frac{1}{V(B^m)} \int_{S^{m-1}} \langle X(v), v \rangle dV_{S^{m-1}} = \frac{1}{V(B^m)} \int_{S^{m-1}} X(v) \cdot v dV_{S^{m-1}},$$

equation (11.6.8) can also be written in the following vector form:

$$R(p) = \frac{1}{V(B^m)} \int_{S^{m-1}} X(v) \cdot dS, \quad (11.6.9)$$

where  $dS = v dV_{S^{m-1}}$ .

## §11.7 Completeness and the Bonnet–Myers Theorem

Before discussing the Bonnet–Myers theorem, it is necessary to understand concepts such as completeness and diameter.

**Definition 11.7.1** Let  $(a_1, a_2, \dots)$  be a sequence of points in a metric space  $(M, d)$ . If for any given  $\varepsilon > 0$ , there exists a positive integer  $N(\varepsilon)$  such that whenever  $n, l > N(\varepsilon)$ , we have

$$d(a_n, a_l) < \varepsilon,$$

where  $d(a_n, a_l)$  denotes the distance between  $a_n$  and  $a_l$ , then  $(a_1, a_2, \dots)$  is called a **Cauchy sequence**.

Intuitively, the elements of a Cauchy sequence become arbitrarily close to one another as the indices increase.

**Example 11.7.1** Consider the sequence

$$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{k}, \dots$$

with the distance defined as

$$d(a_n, a_l) = \frac{1}{l} - \frac{1}{n} = \frac{n-l}{nl},$$

where  $n > l$ . If we arbitrarily specify a positive number  $\varepsilon$ , for example,  $\varepsilon = 0.05$ , we can find a positive integer  $N(0.05) = 6$  such that whenever  $l > 6$ ,  $n > 6$ , say  $l = 8$  and  $n = 10$ , we have

$$d(a_{10}, a_8) = \frac{1}{8} - \frac{1}{10} = \frac{2}{80} = 0.025 < 0.05.$$

Thus, this sequence is a Cauchy sequence.

<sup>2</sup>Chen Weihuan, Li Xingxiao. *An Introduction to Riemannian Geometry* (Volume 1) [M]. Beijing: Peking University Press, 2002: 475.

**Definition 11.7.2** A metric space  $(M, d)$  is said to be **complete** if every Cauchy sequence in  $(M, d)$  converges.

**Example 11.7.2** The open interval  $(0,1)$  is not complete because the sequence

$$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots$$

is a Cauchy sequence belonging to  $(0,1)$ , but it does not converge to any point in  $(0,1)$ .

**Theorem 11.7.1** Every compact metric space is a complete metric space.

**Definition 11.7.3** A connected Riemannian manifold  $(M, g)$  is called a **complete Riemannian manifold** if it is complete as a metric space.

**Theorem 11.7.2** On a complete Riemannian manifold, any two points can be connected by a shortest geodesic.

**Theorem 11.7.3** Let  $M$  be a connected Riemannian manifold. Then the following two conditions are equivalent:

- 1)  $M$  is complete;
- 2) Every geodesic on  $M$  can be extended indefinitely.

**Theorem 11.7.4** A compact connected Riemannian manifold is complete.

According to the above theorem, every geodesic on a compact connected Riemannian manifold can be extended indefinitely.

**Theorem 11.7.5** Isometric transformations preserve the completeness of Riemannian manifolds.

**Definition 11.7.4** Let  $S$  be a set of real numbers. If there exists a real number  $a$  such that  $x \geq a$  for every element  $x \in S$ , then  $a$  is called a **lower bound** of  $S$ . For any given set  $E$  of numbers, the greatest lower bound of  $E$  is called the **infimum** of  $E$ . The symbol  $\inf$  denotes the infimum.

**Example 11.7.3** The set  $\{1,2,3,4,5\}$  has a minimum value of 1. Thus, 1 is a lower bound of the set, and since it is the greatest lower bound, the infimum of the set is 1, denoted by  $\inf\{1,2,3,4,5\} = 1$ . This infimum belongs to the set. For the set  $\{x \in \mathbb{R}, 0 < x < 1\}$ , the infimum is 0, denoted by  $\inf\{x \in \mathbb{R}, 0 < x < 1\} = 0$ , but this infimum does not belong to the set.

**Definition 11.7.5** Let  $(M, g)$  be a connected Riemannian manifold. For any two points  $p, q \in M$ , define

$$d(p, q) = \inf\{L(\gamma); \gamma \text{ is a piecewise smooth curve joining } p \text{ and } q \text{ in } M\},$$

where  $L(\gamma)$  is the arc length of the curve  $\gamma$ . The quantity  $d(p, q)$  is called the **distance** between  $p$  and  $q$  in  $M$ .

**Definition 11.7.6** The **diameter**  $d(M)$  of a Riemannian manifold  $(M, g)$  is defined as

$$d(M) = \max\{d(p, q); p, q \in M\}.$$

If its diameter satisfies  $d(M) < +\infty$ , then the Riemannian manifold  $(M, g)$  is said to be **bounded**.

**Theorem 11.7.6 (Bonnet–Myers Theorem)** Let  $(M, g)$  be a complete  $m$ -dimensional Riemannian manifold whose Ricci curvature is bounded below by a positive constant  $a > 0$ , i.e., for every point  $p \in M$  and any tangent vector  $v \in T_p M$ , we have  $\text{Ric}(v) \geq a$ . Then  $(M, g)$  is a compact Riemannian manifold with diameter not exceeding  $\pi\sqrt{m-1}/\sqrt{a}$ .

From this theorem, it follows that

$$d(M) \leq \frac{\pi\sqrt{m-1}}{\sqrt{a}}. \quad (11.7.1)$$

Let  $\text{Ric}(v) = a$ . According to equation (11.7.1): If the lower bound of the Ricci curvature  $\text{Ric}(v)$  of the Riemannian manifold  $(M, g)$  is larger, the maximum diameter  $d(M)_{\max}$  of  $M$  is smaller. Conversely, if the lower bound of the Ricci curvature  $\text{Ric}(v)$  is smaller, the maximum diameter  $d(M)_{\max}$  of  $M$  is larger. If the lower bound of the Ricci curvature  $\text{Ric}(v)$  approaches zero, the maximum diameter  $d(M)_{\max}$  of  $M$  tends to infinity. Conversely, if the lower bound of the Ricci

curvature  $Ric(v)$  tends to infinity, the maximum diameter  $d(M)_{\max}$  of  $M$  tends to zero.

**Theorem 11.7.7 (Bonnet Theorem)** Let  $(M, g)$  be a complete  $m$ -dimensional Riemannian manifold whose sectional curvature  $K$  has a positive lower bound, i.e.,

$$K \geq \delta > 0,$$

then  $M$  is compact, and its diameter satisfies  $d(M) \leq \frac{\pi}{\sqrt{\delta}}$ .

**Theorem 11.7.8 (Toponogov Maximal Diameter Theorem)** Let  $(M, g)$  be a complete  $m$ -dimensional Riemannian manifold whose sectional curvature  $K$  has a positive lower bound, i.e.,

$$K \geq \delta > 0,$$

and its diameter is  $d(M) = \frac{\pi}{\sqrt{\delta}}$ , then  $M$  is isometric to an  $m$ -dimensional sphere of radius  $\frac{1}{\sqrt{\delta}}$ .

## §11.8 Submanifolds in Euclidean Space

**Theorem 11.8.1 (Hilbert's Theorem)** Let  $(M, g)$  be a complete two-dimensional Riemannian manifold. If its Gauss curvature  $K$  is negative, i.e.,  $K \leq -a^2 < 0$ , then  $(M, g)$  cannot be isometrically embedded into  $R^3$  as an embedded submanifold of  $R^3$ .

**Theorem 11.8.2** The standard sphere  $S^m(r)$  of radius  $r > 0$  in  $R^{m+1}$  has constant sectional curvature  $K = \frac{1}{r^2}$ .

**Theorem 11.8.3** Let  $M$  be a compact connected hypersurface in  $R^{m+1}$ . Then the following three conditions are equivalent:

- 1) The sectional curvature  $K$  of  $M$  is nowhere zero.
- 2) The sectional curvature  $K$  of  $M$  is everywhere positive.
- 3)  $M$  is orientable, and its Gauss map  $G: M \rightarrow S^m$  is a diffeomorphism.

**Theorem 11.8.4** Let  $M$  be a compact embedded hypersurface without boundary in  $R^{m+1}$ . If  $M$  is connected and has constant scalar curvature, then  $M$  is a standard sphere in  $R^{m+1}$ .

## §11.9 Induced Metrics and Space Forms

### 1. Induced Metric

**Theorem 11.9.1** Let  $M$  and  $N$  be two smooth manifolds, and let  $f: M \rightarrow N$  be a smooth map.

1) If  $\varphi$  is a smooth  $r$ -th order covariant tensor field on  $N$ , then for any  $p \in M$  and any tangent vectors  $X_1, \dots, X_r \in T_p M$ , there exists a smooth  $r$ -th order covariant tensor field  $f^* \varphi$  on  $M$  satisfying

$$(f^* \varphi)|_p(X_1, \dots, X_r) = \varphi|_p(f_{*p}(X_1), \dots, f_{*p}(X_r)). \quad (11.9.1)$$

Moreover, if  $\varphi$  is symmetric, then  $f^* \varphi$  is also symmetric; if  $\varphi$  is antisymmetric, then  $f^* \varphi$  is also antisymmetric.

2) If  $f$  is an immersion and  $h$  is a Riemannian metric on the smooth manifold  $N$ , then  $g = f^* h$  is a Riemannian metric on the smooth manifold  $M$ .

The Riemannian metric referred to in this theorem is positive definite. We call  $g$  the **induced metric** on  $M$  determined by the pullback  $f^*$  of the Riemannian metric  $h$  via the map  $f$ .

## 2. Induced Metric of $S^3$ in $R^4$

The Riemannian metric of four-dimensional Euclidean space  $R^4$  in Cartesian coordinates  $(x, y, z, w)$  is

$$ds^2 = dx^2 + dy^2 + dz^2 + dw^2. \quad (11.9.2)$$

If  $w$  is a constant, we obtain the Riemannian metric of three-dimensional Euclidean space  $R^3$ :

$$ds^2 = dx^2 + dy^2 + dz^2, \quad (11.9.3)$$

called the Riemannian metric **induced** from  $R^4$ .

If spherical coordinates are used:

$$\begin{cases} x = r \sin \omega \sin \theta \cos \varphi, \\ y = r \sin \omega \sin \theta \sin \varphi, \\ z = r \sin \omega \cos \theta, \\ w = r \cos \omega, \end{cases}$$

$$0 \leq r < +\infty, \quad 0 \leq \omega \leq \pi, \quad 0 \leq \theta \leq \pi, \quad 0 \leq \varphi \leq 2\pi,$$

then equation (11.9.2) becomes

$$ds^2 = dr^2 + r^2[d\omega^2 + \sin^2 \omega(d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (11.9.4)$$

Consider the three-dimensional sphere  $S^3$  of radius  $R$  embedded in  $R^4$ :

$$x^2 + y^2 + z^2 + w^2 = R^2,$$

Since the radius  $R$  of  $S^3$  is constant, setting  $r = R$  in the Riemannian metric (11.9.4) yields

$$ds^2 = R^2[d\omega^2 + \sin^2 \omega(d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (11.9.5)$$

Equation (11.9.5) is the Riemannian metric of  $S^3$  in  $R^4$ , called the **induced Riemannian metric** from  $R^4$ .

Using the Riemannian metric (11.9.5), we can compute the surface area of the three-dimensional sphere  $S^3$  in  $R^4$ . The metric determinant of (11.9.5) is

$$\det(g_{ij}) = \begin{vmatrix} R^2 & 0 & 0 \\ 0 & R^2 \sin^2 \omega & 0 \\ 0 & 0 & R^2 \sin^2 \omega \sin^2 \theta \end{vmatrix} = R^6 \sin^4 \omega \sin^2 \theta,$$

and the volume form is

$$\begin{aligned} \Omega &= \sqrt{\det(g_{ij})} dx^1 \wedge dx^2 \wedge dx^3 \\ &= \sqrt{R^6 \sin^4 \omega \sin^2 \theta} d\omega \wedge d\theta \wedge d\varphi = R^3 \sin^2 \omega \sin \theta d\omega d\theta d\varphi. \end{aligned}$$

The surface area of the three-dimensional sphere  $S^3$  is

$$\begin{aligned} V &= \int_{S^3} \Omega = \int_{S^3} R^3 \sin^2 \omega \sin \theta d\omega d\theta d\varphi = R^3 \int_0^\pi \sin^2 \omega d\omega \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &= R^3 \int_0^\pi \frac{1}{2} (1 - \cos 2\omega) d\omega \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &= R^3 \left[ \frac{1}{2} \omega - \frac{1}{4} \sin 2\omega \right]_0^\pi \left[ -\cos \theta \right]_0^\pi \left[ \varphi \right]_0^{2\pi} = 2\pi^2 R^3. \end{aligned} \quad (11.9.6)$$

Integrating the surface area of  $S^3$  with respect to the radial coordinate  $r$  gives the volume enclosed by the three-dimensional sphere  $S^3$ :

$$V = \int_0^R 2\pi^2 r^3 dr = 2\pi^2 \left[ \frac{1}{4} r^4 \right]_0^R = \frac{1}{2} \pi^2 R^4. \quad (11.9.7)$$

Similarly, the surface area of the two-dimensional sphere  $S^2$  is  $4\pi R^2$ , and the volume it

encloses is

$$V = \int_0^R 4\pi r^2 dr = 4\pi \left[ \frac{1}{3} r^3 \right]_0^R = \frac{4}{3} \pi R^3.$$

### 3. Metrics on Open Balls

**Example 11.9.1** For any constant  $K$ , define

$$\rho(K) = \begin{cases} +\infty, & K \geq 0; \\ -\frac{1}{K}, & K < 0. \end{cases}$$

Further, define the open ball in  $R^m$ :

$$U = \left\{ (x^1, \dots, x^m) \in R^m \mid \sum_{i=1}^m (x^i)^2 < \rho(K) \right\}. \quad (11.9.8)$$

If we define the Riemannian metric form on  $U$  as

$$ds^2 = \frac{4 \sum_{i=1}^m (dx^i)^2}{\left( 1 + K \sum_{i=1}^m (x^i)^2 \right)^2}, \quad (11.9.9)$$

then the Riemannian manifold  $(U, ds^2)$  is a space of constant curvature with sectional curvature  $K$ .

The open ball  $U$  in  $R^m$  is non-compact. This example demonstrates that one can define a Riemannian metric with positive or negative sectional curvature on an open ball  $U$ .

**Theorem 11.9.2** Any non-compact manifold  $M$  can be endowed with a Riemannian metric such that its sectional curvature  $K$  satisfies

$$a^2 < K < b^2 \quad \text{or} \quad -a^2 < K < -b^2,$$

where  $a$  and  $b$  are arbitrary distinct positive numbers.

### 4. Space Forms

**Definition 11.9.1** A **space form** is a complete, simply connected space of constant curvature.

**Example 11.9.2** 1) Euclidean space  $R^m$  is an  $m$ -dimensional space form with sectional curvature  $K = 0$ .

2) The standard sphere  $S^m (m \geq 2)$  of radius  $r > 0$  in  $R^{m+1}$  is an  $m$ -dimensional space form with sectional curvature  $K = 1/r^2$ .

3) Hyperbolic space  $H^m(K) (K < 0)$  is an  $m$ -dimensional space form with sectional curvature  $K$ .

When  $K < 0$  in equation (11.9.9), the open ball  $U$  can serve as a model for hyperbolic space  $H^m(K) (K < 0)$ .

**Theorem 11.9.3** Let  $M^m(K)$  be an  $m$ -dimensional space form of constant sectional curvature  $K$ . Then:

- 1) If  $K > 0$ ,  $M^m(K)$  is isometric to the standard sphere  $S^m$  of radius  $r = 1/\sqrt{K}$ .
- 2) If  $K = 0$ ,  $M^m(K)$  is isometric to Euclidean space  $R^m$ .
- 3) If  $K < 0$ ,  $M^m(K)$  is isometric to hyperbolic space  $H^m(K)$ .

**Theorem 11.9.4 (Hadamard's Theorem)** Let  $M$  be a complete, simply connected  $m$ -dimensional Riemannian manifold. If the sectional curvature of  $M$  satisfies  $K \leq 0$ , then for any point  $p \in M$ , the exponential map

$$\exp_p : T_p M \rightarrow M$$

is a diffeomorphism.

### §11.10 Products of Riemannian Manifolds

**Theorem 11.10.1** As shown in Figure 11.10.1, let  $M_1$  and  $M_2$  be two smooth manifolds. Their product manifold is  $M = M_1 \times M_2$ , with points  $(p, q) \in M$ , and let  $\pi_i : M \rightarrow M_i (i=1,2)$  be the natural projections. Define mappings  $f_i : M_i \rightarrow M (i=1,2)$  such that

$$f_1(x) = (x, q), \quad \forall x \in M_1, \quad f_2(y) = (p, y), \quad \forall y \in M_2,$$

then  $\pi_i \circ f_i = \text{id}_{M_i} : M_i \rightarrow M_i (i=1,2)$ , and  $f_1, f_2$  are embeddings. Moreover,

$$T_{(p,q)}M = (f_1)_* (T_p M_1) \oplus (f_2)_* (T_q M_2),$$

which is isomorphic to the direct sum  $T_p M_1 \oplus T_q M_2$ .

**Definition 11.10.1** Let  $(M_1, g_1)$  and  $(M_2, g_2)$  be two Riemannian manifolds. Suppose the tangent space of  $M_1$  at point  $p$  is  $T_p M_1$ , and the tangent space of  $M_2$  at point  $q$  is  $T_q M_2$ . The tangent space of the product manifold  $M_1 \times M_2$  at point  $(p, q)$  is  $T_{(p,q)} M_1 \times M_2$ , which equals the direct sum of  $T_p M_1$  and  $T_q M_2$ , i.e.,

$$T_{(p,q)} M_1 \times M_2 = T_p M_1 \oplus T_q M_2.$$

On the product manifold, a metric  $g = g_1 \times g_2$  can be introduced such that for any  $(p, q) \in M_1 \times M_2$ ,

$$g(X, Y) = g_1(X_1, Y_1) + g_2(X_2, Y_2), \quad (11.10.1)$$

where

$$X_1, Y_1 \in T_p M_1, \quad X_2, Y_2 \in T_q M_2, \quad X, Y \in T_{(p,q)} M_1 \times M_2.$$

The metric  $g$  is called the **Riemannian metric of the product manifold**  $M_1 \times M_2$ .

According to this definition, we have

$$\begin{aligned} g(X_1 + X_2, Y_1 + Y_2) &= g(X_1, Y_1) + g(X_1, Y_2) + g(X_2, Y_1) + g(X_2, Y_2) \\ &= g_1(X_1, Y_1) + g_2(X_2, Y_2), \end{aligned}$$

so we may set

$$g(X_1, Y_1) = g_1(X_1, Y_1), \quad g(X_2, Y_2) = g_2(X_2, Y_2), \quad g(X_1, Y_2) = g(X_2, Y_1) = 0.$$

Moreover, since

$$\begin{aligned} g(X_1 + Y_2, Y_1 + X_2) &= g(X_1, Y_1) + g(X_1, X_2) + g(Y_2, Y_1) + g(X_2, Y_2) \\ &= g_1(X_1, Y_1) + g_2(X_2, Y_2), \end{aligned}$$

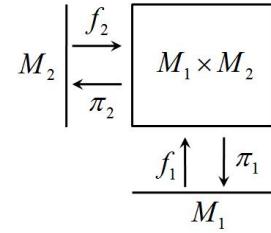
we can set

$$g(X_1, Y_1) = g_1(X_1, Y_1), \quad g(X_2, Y_2) = g_2(X_2, Y_2), \quad g(X_1, X_2) = g(Y_1, Y_2) = 0.$$

This means that on the product manifold, the inner product of two tangent vectors belonging to different submanifolds is zero.

**Example 11.10.1** The torus  $S^1 \times S^1$  can be regarded as a submanifold in three-dimensional Euclidean space  $R^3$ , endowed with the Riemannian metric induced from  $R^3$ . However, the torus can also be viewed as the product manifold  $S^1 \times S^1$  of two circles, forming another Riemannian manifold with a Riemannian metric distinct from that of the torus in  $R^3$ . This product manifold cannot be realized as a submanifold in  $R^3$  but can be embedded into  $R^4$ .

**Theorem 11.10.2** Let  $M_1$  and  $M_2$  be two Riemannian manifolds with Riemannian connections  $D^{(1)}$  and  $D^{(2)}$ , respectively. The product manifold  $M = M_1 \times M_2$  is a



**Figure 11.10.1** Projection maps on a product manifold



Riemannian manifold. Let  $(U; u^i)$  be a local coordinate system on  $M_1$  and  $(V; v^\alpha)$  be a local coordinate system on  $M_2$ . Define

$$D_{\frac{\partial}{\partial u^i}}^{(1)} \frac{\partial}{\partial u^j} = \Gamma_{ji}^{(1)k} \frac{\partial}{\partial u^k}, \quad D_{\frac{\partial}{\partial v^\alpha}}^{(2)} \frac{\partial}{\partial v^\beta} = \Gamma_{\beta\alpha}^{(2)\gamma} \frac{\partial}{\partial v^\gamma}.$$

A local coordinate system on the product manifold  $M$  can be taken as  $(U \times V; u^i, v^\alpha)$ , with natural frames

$$\left\{ (p, q); \frac{\partial}{\partial u^i} \Big|_{(p,q)}, \frac{\partial}{\partial v^\alpha} \Big|_{(p,q)} \right\}.$$

On  $M$ , for any fixed  $q \in V$ ,

$$\frac{\partial}{\partial u^i} \Big|_{(p,q)} = (f_1)_* \left( \frac{\partial}{\partial u^i} \Big|_p \right),$$

and for any fixed  $p \in U$ ,

$$\frac{\partial}{\partial v^\alpha} \Big|_{(p,q)} = (f_2)_* \left( \frac{\partial}{\partial v^\alpha} \Big|_q \right).$$

The Riemannian connection  $D$  on the product manifold  $M$  is given by the following formulas:

$$D_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} = (\Gamma_{ji}^{(1)k} \circ \pi_1) \frac{\partial}{\partial u^k}, \quad (11.10.2)$$

$$D_{\frac{\partial}{\partial v^\alpha}} \frac{\partial}{\partial v^\beta} = (\Gamma_{\beta\alpha}^{(2)\gamma} \circ \pi_2) \frac{\partial}{\partial v^\gamma}, \quad (11.10.3)$$

$$D_{\frac{\partial}{\partial v^\alpha}} \frac{\partial}{\partial u^i} = D_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial v^\alpha} = 0. \quad (11.10.4)$$

From equation (11.10.4), the covariant derivative of a tangent vector field across different factor manifolds is zero. For instance, on the Riemannian manifold  $M_1$ , the covariant derivative of a tangent vector field from  $M_2$  equals zero.

Let a smooth tangent vector field  $X$  be expressed in the local coordinate system  $(U; u^i)$  of  $M_1$  as  $X = x^i \frac{\partial}{\partial u^i}$ , and let a smooth tangent vector field  $Y$  be expressed in the local coordinate system  $(V; v^\alpha)$  of  $M_2$  as  $Y = y^\alpha \frac{\partial}{\partial v^\alpha}$ . Then,

$$D_X Y = D_{x^i \frac{\partial}{\partial u^i}} y^\alpha \frac{\partial}{\partial v^\alpha} = x^i D_{\frac{\partial}{\partial u^i}} y^\alpha \frac{\partial}{\partial v^\alpha} = x^i \left( \frac{\partial y^\alpha}{\partial u^i} \frac{\partial}{\partial v^\alpha} + y^\alpha D_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial v^\alpha} \right).$$

Since  $y^\alpha$  are functions on  $(V; v^\alpha)$  and not on  $(U; u^i)$ , we have  $\frac{\partial y^\alpha}{\partial u^i} = 0$ . Moreover, from

equation (11.10.4),  $D_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial v^\alpha} = 0$ . Therefore, we obtain

$$D_X Y = 0. \quad (11.10.5)$$

Similarly, we have

$$D_Y X = 0. \quad (11.10.6)$$

**Theorem 11.10.3** Let  $M_1$  and  $M_2$  be two Riemannian manifolds with Riemannian metric tensor fields  $g_1$  and  $g_2$ , respectively. Let  $p \in M_1$ ,  $q \in M_2$ ,  $v \in T_p M_1$ ,  $w \in T_q M_2$ , and assume  $v \neq 0$ ,  $w \neq 0$ . If the product manifold is  $M = M_1 \times M_2$ , and  $T_{(p,q)} M = T_p M_1 \oplus T_q M_2$ , then the sectional curvature determined by  $v$  and  $w$  on  $M$  satisfies  $K(v, w) = 0$ .

## §11.11 Spherical Trigonometry

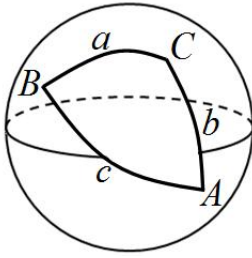
### 1. Cosine Theorem for Sides of a Spherical Triangle

**Cosine Theorem for Sides of a Spherical Triangle** As shown in Figure 11.11.1, consider an arbitrary triangle  $ABC$  on a sphere of radius 1. Denote the three interior angles by  $A$ ,  $B$ , and  $C$ , and let the sides opposite these angles be  $a$ ,  $b$ , and  $c$ , respectively. Then the following relations hold:

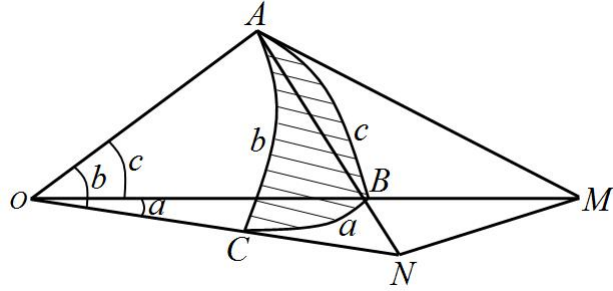
$$\cos a = \cos b \cos c + \sin b \sin c \cos A, \quad (11.11.1)$$

$$\cos b = \cos c \cos a + \sin c \sin a \cos B, \quad (11.11.2)$$

$$\cos c = \cos a \cos b + \sin a \sin b \cos C. \quad (11.11.3)$$



**Figure 11.11.1** Triangle on a sphere



**Figure 11.11.2** Proof of the cosine theorem for the sides of a spherical triangle

**Proof:** As shown in Figure 11.11.2, consider a triangle  $ABC$  (shaded region) on a sphere of radius 1. Connect its vertices to the center  $O$  of the sphere to obtain a spherical triangle  $O-ABC$ . Through vertex  $A$  of triangle  $ABC$ , draw tangents to sides  $b$  and  $c$ , intersecting the extensions of  $OB$  and  $OC$  at points  $M$  and  $N$ , respectively. This yields two right triangles  $\triangle OAM$  and  $\triangle OAN$ .

Since the measure of an angle in radians equals arc length divided by radius, when the sphere's radius is 1, we have:

$$\angle BOC = a \text{ (arc length of } BC), \angle AOC = b \text{ (arc length of } AC), \angle AOB = c \text{ (arc length of } AB).$$

In the planar triangle  $\triangle OMN$ , by the planar cosine theorem:

$$MN^2 = OM^2 + ON^2 - 2OM \cdot ON \cos a.$$

Similarly, in the planar triangle  $\triangle AMN$ , by the planar cosine theorem:

$$MN^2 = AM^2 + AN^2 - 2AM \cdot AN \cos A.$$

Thus, we obtain:

$$OM^2 + ON^2 - 2OM \cdot ON \cos a = AM^2 + AN^2 - 2AM \cdot AN \cos A,$$

i.e.,

$$\begin{aligned} 2OM \cdot ON \cos a &= (OM^2 - AM^2) + (ON^2 - AN^2) + 2AM \cdot AN \cos A \\ &= OA^2 + OA^2 + 2AM \cdot AN \cos A, \end{aligned}$$

or

$$\cos a = \frac{OA}{ON} \frac{OA}{OM} + \frac{AN}{ON} \frac{AM}{OM} \cos A.$$

Substituting

$$\cos b = \frac{OA}{ON}, \quad \cos c = \frac{OA}{OM}, \quad \sin b = \frac{AN}{ON}, \quad \sin c = \frac{AM}{OM}$$

into the above equation yields

$$\cos a = \cos b \cos c + \sin b \sin c \cos A.$$

This is equation (11.11.1). The other two cosine formulas can be obtained by cyclic permutation of the sides and angles.

If the sphere's radius is not 1 but  $r$  ( $r \neq 1$ ), then

$$\angle BOC = \frac{a}{r}, \quad \angle AOC = \frac{b}{r}, \quad \angle AOB = \frac{c}{r},$$

and equation (11.11.1) should be modified to

$$\cos \frac{a}{r} = \cos \frac{b}{r} \cos \frac{c}{r} + \sin \frac{b}{r} \sin \frac{c}{r} \cos A.$$

## 2. General Cosine Theorem

Another important and general cosine theorem is available for reference:

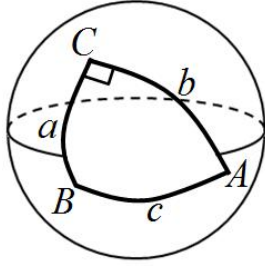
**Cosine Theorem** Let  $N_k^n$  be a space of constant curvature with dimension  $n$  and sectional curvature  $k$ . Consider a triangle in  $N_k^n$  with side lengths  $a, b, c$ . If angle  $A$  is opposite side  $a$ , then

$$\begin{aligned} k = 0 & \quad a^2 = b^2 + c^2 - 2bc \cos A, \\ k = -1 & \quad \cosh a = \cosh b \cosh c - \sinh b \sinh c \cos A, \\ k = 1 & \quad \cos a = \cos b \cos c + \sin b \sin c \cos A. \end{aligned}$$

Based on this general cosine theorem, the cosine theorem for sides of a spherical triangle applies to both two-dimensional and three-dimensional spheres.

## 3. Formulas for a Spherical Right Triangle

As shown in Figure 11.11.3, consider a right triangle  $ABC$  on a sphere of radius 1. Denote its three angles by  $A, B$ , and  $C$ , with angle  $C$  being the right angle. Let the sides opposite angles  $A, B$ , and  $C$  be  $a, b$ , and  $c$ , respectively. Then the following relations hold:



- 1)  $\cos c = \cos a \cos b$  ;
- 2)  $\cos A = \cos b \sin B$  ;
- 3)  $\cos B = \cos a \sin A$  ;
- 4)  $\cos c = \operatorname{ctg} A \operatorname{ctg} B$  ;
- 5)  $\cos A = \operatorname{tg} b \operatorname{ctg} c$  ;
- 6)  $\cos B = \operatorname{tg} a \operatorname{ctg} c$  ;
- 7)  $\sin b = \sin B \sin c$  ;
- 8)  $\sin a = \sin A \sin c$  ;
- 9)  $\sin b = \operatorname{tg} a \operatorname{tg} A$  ;
- 10)  $\sin a = \operatorname{tg} b \operatorname{tg} B$  .

(11. 11. 4)

**Figure 11.11.3** Right triangle on a sphere

In equation (11.11.3), setting  $C = \pi/2$  gives  $\cos C = 0$ , so equation (11.11.3) becomes  $\cos c = \cos a \cos b$ ,

which is equation (11.11.4).

## §11.12 Toponogov Comparison Theorem

In this section, we introduce the renowned Toponogov comparison theorem.

**Definition 11.12.1** A **triangle** on an  $m$ -dimensional Riemannian manifold  $M$  is a set consisting of three vertices  $p, q, r$  and three segments connecting each pair of distinct vertices, joined end to end, denoted as  $\Delta pqr$ , as shown in Figure 11.12.1.

If one side of a triangle is removed, it is called a **hinge**.

**Definition 11.12.2** As shown in Figure 11.12.2, a **hinge** on an  $m$ -dimensional Riemannian manifold  $M$  consists of two segments  $\gamma_1$  and  $\gamma_2$  emanating from a common point  $p$ , forming an angle  $\alpha$ . The requirements are

$$\gamma_1(l(\gamma_1)) = p = \gamma_2(0),$$

and the angle

$$\alpha = \pi - \angle(\dot{\gamma}_1(l(\gamma_1)), \dot{\gamma}_2(0)).$$

Here,  $l(\gamma_1)$  denotes the arc length of segment  $\gamma_1$ . This definition requires that the endpoint  $\gamma_1(l(\gamma_1)) = p$  of segment  $\gamma_1$  is the starting point  $\gamma_2(0)$  of segment  $\gamma_2$ , and the angle  $\alpha$  is the

interior angle between the two segments  $\gamma_1$  and  $\gamma_2$ . A hinge is denoted as  $(p, \gamma_1, \gamma_2)$ .

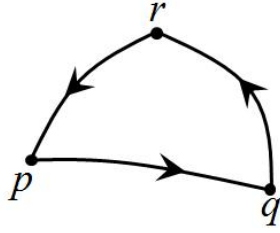


Figure 11.12.1 Triangle

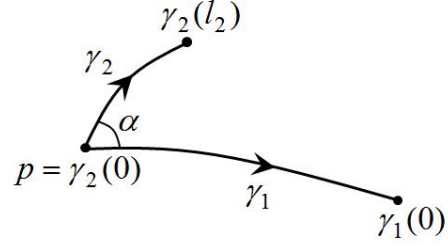


Figure 11.12.2 Hinge

**Theorem 11.12.1** Let  $(M, g)$  be an  $m$ -dimensional complete Riemannian manifold whose sectional curvature satisfies  $K_M \geq k$ , where  $k$  is a fixed constant. Let  $N_k^m$  be an  $m$ -dimensional space of constant curvature  $k$ . Then for every hinge (or triangle) on  $M$ , we can find a corresponding hinge (or triangle) on  $N_k^m$  for comparison, such that their corresponding segments have the same length and the angle is the same (all corresponding segments have the same length).

**Theorem 11.12.2 (Toponogov Comparison Theorem)** Let  $(M, g)$  be an  $m$ -dimensional complete Riemannian manifold whose sectional curvature satisfies  $K_M \geq k$ , where  $k$  is a fixed constant.

**1)Hinge Version:** For any hinge  $(p, \gamma_1, \gamma_2)$  on  $(M, g)$ , there exists a comparison hinge  $(\bar{p}, \bar{\gamma}_1, \bar{\gamma}_2)$  on  $N_k^m$  such that the distance between points  $\gamma_1(0)$  and  $\gamma_2(l_2)$  is less than or equal to the distance between points  $\bar{\gamma}_1(0)$  and  $\bar{\gamma}_2(l_2)$ , i.e.,

$$d(\gamma_1(0), \gamma_2(l_2)) \leq d(\bar{\gamma}_1(0), \bar{\gamma}_2(l_2)).$$

**2)Triangle Version:** For any triangle  $\Delta pqr$  on  $(M, g)$ , there exists a comparison triangle  $\Delta \bar{p}\bar{q}\bar{r}$  on  $N_k^m$  such that the corresponding sides of the two triangles are equal, but the interior angles of triangle  $\Delta pqr$  are greater than or equal to the corresponding angles of triangle  $\Delta \bar{p}\bar{q}\bar{r}$ , i.e.,

$$\alpha \geq \bar{\alpha}, \quad \beta \geq \bar{\beta}, \quad \gamma \geq \bar{\gamma}.$$

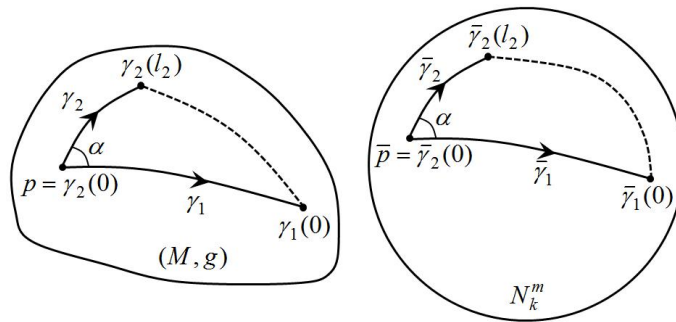


Figure 11.12.3 Comparison of hinges

As shown in Figure 11.12.3, the hinge  $(p, \gamma_1, \gamma_2)$  on the Riemannian manifold  $(M, g)$  and the hinge  $(\bar{p}, \bar{\gamma}_1, \bar{\gamma}_2)$  on the constant curvature space  $N_k^m$  have equal lengths for corresponding sides  $\gamma_1$  and  $\bar{\gamma}_1$ , as well as for corresponding sides  $\gamma_2$  and  $\bar{\gamma}_2$ . The corresponding angles at vertices  $p$  and  $\bar{p}$  are also equal. The distance between vertices  $\gamma_1(0)$  and  $\gamma_2(l_2)$  is  $d(\gamma_1(0), \gamma_2(l_2))$ , while the distance between vertices  $\bar{\gamma}_1(0)$  and  $\bar{\gamma}_2(l_2)$  is  $d(\bar{\gamma}_1(0), \bar{\gamma}_2(l_2))$ . Since the sectional curvature of the Riemannian manifold  $(M, g)$  is greater than or equal to that

of the constant curvature space  $N_k^m$ , according to the Toponogov comparison theorem,

$$d(\gamma_1(0), \gamma_2(l_2)) \leq d(\bar{\gamma}_1(0), \bar{\gamma}_2(l_2)).$$

As shown in Figure 11.12.4, the triangle  $\Delta pqr$  on the Riemannian manifold  $(M, g)$  and the triangle  $\Delta \bar{p}\bar{q}\bar{r}$  on the constant curvature space  $N_k^m$  have equal corresponding sides, i.e.,

$$d(p, q) = d(\bar{p}, \bar{q}), \quad d(q, r) = d(\bar{q}, \bar{r}), \quad d(r, p) = d(\bar{r}, \bar{p}).$$

Since the sectional curvature of the Riemannian manifold  $(M, g)$  is greater than or equal to that of the constant curvature space  $N_k^m$ , according to the Toponogov comparison theorem, the three interior angles of triangle  $\Delta pqr$  are greater than or equal to the corresponding angles of triangle  $\Delta \bar{p}\bar{q}\bar{r}$ , i.e.,

$$\alpha \geq \bar{\alpha}, \quad \beta \geq \bar{\beta}, \quad \gamma \geq \bar{\gamma}.$$

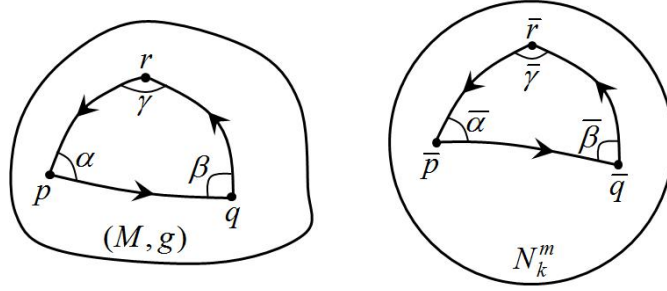


Figure 11.12.4 Comparison of triangles

### §11.13 Hodge Star Operator

Suppose in a local coordinate system  $(U; x^i)$ , the components of the metric tensor  $g$  of an  $m$ -dimensional oriented Riemannian manifold  $(M, g)$  are  $g_{ij}$ ,

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right), \quad (11.13.1)$$

and its volume form is

$$\Omega|_U = \sqrt{G} dx^1 \wedge \cdots \wedge dx^m, \quad (11.13.2)$$

where  $G = \det(g_{ij})$ .

The generalized Kronecker delta is defined as

$$\delta_{j_1 \cdots j_r}^{i_1 \cdots i_r} = \det \begin{pmatrix} \delta_{j_1}^{i_1} & \delta_{j_2}^{i_1} & \cdots & \delta_{j_r}^{i_1} \\ \delta_{j_1}^{i_2} & \delta_{j_2}^{i_2} & \cdots & \delta_{j_r}^{i_2} \\ \cdots & \cdots & \cdots & \cdots \\ \delta_{j_1}^{i_r} & \delta_{j_2}^{i_r} & \cdots & \delta_{j_r}^{i_r} \end{pmatrix} \quad (1 \leq r \leq m). \quad (11.13.3)$$

When  $r=1$ ,  $\delta_{j_1 \cdots j_r}^{i_1 \cdots i_r}$  is the usual Kronecker  $\delta_j^i$ . According to equation (11.13.3), the generalized Kronecker delta possesses the following properties:

- 1)  $\delta_{j_1 \cdots j_r}^{i_1 \cdots i_r}$  is antisymmetric with respect to the indices  $(i_1, \cdots, i_r)$  or  $(j_1, \cdots, j_r)$ .
- 2) If  $i_1 < \cdots < i_r$ ,  $j_1 < \cdots < j_r$ , then  $\delta_{j_1 \cdots j_r}^{i_1 \cdots i_r} = \delta_{j_1}^{i_1} \cdots \delta_{j_r}^{i_r}$ .
- 3)  $\delta_{j_1 \cdots j_r}^{i_1 \cdots i_r} = 1$  when  $i_1, \cdots, i_r$  are all distinct, and  $(j_1, \cdots, j_r)$  is an even permutation of  $(i_1, \cdots, i_r)$ .

$\delta_{j_1 \dots j_r}^{i_1 \dots i_r} = -1$  when  $i_1, \dots, i_r$  are all distinct, and  $(j_1, \dots, j_r)$  is an odd permutation of  $(i_1, \dots, i_r)$ .

$\delta_{j_1 \dots j_r}^{i_1 \dots i_r} = 0$  in all other cases.

**Example 11.13.1** 1) For the generalized Kronecker delta, if the elements of the sets  $(j_1, \dots, j_r)$  and  $(i_1, \dots, i_r)$  are not exactly the same, then  $\delta_{j_1 \dots j_r}^{i_1 \dots i_r} = 0$ . For example,  $\delta_{5678}^{1234} = 0$ .

2)  $\delta_{2134}^{1234} = -1$ , which can be understood as swapping the first and second columns of the matrix determinant (11.13.3), thus resulting in  $-1$ .

$\delta_{2143}^{1234} = 1$ , which can be understood as swapping the first and second columns of the matrix determinant (11.13.3), followed by swapping the third and fourth columns, thus resulting in  $+1$ .

The set of  $r$ -forms on a smooth manifold  $M$  is denoted by  $A^r(M)$ . Let  $\varphi, \omega \in A^r(M)$ . In a local coordinate system  $(U; x^i)$ , they can be expressed as

$$\varphi|_U = \frac{1}{r!} \varphi_{i_1 \dots i_r} dx^{i_1} \wedge \dots \wedge dx^{i_r}, \quad (11.13.4)$$

$$\omega|_U = \frac{1}{r!} \omega_{i_1 \dots i_r} dx^{i_1} \wedge \dots \wedge dx^{i_r}. \quad (11.13.5)$$

**Definition 11.13.1** The **local inner product**  $\langle \varphi, \omega \rangle$  of the  $r$ -forms  $\varphi$  and  $\omega$  is defined as

$$\langle \varphi, \omega \rangle = \frac{1}{r!} \varphi^{i_1 \dots i_r} \omega_{i_1 \dots i_r} = \sum_{i_1 < \dots < i_r} \varphi^{i_1 \dots i_r} \omega_{i_1 \dots i_r}, \quad (11.13.6)$$

where

$$\varphi^{i_1 \dots i_r} = \varphi_{j_1 \dots j_r} g^{j_1 i_1} \dots g^{j_r i_r}. \quad (11.13.7)$$

Since both  $\varphi$  and  $\omega$  are tensor fields on the smooth manifold  $M$ , equation (11.13.6) is independent of the choice of coordinates. Thus,  $\langle \varphi, \omega \rangle$  is a smooth function globally defined on the smooth manifold  $M$ .

The squared norm of the differential form  $\varphi$  at each point of the smooth manifold  $M$  is

$$|\varphi|^2 = \langle \varphi, \varphi \rangle = \frac{1}{r!} \varphi^{i_1 \dots i_r} \varphi_{i_1 \dots i_r} \geq 0. \quad (11.13.8)$$

**Definition 11.13.2** Let  $(M, g)$  be a compact  $m$ -dimensional Riemannian manifold. The **global inner product**  $(\varphi, \omega)$  of the  $r$ -forms  $\varphi$  and  $\omega$  is defined as

$$(\varphi, \omega) = \int_M \langle \varphi, \omega \rangle \Omega.$$

The squared norm of the differential form  $\varphi$  over the entire Riemannian manifold  $M$  is

$$\|\varphi\|^2 = \int_M \langle \varphi, \varphi \rangle \Omega.$$

Let

$$*\omega|_U = \frac{1}{(m-r)!} \frac{\sqrt{G}}{r!} \delta_{i_1 \dots i_m}^{1 \dots m} \omega_{i_1 \dots i_r} dx^{i_{r+1}} \wedge \dots \wedge dx^{i_m}. \quad (11.13.9)$$

The right-hand side of equation (11.13.9) is invariant under orientation-preserving coordinate transformations. Therefore,  $*\omega$  are an  $(m-r)$ -forms defined on the entire Riemannian manifold  $M$ .

**Definition 11.13.3** The linear map  $*$ :  $A^r(M) \rightarrow A^{m-r}(M)$  is called the **Hodge star operator** on the oriented Riemannian manifold  $(M, g)$ .

From equations (11.13.7) and (11.13.9), we obtain

$$*\omega|_U = \frac{1}{(m-r)!} \frac{\sqrt{G}}{r!} \delta_{i_1 \dots i_m}^{1 \dots m} \omega_{j_1 \dots j_r} g^{j_1 i_1} \dots g^{j_r i_r} dx^{i_{r+1}} \wedge \dots \wedge dx^{i_m}. \quad (11.13.10)$$

**Theorem 11.13.1** Let  $(M, g)$  be a compact oriented  $m$ -dimensional Riemannian manifold. Then the Hodge star operator  $*$  has the following properties:

1) For any  $\varphi, \omega \in A^r(M)$ , the following holds:

$$\varphi \wedge *\omega = \langle \varphi, \omega \rangle \Omega. \quad (11.13.11)$$

2)  $*\Omega = 1$ ,  $*1 = \Omega$ .

3) For any  $\varphi \in A^r(M)$ , the following holds:

$$* \circ * \varphi = (-1)^{rm+r} \varphi,$$

i.e.,

$$* \circ * \varphi = (-1)^{rm+r} \cdot \text{id} : A^r(M) \rightarrow A^r(M). \quad (11.13.12)$$

4) For any  $\varphi, \omega \in A^r(M)$ , the following holds:

$$(*\varphi, *\omega) = (\varphi, \omega). \quad (11.13.13)$$

If the Riemannian manifold  $(M, g)$  is three-dimensional, i.e.,  $m = 3$ , then the Hodge star operator  $*$  maps a 2-form  $\omega$  to a 1-form  $\lambda$ . According to equation (11.13.13), the norms of  $\omega$  and  $\lambda$  over the entire  $(M, g)$  are equal.

**Example 11.13.2** Let  $M$  be a regular surface in  $R^3$  with metric form

$$ds^2 = g_{\alpha\beta} du^\alpha \wedge du^\beta, \quad 1 \leq \alpha, \beta \leq 2.$$

We compute  $*du^1$ ,  $*du^2$ ,  $*1$ ,  $*(du^1 \wedge du^2)$ .

First, from

$$\begin{pmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

we obtain

$$\begin{pmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}^{-1} = \frac{\begin{pmatrix} (-1)^{1+1} g_{22} & (-1)^{1+2} g_{21} \\ (-1)^{2+1} g_{12} & (-1)^{2+2} g_{11} \end{pmatrix}}{\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix}} = \frac{1}{G} \begin{pmatrix} g_{22} & -g_{21} \\ -g_{12} & g_{11} \end{pmatrix},$$

$$G = g_{11}g_{22} - g_{12}^2.$$

Thus,

$$g^{11} = \frac{1}{G} g_{22}, \quad g^{12} = -\frac{1}{G} g_{21}, \quad g^{21} = -\frac{1}{G} g_{12}, \quad g^{22} = \frac{1}{G} g_{11}.$$

1)  $du^1$  is a 1-form, so  $r = 1$ . The Hodge star operator  $*$  maps a 1-form  $du^1$  to another 1-form because  $m - r = 2 - 1 = 1$ . According to equation (11.13.5), a general 1-form can be expressed as

$$\omega = \frac{1}{1!} \omega_{j_1} dx^{j_1} = \omega_1 dx^1 + \omega_2 dx^2 = \omega_1 du^1 + \omega_2 du^2.$$

Thus, for  $\omega = du^1$ , we have  $\omega_1 = 1$ ,  $\omega_2 = 0$ . From equation (11.13.10), we obtain

$$\begin{aligned} *du^1 &= \frac{1}{(2-1)!} \frac{\sqrt{G}}{1!} \delta_{i_1 i_2}^{12} \omega_{j_1} g^{j_1 i_1} du^{i_2} \\ &= \sqrt{G} \delta_{i_1 i_2}^{12} \omega_{j_1} g^{j_1 i_1} du^{i_2} = \sqrt{G} (\delta_{12}^{12} \omega_{j_1} g^{j_1 1} du^2 + \delta_{21}^{12} \omega_{j_1} g^{j_1 2} du^1) \\ &= \sqrt{G} (\omega_1 g^{11} du^2 + \omega_2 g^{21} du^2 - \omega_1 g^{12} du^1 - \omega_2 g^{22} du^1) \\ &= \sqrt{G} (g^{11} du^2 - g^{12} du^1) = \sqrt{G} \left( \frac{1}{G} g_{22} du^2 + \frac{1}{G} g_{21} du^1 \right) \\ &= \frac{1}{\sqrt{G}} (g_{22} du^2 + g_{21} du^1). \end{aligned}$$

2)  $du^2$  is a 1-form, so  $r = 1$ . For  $\omega = du^2$ , we have  $\omega_1 = 0$ ,  $\omega_2 = 1$ . From equation

(11.13.10), we obtain

$$\begin{aligned}
*du^2 &= \frac{1}{(2-1)!} \frac{\sqrt{G}}{1!} \delta_{i_1 i_2}^{12} \omega_{j_1} g^{j_1 i_1} du^{i_2} \\
&= \sqrt{G} \delta_{i_1 i_2}^{12} \omega_{j_1} g^{j_1 i_1} du^{i_2} = \sqrt{G} (\delta_{12}^{12} \omega_{j_1} g^{j_1 1} du^2 + \delta_{21}^{12} \omega_{j_1} g^{j_1 2} du^1) \\
&= \sqrt{G} (\omega_1 g^{11} du^2 + \omega_2 g^{21} du^2 - \omega_1 g^{12} du^1 - \omega_2 g^{22} du^1) \\
&= \sqrt{G} (g^{21} du^2 - g^{22} du^1) = \sqrt{G} \left( -\frac{1}{G} g_{12} du^2 - \frac{1}{G} g_{11} du^1 \right) \\
&= -\frac{1}{\sqrt{G}} (g_{12} du^2 + g_{11} du^1).
\end{aligned}$$

3) The real number 1 is a 0-form, so  $r = 0$ . The Hodge star operator  $*$  maps the 0-form 1 to a 2-form because  $m - r = 2 - 0 = 2$ . Since 1 is a 0-form, no index raising or lowering is needed, so the sum over indices in equation (11.13.10) reduces to a single term, i.e.,  $\omega_{j_1 \dots j_r} g^{j_1 i_1} \dots g^{j_r i_r} = 1$ . From equation (11.13.10), we obtain

$$\begin{aligned}
*1 &= \frac{1}{(2-0)!} \frac{\sqrt{G}}{0!} \delta_{i_1 i_2}^{12} du^{i_1} \wedge du^{i_2} \\
&= \frac{1}{2} \sqrt{G} (\delta_{12}^{12} du^1 \wedge du^2 + \delta_{21}^{12} du^2 \wedge du^1) = \frac{1}{2} \sqrt{G} (du^1 \wedge du^2 - du^2 \wedge du^1) \\
&= \frac{1}{2} \sqrt{G} (du^1 \wedge du^2 + du^1 \wedge du^2) = \sqrt{G} du^1 \wedge du^2.
\end{aligned}$$

4) A general 2-form  $\omega$  can be expressed as

$$\begin{aligned}
\omega &= \frac{1}{2!} \omega_{j_1 j_2} dx^{j_1} \wedge dx^{j_2} = \frac{1}{2} (\omega_{1j_2} dx^1 \wedge dx^{j_2} + \omega_{2j_2} dx^2 \wedge dx^{j_2}) \\
&= \frac{1}{2} (\omega_{11} dx^1 \wedge dx^1 + \omega_{12} dx^1 \wedge dx^2 + \omega_{21} dx^2 \wedge dx^1 + \omega_{22} dx^2 \wedge dx^2) \\
&= \frac{1}{2} (\omega_{12} dx^1 \wedge dx^2 + \omega_{12} dx^1 \wedge dx^2) = \omega_{12} dx^1 \wedge dx^2.
\end{aligned}$$

For the 2-form  $\omega = du^1 \wedge du^2$ , we have  $r = 2$ , and the coefficient in front of  $du^1 \wedge du^2$  is  $\omega_{12} = 1$ . Since  $\omega_{j_1 j_2}$  is antisymmetric,  $\omega_{21} = -1$ . The Hodge star operator  $*$  maps this 2-form to a 0-form because  $m - r = 2 - 2 = 0$ . From equation (11.13.10), we obtain

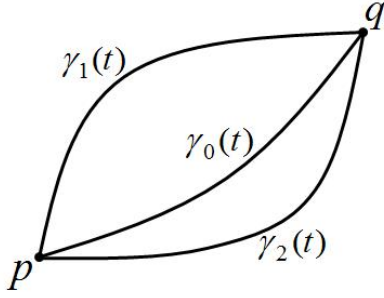
$$\begin{aligned}
*(du^1 \wedge du^2) &= \frac{1}{(2-2)!} \frac{\sqrt{G}}{2!} \delta_{i_1 i_2}^{12} \omega_{j_1 j_2} g^{j_1 i_1} g^{j_2 i_2} \\
&= \frac{1}{2} \sqrt{G} (\delta_{12}^{12} \omega_{j_1 j_2} g^{j_1 1} g^{j_2 2} + \delta_{21}^{12} \omega_{j_1 j_2} g^{j_1 2} g^{j_2 1}) \\
&= \frac{1}{2} \sqrt{G} (\omega_{j_1 j_2} g^{j_1 1} g^{j_2 2} - \omega_{j_1 j_2} g^{j_1 2} g^{j_2 1}) \\
&= \frac{1}{2} \sqrt{G} (\omega_{12} g^{11} g^{22} + \omega_{21} g^{21} g^{12} - \omega_{12} g^{12} g^{21} - \omega_{21} g^{22} g^{11}) \\
&= \frac{1}{2} \sqrt{G} (\omega_{12} g^{11} g^{22} - \omega_{12} g^{21} g^{12} - \omega_{12} g^{12} g^{21} + \omega_{12} g^{22} g^{11}) \\
&= \sqrt{G} (g^{11} g^{22} - g^{21} g^{12}) = \sqrt{G} \left[ \frac{1}{G} g_{22} \frac{1}{G} g_{11} - \left( -\frac{1}{G} g_{12} \right) \left( -\frac{1}{G} g_{21} \right) \right] \\
&= \frac{\sqrt{G}}{G^2} (g_{22} g_{11} - g_{12} g_{21}) = \frac{\sqrt{G}}{G^2} G = \frac{1}{\sqrt{G}}.
\end{aligned}$$



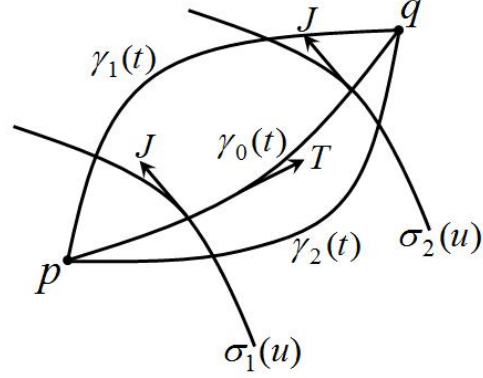
## §11.14 Jacobi Fields

Let  $(M, g)$  be an  $m$ -dimensional Riemannian manifold, and let  $\gamma : [a, b] \rightarrow M$  be a geodesic connecting points  $p$  and  $q$  on  $M$ , as shown in Figure 11.14.1. Besides the geodesic  $\gamma$ , there may be infinitely many other geodesics connecting  $p$  and  $q$ , forming a one-parameter family of geodesics  $\{\gamma = \gamma_0(t), \gamma_1(t), \gamma_2(t), \dots\}$ , expressed as

$$\varphi : [a, b] \times (-\varepsilon, \varepsilon) \rightarrow M.$$



**Figure 11.14.1** Family of geodesics between points  $p$  and  $q$



**Figure 11.14.2** Jacobi field  $J$

Let  $t \in [a, b]$ ,  $u \in (-\varepsilon, \varepsilon)$ . When  $u$  is fixed,  $\varphi(t, u) = \gamma_u(t)$  is a geodesic from  $p$  to  $q$ . When  $u = 0$ , it is the original geodesic

$$\gamma(t) = \gamma_0(t) = \varphi(t, 0).$$

Define

$$T = \varphi_* \left( \frac{\partial}{\partial t} \right), \quad J = \varphi_* \left( \frac{\partial}{\partial s} \right).$$

$T$  is the tangent vector field along the geodesic  $\gamma_u(t) = \varphi(t, u)$ .  $J$  is the tangent vector field along the curve  $\sigma_i(u) = \varphi(t, u)$  when  $t$  is fixed, as shown in Figure 11.14.2.

Since  $t$  and  $u$  are coordinate bases and independent of each other, we have  $\left[ \frac{\partial}{\partial t}, \frac{\partial}{\partial s} \right] = 0$ , and thus

$$[T, J] = \left[ \varphi_* \left( \frac{\partial}{\partial t} \right), \varphi_* \left( \frac{\partial}{\partial s} \right) \right] = \varphi_* \left[ \frac{\partial}{\partial t}, \frac{\partial}{\partial s} \right] = 0. \quad (11.14.1)$$

Taking the connection on the Riemannian manifold  $(M, g)$  to be the Levi-Civita connection, which is torsion-free and compatible, the torsion tensor vanishes. Therefore, from equation (10.3.14), we obtain

$$T(T, J) = D_T J - D_J T - [T, J] = 0, \quad (11.14.2)$$

and together with equation (11.14.1), this yields

$$D_T J = D_J T. \quad (11.14.3)$$

From equation (10.3.2), we have

$$R(T, J)T = D_T D_J T - D_J D_T T - D_{[T, J]} T. \quad (11.14.4)$$

Since  $T$  is the tangent vector field of the geodesic  $\gamma$ , we have

$$D_T T = 0. \quad (11.14.5)$$

Substituting equations (11.14.1), (11.14.3), and (11.14.5) into equation (11.14.4) gives

$$D_T D_T J = R(T, J)T. \quad (11.14.6)$$

Equation (11.14.6) is called the **Jacobi equation**.

Let  $J = J(t)$  be a smooth tangent vector field defined along the geodesic  $\gamma$ , which can be

expressed as

$$J(t) = \sum_{i=1}^m J^i(t) e_i(t),$$

then equation (11.14.6) can be written in component form as

$$(J^i(t))'' = \sum_j J^j(t) R_{mimj}, \quad 1 \leq i \leq m, \quad (11.14.7)$$

where  $R_{mimj} = \langle R(e_m, e_j) e_m, e_i \rangle$ .

**Definition 11.14.1** Let  $\gamma: [a, b] \rightarrow M$  be a geodesic on an  $m$ -dimensional Riemannian manifold  $(M, g)$ , and let  $T$  be the tangent vector field of this geodesic. If a smooth tangent vector field  $J$  along  $\gamma$  satisfies the Jacobi equation, then  $J$  is called a **Jacobi field** along the geodesic  $\gamma$ .

**Theorem 11.14.1** Let  $\gamma: [a, b] \rightarrow M$  be a geodesic on an  $m$ -dimensional Riemannian manifold  $(M, g)$ , let  $T$  be the tangent vector field of this geodesic, and let  $J = J(t)$  be a Jacobi field along  $\gamma$ . Then  $J$  possesses the following properties:

- 1) All Jacobi fields along  $\gamma$  form a  $2m$ -dimensional linear space.
- 2) The condition that  $J$  is orthogonal to  $T$ , i.e.,  $J \cdot T = 0$ , holds if and only if

$$J(a) \cdot T(a) = 0, \quad (D_T J(a)) \cdot T(a) = 0.$$

**Definition 11.14.2** Let  $J = J(t)$  be a Jacobi field along a geodesic  $\gamma$  on a Riemannian manifold  $(M, g)$ . If  $J$  is everywhere orthogonal to  $\gamma$ , then  $J$  is called a **normal Jacobi field** along  $\gamma$ .

**Definition 11.14.3** Let  $\gamma: [a, b] \rightarrow M$  be a geodesic on a Riemannian manifold  $(M, g)$ . If its tangent vector  $T$  is a unit vector, i.e.,  $\langle T, T \rangle = 1$ , then the geodesic is called a **normal geodesic**.

Let  $(M, g)$  be an  $m$ -dimensional space of constant curvature with sectional curvature  $K$ . From Definition 11.2.1 of sectional curvature, we have

$$R(X, Y, X, Y) = -Kg(X, Y, X, Y),$$

Substituting equation (11.2.4) yields

$$\begin{aligned} (R(X, Y)X) \cdot Y &= R(X, Y, X, Y) = -Kg(X, Y, X, Y) \\ &= -K[g(X, X)g(Y, Y) - g(X, Y)g(Y, X)], \end{aligned}$$

and thus

$$R(X, Y)X = -K[g(X, X)Y - g(Y, X)X]. \quad (11.14.8)$$

Let  $\gamma: [a, b] \rightarrow M$  be a normal geodesic on an  $m$ -dimensional constant curvature space  $(M, g)$ , with tangent vector  $T$ , and let  $J = J(t)$  be a normal Jacobi field along  $\gamma$ . Then

$$g(T, T) = 1, \quad g(T, J) = 0.$$

From equation (11.14.8), we obtain

$$R(T, J)T = -K[g(T, T)J - g(J, T)T] = -KJ. \quad (11.14.9)$$

**Theorem 11.14.2** Let  $\gamma: [a, b] \rightarrow M$  be a normal geodesic on an  $m$ -dimensional constant curvature space  $(M, g)$ , and let  $J = J(t)$  be a normal Jacobi field along  $\gamma$ . Then the Jacobi equation satisfied by  $J$  becomes

$$D_T D_T J = -KJ, \quad (11.14.10)$$

where  $K$  is the sectional curvature of the constant curvature space  $(M, g)$ .

Along the normal geodesic  $\gamma$ , choose a parallel orthonormal frame field  $\{e_1(t), \dots, e_m(t)\}$  such that  $e_m(t) = T$ . Write

$$J(t) = \sum_i J^i(t) e_i(t),$$

then we obtain the system of linear homogeneous ordinary differential equations with constant coefficients for  $J^i(t)$ :

$$(J^i(t))'' + KJ^i(t) = 0, \quad 1 \leq i \leq m-1. \quad (11.14.11)$$

Since  $J$  is orthogonal to  $e_m(t) = T$ , the projection of  $J$  onto  $e_m(t)$  is zero, so  $J^m(t) = 0$ .

Therefore, the system (11.14.11) consists of only  $m - 1$  equations.

## Chapter 12 Lie Groups

### §12.1 Topological Groups

**Definition 12.1.1** Let  $G$  be both a topological space and a group, such that the group multiplication and inversion operations are continuous. Then  $G$  is called a **topological group**.

**Theorem 12.1.1** Let  $G$  be a topological group, and let  $N$  be the connected component of the identity element  $e$  in  $G$ . Then  $N$  is a **normal subgroup** of  $G$ .

**Theorem 12.1.2** Let  $G$  be a connected topological group, and let  $U$  be an open neighborhood of the identity element  $e$ . Then  $G$  is generated by  $U$ , i.e., for any  $g \in G$ , there exist finitely many elements  $u_i \in U$  such that  $g = u_1 \cdots u_i$ .

**Theorem 12.1.3** Let  $G$  be a connected topological group. Then for any open neighborhood  $U$  of the identity element  $e$ , there exists an open neighborhood  $V$  of  $e$  such that  $V \subset U$  and  $G = \bigcup_{i=1}^{\infty} V^i$ .

**Theorem 12.1.4** Let  $G$  be a topological group. If  $H$  is an open subgroup of  $G$ , then  $H$  is also a closed subgroup of  $G$ .

**Theorem 12.1.5** If  $G$  is a topological manifold and both the multiplication and inversion operations are continuous, then there exists an analytic structure on  $G$  such that the multiplication and inversion operations are analytic, making  $G$  a Lie group.

A Lie group is also a topological group. The following sections cover topics related to Lie groups.

### §12.2 Lie Groups and Lie Algebras

#### 1. Lie Groups

**Definition 12.2.1** Let  $G$  be a non-empty set.  $G$  is called an  **$r$ -dimensional Lie group** if it satisfies the following three conditions:

- 1)  $G$  is a group (group operation denoted as multiplication);
- 2)  $G$  is an  $r$ -dimensional smooth manifold;
- 3) The inverse map  $\tau: G \rightarrow G$ , defined by  $\tau(g) = g^{-1}$ , and the multiplication map  $\varphi: G \times G \rightarrow G$ , defined by  $\varphi(g_1, g_2) = g_1 \cdot g_2$ , are both smooth maps.

According to the definition of a Lie group, the existence of the inverse map  $\tau: G \rightarrow G$ , defined by  $\tau(g) = g^{-1}$ , implies

$$\begin{aligned}\tau(\tau(g)) &= \tau(g^{-1}) = (g^{-1})^{-1} = g, \\ \tau^2(g) &= g,\end{aligned}$$

i.e.,

$$\tau^2 = id: G \rightarrow G.$$

$\tau^2$  is the identity map. Since  $\tau$  is a smooth map and  $\tau^{-1} = \tau$  is also smooth, the inverse map  $\tau$  is a diffeomorphism from  $G$  to itself.

#### 2. Right and Left Translations

Since  $G$  is a group, multiplication operations exist among its elements. Let  $g_1, g_2 \in G$ . Then  $g_2$  can be multiplied on the left by  $g_1$ , i.e.,  $g_1 g_2$ , or on the right by  $g_1$ , i.e.,  $g_2 g_1$ . This gives rise to two families of diffeomorphisms on the group  $G$ , called **left translations** and **right translations**.

Let  $g \in G$  be an element of  $G$ . The multiplication of  $g$  with another element  $x$  (including  $g$  itself) on the right is  $x \cdot g$ . Denote

$$R_g(x) = \varphi(x, g) = x \cdot g.$$

Since

$$R_{g^{-1}} \circ R_g x = R_{g^{-1}}(R_g(x)) = \varphi(\varphi(x, g), g^{-1}) = (x \cdot g) \cdot g^{-1} = x,$$

$$R_g \circ R_{g^{-1}} x = R_g(R_{g^{-1}}(x)) = \varphi(\varphi(x, g^{-1}), g) = (x \cdot g^{-1}) \cdot g = x,$$

we have

$$R_g \circ R_{g^{-1}} = R_{g^{-1}} \circ R_g = \text{id},$$

or

$$(R_g)^{-1} = R_{g^{-1}}.$$

According to the definition, both  $R_g$  and  $R_{g^{-1}}$  are smooth maps. Since  $(R_g)^{-1} = R_{g^{-1}}$ , the inverse map  $(R_g)^{-1}$  is also smooth. Therefore,  $R_g : G \rightarrow G$  is a smooth diffeomorphism.

**Definition 12.2.2** Let  $g \in G$ . Multiplying every element of  $G$  on the right by  $g$  is denoted as

$$R_g(x) = \varphi(x, g) = x \cdot g,$$

and the resulting mapping

$$R_g : G \rightarrow G$$

is a diffeomorphism, called the **right translation** on  $G$  induced by  $g$ .

Similarly, left translations can be defined.

**Definition 12.2.3** Let  $g \in G$ . Multiplying every element of  $G$  on the left by  $g$  is denoted as

$$L_g(x) = \varphi(g, x) = g \cdot x,$$

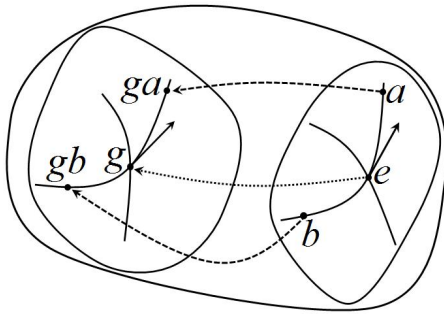
and the resulting mapping

$$L_g : G \rightarrow G$$

is a diffeomorphism, called the **left translation** on  $G$  induced by  $g$ .

### 3. Lie Algebras

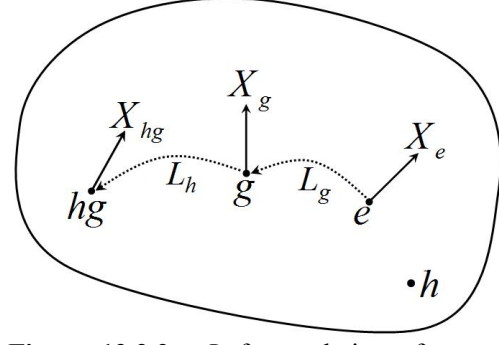
Let the identity element of the  $r$ -dimensional Lie group  $G$  be  $e$ , and let  $T_e G$  and  $T_g G$  denote the tangent spaces of the smooth manifold  $G$  at points  $e$  and  $g$ , respectively. Since for each element  $g \in G$ , the left translation  $L_g$  is a diffeomorphism and satisfies  $L_g(g^{-1}) = e$ , the induced tangent map  $(L_g)_* : T_e G \rightarrow T_g G$  is a linear isomorphism.



**Figure 12.2.1** Left translation of points

As shown in Figure 12.2.1,  $g$  induces a left translation  $L_g$ , which maps the identity element  $e$  to the element  $g$ , maps a neighborhood of  $e$  to a neighborhood of  $g$ , maps a point  $a$  to  $ga$ , maps a point  $b$  to  $gb$ , and consequently maps a curve connecting  $a$  and  $b$  to a curve connecting  $ga$  and  $gb$ . It also maps a tangent vector at point  $e$  to a tangent vector at point  $g$ .

As shown in Figure 12.2.2, let a tangent vector  $X_e \in T_e G$ . An element  $g \in G$  induces a left translation  $L_g$ , which maps the tangent vector  $X_e$  to  $X_g \in T_g G$ . Similarly, another element  $h \in G$  induces another left translation  $L_h$ , mapping the tangent vector  $X_g$  to  $X_{hg} \in T_{hg} G$ . Other elements in  $G$  can also generate similar translation processes. Thus, an element  $X_e \in T_e G$ , through left translations, generates a tangent vector field on  $G$ , denoted as  $X(a)$ , where  $a$  is an arbitrary element of the Lie group  $G$ .



**Figure 12.2.2** Left translation of tangent vectors

When  $a$  is an arbitrary element of the Lie group  $G$ , we use  $X(a)$  to denote the tangent vector field  $X$  on  $G$ . When  $a$  represents a specific point (or element) on the Lie group  $G$ ,  $X_a$  denotes the tangent vector at point  $a$ .

Since any element  $a \in G$  can left-translate any tangent vector on  $G$  to other points, it can also left-translate tangent vector fields on  $G$ .

**Definition 12.2.4** Let  $g, h$  be arbitrary elements of the Lie group  $G$ ,  $g, h \in G$ , and let  $X$  be a smooth tangent vector field on  $G$ . If the left translation  $L_h$  maps the tangent vector  $X_g$  of the smooth vector field  $X$  at point  $g$  to the tangent vector  $X_h$  of  $X$  at point  $h$ , i.e.,

$$(L_h)_* X_g = X_h,$$

then  $X$  is called a **left-invariant vector field** on the Lie group  $G$ .

Since  $X_g$  and  $X_h$  belong to the same smooth tangent vector field  $X$ , a left-invariant vector field can also be succinctly defined as:

Let  $X$  be a smooth tangent vector field on the Lie group  $G$ . If for any  $h \in G$ ,

$$(L_h)_* X = X,$$

then  $X$  is called a left-invariant tangent vector field on  $G$ .

An arbitrary element  $g$  in the Lie group  $G$  gives rise to a left translation  $L_g$ . Let the tangent vector field generated by left-translating the tangent vector  $X_e \in T_e G$  be denoted as  $X(g)$ , i.e.,

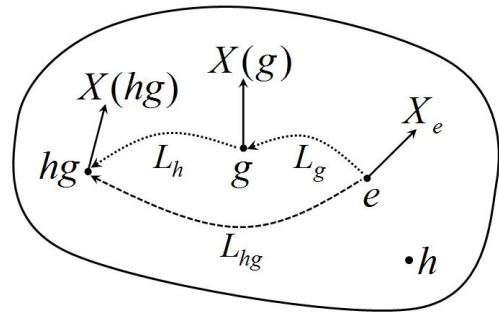
$$X(g) = (L_g)_* X_e.$$

Another element  $h$  in  $G$  gives rise to a left translation  $L_h$ . Applying  $L_h$  to  $X(g)$  yields

$$(L_h)_* X(g) = (L_h)_* \circ (L_g)_* X_e = (L_h \circ L_g)_* X_e = (L_{hg})_* X_e = X(hg).$$

Both  $X(g)$  and  $X(hg)$  belong to the same tangent vector field. Therefore, the tangent vector field  $X(g)$  is invariant under the left translation  $L_h$  ( $\forall h \in G$ ).

As shown in Figure 12.2.3, the product of two elements  $h$  and  $g$  in the Lie group  $G$  is  $hg$ , which induces a left translation  $L_{hg}$ . The tangent vector field obtained by left-translating the element  $X_e$  from  $T_e G$  via  $L_{hg}$  is exactly  $(L_{hg})_* X_e = X(hg)$ . This is the same tangent vector field as that generated by the left translation  $L_g$  (induced by the element  $g$  in  $G$ ) acting on the tangent vector  $X_e$  to produce  $X(g)$ . Therefore, the tangent vector field  $X$  generated on  $G$  by left-translating the tangent vector  $X_e \in T_e G$  is a left-invariant vector field.



**Figure 12.2.3** Composite left translation

Let  $\forall g \in G$ , and  $Y(g)$  be left-invariant vector fields on the Lie group  $G$ . An element  $h$  in

the group  $G$  induces a left translation  $L_h$ , then

$$(L_h)_* Y(g) = Y(hg), \quad \forall g \in G.$$

If we set  $g = e$ , we obtain

$$(L_h)_* Y(e) = Y(he) = Y(h).$$

Thus, let  $\forall g \in G$ , the left-invariant vector fields  $Y(g)$  can be generated by left-translating the elements  $Y_e$  from the tangent space  $T_e G$  at the identity element  $e$ .

In summary, we have the following theorem.

**Theorem 12.2.1** Let  $G$  be an  $r$ -dimensional Lie group. A smooth tangent vector field  $X$  generated by left-translating a tangent vector from  $T_e G$  is invariant under left translations on the Lie group  $G$ . Conversely, any tangent vector field on  $G$  that is invariant under left translations is generated by left-translating some tangent vector from  $T_e G$ .

Let the basis of the tangent space  $T_e G$  at the identity element  $e$  be  $\{\delta_1, \dots, \delta_r\}$ . Denote by  $X_i$  the left-invariant vector field generated by left-translating  $\delta_i \in T_e G$  ( $1 \leq i \leq r$ ). Then there are  $r$  everywhere linearly independent left-invariant tangent vector fields  $X_i$  ( $1 \leq i \leq r$ ) on  $G$ , forming a basis for an  $r$ -dimensional vector space, denoted as  $\text{Lie}(G)$ .

**Theorem 12.2.2** Let  $G$  be an  $r$ -dimensional Lie group, and let  $\{X_1, \dots, X_r\}$  be a basis of  $\text{Lie}(G)$ . Then any left-invariant vector field on  $G$  can be expressed as a constant-coefficient linear combination of the  $X_i$  ( $1 \leq i \leq r$ ). Conversely, any smooth tangent vector field that is a constant-coefficient linear combination of the  $X_i$  is a left-invariant tangent vector field on the Lie group  $G$ .

The set of left-invariant vector fields on  $G$ , denoted as  $\text{Lie}(G)$ , is isomorphic to  $T_e G$ .

**Theorem 12.2.3** If  $X$  and  $Y$  are left-invariant vector fields on the Lie group  $G$ , then the Poisson bracket  $[X, Y]$  is also a left-invariant vector field on  $G$ .

According to Theorem 12.2.2,  $[X, Y]$  can be expressed as a constant-coefficient linear combination of other left-invariant vector fields:

$$[X_j, X_k] = \sum_{i=1}^r C_{j\ k}^i X_i,$$

where the  $C_{j\ k}^i$  are constants. Therefore, the Poisson bracket is closed in  $\text{Lie}(G)$ , defining a multiplication operation in  $\text{Lie}(G)$ . This multiplication satisfies the following three laws:

- 1) Distributivity  $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ ;
- 2) Anticommutativity  $[X, Y] = -[Y, X]$ ;
- 3) Jacobi Identity  $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ .

An  $n$ -dimensional vector space equipped with a multiplication operation satisfying the above three laws is called an  **$n$ -dimensional Lie algebra**. The set of all smooth tangent vector fields on an  $m$ -dimensional manifold  $M$  forms an infinite-dimensional vector space, which becomes an infinite-dimensional Lie algebra under the Poisson bracket. The vector space consisting of all left-invariant vector fields on an  $r$ -dimensional Lie group  $G$  forms an  $r$ -dimensional Lie algebra under the Poisson bracket, denoted by  $\text{Lie}(G)$ . We call the Lie algebra  $\text{Lie}(G)$  the **Lie algebra of the Lie group  $G$** .

The constants  $C_{j\ k}^i$  are called the **structure constants** of the Lie group. There are  $r^3$  such constants in total, and they satisfy the following identities:

- 1)  $C_{j\ k}^i + C_{k\ j}^i = 0$ ;
- 2)  $\sum_{j=1}^r (C_{j\ k}^i C_{h\ l}^j + C_{j\ h}^i C_{l\ k}^j + C_{j\ l}^i C_{k\ h}^j) = 0$ .

Here,  $C_{j\ k}^i$  is a tensor of type  $(1, 2)$ .

Since the set  $\text{Lie}(G)$  of left-invariant vector fields on  $G$  is isomorphic to  $T_e G$ , we can also define

$$[\delta_j, \delta_k] = \sum_{i=1}^r C_{j\ k}^i \delta_i \quad (12.2.1)$$

in  $T_e G$ , making  $T_e G$  an  $r$ -dimensional Lie algebra as well. Thus,  $T_e G$  and  $\text{Lie}(G)$  are not only linearly isomorphic as vector spaces but also isomorphic as Lie algebras. Therefore, we may also refer to the Lie algebra  $T_e G$  as the **Lie algebra of the Lie group  $G$** .

On a Lie group, one can also discuss **right-invariant vector fields**  $\tilde{X}_i$  and define the Poisson bracket for right-invariant vector fields:

$$\begin{aligned} [\tilde{X}_j, \tilde{X}_k] &= \sum_{i=1}^r \tilde{C}_{j\ k}^i \tilde{X}_i, \\ [\delta_j, \delta_k]_{\text{right}} &= \sum_{i=1}^r \tilde{C}_{j\ k}^i \delta_i. \end{aligned}$$

However, the structure constants  $\tilde{C}_{j\ k}^i$  for right-invariant vector fields differ by a sign from those for left-invariant vector fields:

$$C_{j\ k}^i = -\tilde{C}_{j\ k}^i, \quad (12.2.2)$$

and therefore

$$\begin{aligned} [\delta_i, \delta_j]_{\text{left}} &= [X_i, X_j]_e, \\ [\delta_i, \delta_j]_{\text{right}} &= [\tilde{X}_i, \tilde{X}_j]_e, \\ [\delta_i, \delta_j]_{\text{right}} &= -[\delta_i, \delta_j]_{\text{left}}. \end{aligned}$$

Here,  $[\ , \ ]_{\text{right}}$  and  $[\ , \ ]_{\text{left}}$  denote the bracket products defined on right-invariant and left-invariant vector fields, respectively.

We adopt the convention that when  $T_e G$  is referred to as the Lie algebra of the Lie group  $G$ , its multiplication operation is denoted by  $[\ , \ ]_{\text{left}}$ , i.e., the multiplication given by equation (12.2.1).

#### 4. The Tangent Bundle of a Lie Group is Trivial

**Theorem 12.2.4** The tangent bundle  $TG$  of an  $r$ -dimensional Lie group  $G$  is trivial, i.e.,

$$TG \cong G \times T_e G \cong G \times \mathbb{R}^r.$$

The triviality of the tangent bundle  $TG$  of a Lie group  $G$  means that, globally, the tangent bundle  $TG$  equals the product of the Lie group  $G$  and the tangent space  $T_e G$  at the identity element  $e$ , which is also equal to the product of  $G$  and the  $r$ -dimensional vector space  $\mathbb{R}^r$ . Therefore, the tangent space  $T_e G$  at the identity element  $e$  of the Lie group  $G$  holds a typical representational role. Because the tangent bundle  $TG$  of the Lie group  $G$  is trivial, when we refer to the **Lie algebra of the Lie group  $G$** , we sometimes mean the tangent space  $T_e G$  at the identity element  $e$ , and other times mean the vector space formed by the left-invariant (or right-invariant) vector fields generated on the entire Lie group  $G$  by left (or right) translating elements of  $T_e G$ . The vector space consisting of all left-invariant (or right-invariant) vector fields on  $G$  is isomorphic to  $T_e G$ .

### §12.3 Local Lie Groups

The elements in a neighborhood (or open set) of the identity element  $e$  of a Lie group form a group, called a **local Lie group** or an **infinitesimal group**. Below, we demonstrate that a neighborhood of  $e$  indeed constitutes a group.



Suppose there exists a homeomorphism  $g$  between a neighborhood  $U = O(e)$  of  $e$  and a neighborhood  $E = O(0)$  of the origin  $0$  in the  $r$ -dimensional Euclidean space  $R^r$  :

$$g : E \rightarrow U, \text{ such that } \begin{cases} g(0) = e, \\ g(x) = g(x^1, \dots, x^r) \in U. \end{cases}$$

Let  $g(x^1, \dots, x^r)$  be a function defined on the neighborhood  $E = O(0)$ . The coordinates  $(x^1, \dots, x^r)$  of a point  $x \in E$  are mapped via  $g$  to serve as the coordinates of the point  $p \in O(e) = U$ .

Assuming  $g(x^1, \dots, x^r)$  is an analytic function, it can be expanded into a Taylor series:

$$g(x) = g(x^1, \dots, x^r) = e + \sum_{i=1}^r x^i \left. \frac{\partial g(x)}{\partial x^i} \right|_{\{x^i=0\}} + O(x^2). \quad (12.3.1)$$

Let  $|x^i|$  be sufficiently close to zero, assuming their magnitudes are of order  $\varepsilon$ . Then we obtain

$$g(x) = g(x^1, \dots, x^r) = e + \varepsilon A_x. \quad (12.3.2)$$

The group elements in the Lie group  $G$  that are sufficiently close to the identity element  $e$  form a set  $O(e)$ . Below, we use equation (12.3.2) to verify that  $O(e)$ , together with the ordinary multiplication, satisfies the four group axioms, making  $O(e)$  a group known as a **local Lie group** or **infinitesimal group**.

**1)Closure:** Suppose when  $a, b \in O(0)$ ,  $g(a) \in O(e)$  and  $g(b) \in O(e)$ . Then

$$\begin{aligned} g(a)g(b) &= (e + \varepsilon A_a)(e + \varepsilon A_b) \\ &= [e + \varepsilon(A_a + A_b) + O(\varepsilon^2)] \in O(e). \end{aligned}$$

**2)Associativity:** Suppose when  $c \in O(0)$ ,  $g(c) \in O(e)$ . Then

$$\begin{aligned} g(a)[g(b)g(c)] &= (e + \varepsilon A_a)[(e + \varepsilon A_b)(e + \varepsilon A_c)] \\ &= (e + \varepsilon A_a)[e + \varepsilon(A_b + A_c) + O(\varepsilon^2)] \\ &= [e + \varepsilon(A_a + A_b + A_c) + O(\varepsilon^2)] \in O(e), \\ [g(a)g(b)]g(c) &= [(e + \varepsilon A_a)(e + \varepsilon A_b)](e + \varepsilon A_c) \\ &= [e + \varepsilon(A_a + A_b) + O(\varepsilon^2)](e + \varepsilon A_c) \\ &= [e + \varepsilon(A_a + A_b + A_c) + O(\varepsilon^2)] \in O(e). \end{aligned}$$

Comparing both sides, we find

$$g(a)[g(b)g(c)] = [g(a)g(b)]g(c),$$

thus associativity holds up to the considered order.

**3)Existence of the identity element  $e$ :**  $g(0) = e$ .

**4)Existence of inverse elements:** For any element  $g(a) \in O(e)$ , there exists an inverse element

$$g^{-1}(a) = (e - \varepsilon A_a) \in O(e),$$

satisfying

$$\begin{aligned} g(a)g^{-1}(a) &= (e + \varepsilon A_a)(e - \varepsilon A_a) = [e + O(\varepsilon^2)] \in O(e), \\ g^{-1}(a)g(a) &= (e - \varepsilon A_a)(e + \varepsilon A_a) = [e + O(\varepsilon^2)] \in O(e), \\ g(a)g^{-1}(a) &= g^{-1}(a)g(a). \end{aligned}$$

Therefore,  $O(e)$  is a group.

How large is a local Lie group? Since the magnitude  $\varepsilon$  can be arbitrarily chosen, the size (or extent) of a local Lie group is not fixed. In fact, the entire Lie group can also be regarded as a local Lie group.

From equation (12.3.1), let

$$X_i = \left. \frac{\partial g(x)}{\partial x^i} \right|_{\{x^i=0\}}, \quad i = 1, 2, \dots, r. \quad (12.3.3)$$

These  $X_i$  are called **infinitesimal generators**. The commutation relations among these infinitesimal generators are

$$[X_i, X_j] = \sum_{k=1}^r C_{ij}^k X_k,$$

where  $C_{ij}^k$  are the structure constants. An element (infinitesimal element) in the infinitesimal group  $O(e)$  can be expressed as

$$g(\delta a) = e + \sum_{i=1}^r \delta a^i X_i. \quad (12.3.4)$$

Assume  $G$  is a connected Lie group. From equation (12.3.4), we have

$$g(2\delta a) = e + 2 \sum_{i=1}^r \delta a^i X_i.$$

Since

$$\left( e + \sum_{i=1}^r \delta a^i X_i \right)^2 = e + 2 \sum_{i=1}^r \delta a^i X_i + \left( \sum_{i=1}^r \delta a^i X_i \right)^2 = e + 2 \sum_{i=1}^r \delta a^i X_i$$

(Omit the quadratic terms in the expression),

we obtain

$$g(2\delta a) = \left( e + \sum_{i=1}^r \delta a^i X_i \right)^2.$$

Similarly, we can derive

$$g(N\delta a) = \left( e + \sum_{i=1}^r \delta a^i X_i \right)^N.$$

Let  $\delta a_i = \frac{a_i}{N}$ . Then from the above equations, we obtain

$$g(a) = \left( e + \frac{1}{N} \sum_{i=1}^r a^i X_i \right)^N. \quad (12.3.5)$$

However, if  $N$  is infinitely large, we should take the limit  $N \rightarrow \infty$  in equation (12.3.5), yielding

$$g(a) = \lim_{N \rightarrow \infty} \left( e + \frac{1}{N} \sum_{i=1}^r a^i X_i \right)^N = \exp \left[ \sum_{i=1}^r a^i X_i \right]. \quad (12.3.6)$$

**Theorem 12.3.2** Let  $\text{Lie}(G)$  be the Lie algebra of an  $r$ -dimensional Lie group  $G$ , and let  $r \geq 1$ ,  $X_1, \dots, X_r \in \text{Lie}(G)$ . Then for sufficiently small  $|t|$ , the following holds:

$$\exp(tX_1) \cdots \exp(tX_r) = \exp \left( t \sum_{1 \leq i \leq r} X_i + \frac{t^2}{2} \sum_{1 \leq i < j \leq r} [X_i, X_j] + O(t^3) \right).$$

**Theorem 12.3.2** Let  $\text{Lie}(G)$  be the Lie algebra of an  $r$ -dimensional Lie group  $G$ , and let  $r \geq 1$ ,  $\{X_1, \dots, X_r\}$  be a basis of  $\text{Lie}(G)$ . Then there exist compatible local coordinate systems  $(U_\alpha, \varphi_\alpha)$ ,  $\alpha = 1, 2, 3$  around the identity element  $e$  of  $G$  such that:

- 1)  $U_1 = \exp \left( \sum_{i=1}^r x^i X_i \right)$ ,  $|x^1| < \varepsilon, \dots, |x^r| < \varepsilon$ ,  
 $\varphi_1 \left( \exp \left( \sum_{i=1}^r x^i X_i \right) \right) = (x^1, \dots, x^r)$ ;
- 2)  $U_2 = \exp(x^1 X_1) \cdots \exp(x^r X_r)$ ,  $|x^1| < \varepsilon, \dots, |x^r| < \varepsilon$ ,  
 $\varphi_2(\exp(x^1 X_1) \cdots \exp(x^r X_r)) = (x^1, \dots, x^r)$ ;

$$3) U_3 = \exp\left(\sum_{i=1}^k x^i X_i\right) \exp\left(\sum_{i=k+1}^r x^i X_i\right), \quad |x^1| < \varepsilon, \dots, |x^r| < \varepsilon,$$

$$\varphi_3\left(\exp\left(\sum_{i=1}^k x^i X_i\right) \exp\left(\sum_{i=k+1}^r x^i X_i\right)\right) = (x^1, \dots, x^r), \quad 1 < k < r.$$

The local coordinate systems  $(U_1, \varphi_1)$ ,  $(U_2, \varphi_2)$ , and  $(U_3, \varphi_3)$  are called the **first**, **second**, and **third kind of canonical coordinates** of the Lie group  $G$  at the identity element  $e$ , respectively.

## §12.4 Exponential Matrix

$X_i$  in equation (12.3.6) can be represented in matrix form. If  $X_i$  are matrices, then  $g(a)$  is also a matrix, called an **exponential matrix**.

**Definition 12.4.1** Let  $A$  be an  $n \times n$  matrix, and denote the  $n \times n$  identity matrix as  $E$ . Define the exponential function with matrix  $A$  as the variable:

$$e^A = E + A + \frac{1}{2!}A^2 + \dots + \frac{1}{n!}A^n + \dots \quad (12.4.1)$$

The matrix  $e^A$  is called the **exponential matrix**.

The exponential matrix possesses the following properties:

1) If the elements  $a_{ij}$  of matrix  $A$  satisfy  $|a_{ij}| \leq a < \infty$ , then the series (12.4.1) converges.

2) If matrices  $A$  and  $B$  commute, i.e.,  $[A, B] = 0$ , then

$$e^{A+B} = e^A e^B.$$

3) Suppose  $S$  is a nonsingular  $n \times n$  matrix. Then

$$e^{SAS^{-1}} = S e^A S^{-1}.$$

4) Suppose the eigenvalues of  $A$  are  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Then the eigenvalues of  $e^A$  are  $e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}$ .

5)  $\det e^A = e^{\text{tr}A}$ , and thus

$$|g(a)| = \det\left[\exp\left(\sum_{i=1}^r a^i X_i\right)\right] = \exp\left[\sum_{i=1}^r a^i \text{tr}X_i\right].$$

6)  $\ln(\det A) = \text{tr}(\ln A)$ .

Since  $\det e^A = e^{\text{tr}A}$ , we have

$$\det e^{\ln A} = \det A = e^{\text{tr} \ln A},$$

taking the logarithm on both sides yields

$$\ln(\det A) = \text{tr}(\ln A). \quad (12.4.2)$$

7)  $(e^A)^{-1} = e^{-A}$ .

8)  $(e^A)^T = e^{A^T}$  (where  $T$  denotes matrix transpose).

9)  $(e^A)^+ = e^{A^+}$  (where  $+$  denotes the conjugate transpose of the matrix).

10)  $\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A$ .

## §12.5 Homomorphisms and Subgroups of Lie Groups

### 1. Homomorphisms and Isomorphisms of Lie Groups

**Definition 12.5.1** Let  $G$  and  $G'$  be two groups. If there exists a map  $f: G \rightarrow G'$  that

preserves the group multiplication, i.e.,

$$f(g_i g_j) = f(g_i) f(g_j), \quad \forall g_i, g_j \in G,$$

then  $f$  is called a **homomorphism** from  $G$  to  $G'$ . If  $f$  is also bijective, it is called an **isomorphism** from  $G$  to  $G'$ .

If  $f$  is a homomorphism, then

$$f(eg) = f(g) = f(e)f(g),$$

thus

$$f(e) = e'.$$

From

$$f(g^{-1}g) = f(e) = e' = f(g^{-1})f(g),$$

it follows that

$$f(g^{-1}) = f(g)^{-1}.$$

An isomorphism is an equivalence relation, but a homomorphism is not; homomorphisms only possess reflexivity and transitivity.

**Definition 12.5.2** If  $f: G \rightarrow H$  is both a smooth map from the Lie group  $G$  to the Lie group  $H$  and a group homomorphism from  $G$  to  $H$ , then  $f$  is called a **homomorphism of Lie groups** from  $G$  to  $H$ .

**Definition 12.5.3** If a homomorphism of Lie groups  $f: G \rightarrow H$  is also a diffeomorphism, then  $f$  is called an **isomorphism of Lie groups** from  $G$  to  $H$ . If  $G = H$ , then  $f$  is called an **automorphism** of the Lie group  $G$ .

## 2. Homomorphisms and Isomorphisms of Lie Algebras

**Definition 12.5.4** Let  $\varphi: \text{Lie}(G) \rightarrow \text{Lie}(H)$  be a linear map from the Lie algebra  $\text{Lie}(G)$  to the Lie algebra  $\text{Lie}(H)$ . If  $\varphi$  preserves the Poisson bracket, i.e.,

$$\varphi([X, Y]) = [\varphi(X), \varphi(Y)], \quad \forall X, Y \in \text{Lie}(G),$$

then  $\varphi$  is called a **homomorphism** from  $\text{Lie}(G)$  to  $\text{Lie}(H)$ . If  $\varphi$  is bijective, it is called an **isomorphism** from  $\text{Lie}(G)$  to  $\text{Lie}(H)$ . If  $\text{Lie}(G) = \text{Lie}(H)$ , then  $\varphi$  is called an **automorphism** of the Lie algebra  $\text{Lie}(G)$ .

**Theorem 12.5.1** If  $f: G \rightarrow H$  is a homomorphism of Lie groups, then  $f$  induces a homomorphism  $f_*: \text{Lie}(G) \rightarrow \text{Lie}(H)$  between their Lie algebras. If  $f$  is an isomorphism of Lie groups, then  $f_*$  is an isomorphism of their Lie algebras.

**Theorem 12.5.2** Let  $G$  and  $H$  be two connected Lie groups with Lie algebras  $\text{Lie}(G)$  and  $\text{Lie}(H)$ , respectively, and let  $\psi: \text{Lie}(G) \rightarrow \text{Lie}(H)$  be a homomorphism. If  $G$  is simply connected, then there exists a unique Lie group homomorphism  $f: G \rightarrow H$  such that  $f_* = \psi$ . If  $\psi$  is an isomorphism and both  $G$  and  $H$  are simply connected, then the Lie groups  $G$  and  $H$  are isomorphic.

**Theorem 12.5.3** Every Lie algebra is the Lie algebra of one and only one simply connected Lie group.

## 3. Lie Subgroups

**Theorem 12.5.4** Let  $G$  be a topological group, and let  $H$  be an open subgroup of  $G$ . Then  $H$  is also a closed subgroup of  $G$ .

An open subgroup, as a set, is an open subset of the group  $G$ . A closed subgroup, as a set, is a closed subset of the group  $G$ .

**Definition 12.5.5** Let  $G$  be a Lie group. If  $H$  is both a subgroup of  $G$  and an embedded submanifold of  $G$ , then  $H$  is called a **Lie subgroup** of  $G$ .

**Theorem 12.5.5** Let  $G$  be a Lie group, and let  $\text{Lie}(G)$  be its Lie algebra. If  $\text{lie}(G)$  is any Lie subalgebra of  $\text{Lie}(G)$ , then there exists a unique connected Lie subgroup  $H$  of  $G$  whose Lie algebra is  $\text{lie}(G)$ .

**Theorem 12.5.6** Let  $G$  be a Lie group, and let  $H$  be a closed subgroup of  $G$ . Then there exists a unique smooth structure on the closed subgroup  $H$  such that  $H$  becomes a Lie group, and  $i: H \rightarrow G$  is a Lie subgroup of  $G$ . Furthermore,  $(i, H)$  is an embedded submanifold of  $G$ .

#### 4. One-Parameter Subgroups and the Exponential Map

**Definition 12.5.6** Let  $G$  be a Lie group. A **one-parameter subgroup** of  $G$  is a homomorphism  $\sigma: R \rightarrow G$ , where  $R$  is the one-dimensional Lie group of all real numbers under addition.

**Definition 12.5.7** Let  $G$  be a Lie group. For any  $X \in T_e G$ , denote by  $\sigma_X: R \rightarrow G$  the one-parameter subgroup of  $G$  determined by  $X$ . Define the map  $\exp: T_e G \rightarrow G$  such that for any  $X \in T_e G$ ,

$$\exp(X) = \sigma_X(1).$$

This map is called the **exponential map** of the Lie group  $G$ .

The exponential map  $\exp: T_e G \rightarrow G$  of a Lie group  $G$  possesses the following properties:

**Theorem 12.5.7** Let  $G$  be a Lie group, and let  $\text{Lie}(G)$  be its Lie algebra.

1) The exponential map is a smooth map from  $\text{Lie}(G)$  to  $G$ , and

$$(\exp)_{*0} = id: T_e G \rightarrow T_e G.$$

2) For any  $X \in \text{Lie}(G)$  and  $s, t \in R$ ,  $\exp(s+t)X = \exp s X \exp t X$ .

3) For any  $X \in \text{Lie}(G)$ ,  $(\exp X)^{-1} = \exp(-X)$ .

4) For any  $X \in \text{Lie}(G)$  and  $n \in Z$ ,  $(\exp X)^n = \exp(nX)$ .

5) The exponential map is a diffeomorphism from a neighborhood of the zero vector in  $\text{Lie}(G)$  to a neighborhood of the identity element  $e$  in  $G$ .

Let  $X$  be a left-invariant tangent vector field on the Lie group  $G$ , generated by left-translating a tangent vector  $X_e \in T_e G$ . Then the integral curve of  $X$  passing through point  $e$  has a unique tangent vector at point  $e$ , which is precisely  $X_e$ . Let the parameter of the integral curve of  $X$  be  $t$ , with  $t=0$  corresponding to the identity element  $e$ . Then the points of  $G$  on this curve can be expressed as

$$g_{X_e}(t) = \exp(tX)|_e. \quad (12.5.1)$$

From the properties of the exponential map, we have

$$g_{X_e}(t_2)g_{X_e}(t_1) = \exp(t_2X)\exp(t_1X)|_e = \exp[(t_1+t_2)X]|_e = g_{X_e}(t_1+t_2).$$

It can be verified that the points on the integral curve of  $X$  form a one-parameter subgroup of the Lie group  $G$ , which is an **abelian group** because

$$g_{X_e}(t_1)g_{X_e}(t_2) = \exp(t_1X)\exp(t_2X)|_e = \exp[(t_1+t_2)X]|_e = g_{X_e}(t_2)g_{X_e}(t_1).$$

An **abelian group** is a group whose multiplication satisfies commutativity. If all structure constants of the Lie group are zero, then the Lie group is abelian. This conclusion can be derived from Theorem 6.3.2.

Since every tangent vector in  $T_e G$  can generate a left-invariant vector field via left translation, each tangent vector in  $T_e G$  corresponds to a one-parameter subgroup, and their integral curves are smooth curves in  $G$  passing through the identity element  $e$ .

Similarly, every right-invariant vector field also corresponds to a one-parameter subgroup.

**Theorem 12.5.8** The integral curves of a right-invariant vector field and a left-invariant vector field on a Lie group  $G$  passing through the identity element  $e$  coincide. However, their integral curves passing through other elements generally do not coincide unless the group is abelian.

**Theorem 12.5.9** Any right-invariant vector field  $X$  on a Lie group  $G$  determines a one-parameter subgroup  $a_t$  of  $G$ . The one-parameter transformation group  $\varphi_t$  on  $G$  determined by the right-invariant vector field  $X$  is exactly the left translation induced by  $a_t$  on  $G$ , i.e., for any  $x \in G$ ,  $\varphi_t(x) = a_t \cdot x$ .

**Theorem 12.5.10** Any left-invariant vector field  $X$  on a Lie group  $G$  determines a

one-parameter subgroup  $a_t$  of  $G$ . The one-parameter transformation group  $\varphi_t$  on  $G$  determined by the left-invariant vector field  $X$  is exactly the right translation induced by  $a_t$  on  $G$ , i.e., for any  $x \in G$ ,  $\varphi_t(x) = x \cdot a_t$ .

**Theorem 12.5.11** If a Lie group  $G$  is abelian, then its Lie algebra  $\text{Lie}(G)$  is also abelian.

**Theorem 12.5.12** Let  $G$  be a Lie group. Then:

1) The image of a one-parameter subgroup of  $G$  is precisely a connected Lie subgroup of  $G$  with dimension less than or equal to 1.

2) The image of each one-parameter subgroup, as a Lie group, is isomorphic to one of the following: the real line  $R$ , the circle  $S^1$ , or the trivial group  $\{e\}$ .

### 5. Sophus Lie's Fundamental Theorems

**Theorem 12.5.13** The set of left-invariant tangent vector fields on a local Lie group forms a finite-dimensional Lie algebra under the Lie bracket operation. Moreover, if two Lie algebras are isomorphic, then their corresponding local Lie groups are locally isomorphic.

**Theorem 12.5.14** Given a finite-dimensional abstract Lie algebra  $\text{Lie}(G)$ , then there exists a local Lie group such that the Lie algebra formed by its left-invariant tangent vector fields is isomorphic to the abstract Lie algebra  $\text{Lie}(G)$ .

These two theorems are collectively referred to as **Sophus Lie's fundamental theorems**.

## §12.6 Lie Transformation Groups

### 1. General Lie Transformation Groups

**Definition 12.6.1** Let  $M$  be an  $m$ -dimensional smooth manifold,  $G$  an  $r$ -dimensional Lie group,  $g \in G, x \in M$ , and  $\sigma: G \times M \rightarrow M$  a smooth map, denoted as

$$\sigma(g, x) = g \cdot x, \quad (g, x) \in G \times M,$$

such that:

1) If  $e$  is the identity element of the Lie group  $G$ , then for any  $x \in M$ ,

$$e \cdot x = x;$$

2) If  $g_1, g_2 \in G$ , then for any  $x \in M$ ,

$$g_1 \cdot (g_2 \cdot x) = (g_1 \cdot g_2) \cdot x.$$

Then the Lie group  $G$  is said to be a **Lie transformation group acting on the left** on the smooth manifold  $M$ .

We can also define a **right action** of the Lie group  $G$  on the smooth manifold  $M$  by simply modifying the direction of the group element action in the above definition to a right action.

**Definition 12.6.2** Let  $M$  be an  $m$ -dimensional smooth manifold,  $G$  an  $r$ -dimensional Lie group,  $g \in G, x \in M$ , and  $\theta: M \times G \rightarrow M$  a smooth map, denoted as

$$\theta(x, g) = x \cdot g, \quad (x, g) \in M \times G,$$

such that:

1) If  $e$  is the identity element of the Lie group  $G$ , then for any  $x \in M$ ,

$$x \cdot e = x;$$

2) If  $g_1, g_2 \in G$ , then for any  $x \in M$ ,

$$(x \cdot g_1) \cdot g_2 = x \cdot (g_1 \cdot g_2).$$

Then the Lie group  $G$  is said to be a **Lie transformation group acting on the right** on the smooth manifold  $M$ .

The left action and right action of a Lie group  $G$  on a smooth manifold  $M$  are interchangeable. Suppose we already have a right action  $\theta(x, g) = x \cdot g$  of  $G$  on  $M$ . Then we can define the left action of the Lie group  $G$  on the smooth manifold  $M$  as  $\sigma(g, x) = \theta(x, g^{-1}) = x \cdot g^{-1}$ .

A one-parameter differentiable transformation group is a one-dimensional Lie group, i.e.,

$G = R$ . It can be regarded as acting on the smooth manifold  $M$  either from the left or from the right. A Lie group  $G$  is itself a smooth manifold and also a Lie transformation group. Therefore,  $G$  can act on itself both as left translations and as right translations, i.e.,  $\varphi: G \times G \rightarrow G$ .

**Definition 12.6.3** If for every non-identity element  $g \in G$  acting on the smooth manifold  $M$ , there exists at least one point  $x \in M$  such that  $g \cdot x \neq x$ , then the action of  $G$  on  $M$  is said to be **effective**.

According to this definition, if the action of the Lie group  $G$  on the smooth manifold  $M$  is effective, then for any non-identity element  $g \neq e$ ,  $L_g: M \rightarrow M$  is not the identity mapping.

**Definition 12.6.4** If for every non-identity element  $g \in G$  acting on the smooth manifold  $M$ , it holds that  $g \cdot x \neq x$  for all  $x \in M$ , then the action of  $G$  on  $M$  is said to be **free of fixed points**, or simply **free**.

Take an element  $g \in G$  and fix it. Then act it on all elements of the smooth manifold  $M$ , i.e., define

$$L_g(x) = g \cdot x, \quad x \in M,$$

Then  $L_g: M \rightarrow M$  is a smooth map. This action  $L_g$  transforms the smooth manifold  $M$  using the element  $g$  from the Lie group. Since the element  $g$  has an inverse, this transformation also has an inverse  $L_g^{-1} = L_{g^{-1}}$ :

$$L_g^{-1}(x) = L_{g^{-1}} \circ L_g(x) = g^{-1}(g \cdot x) = x,$$

and thus  $L_g^{-1}: M \rightarrow M$  is also a smooth map. Since  $L_g: M \rightarrow M$  is a homeomorphism,  $L_g$  is a diffeomorphism from  $M$  to itself. If we take all elements in the Lie group  $G$ , the set  $\{L_g, g \in G\}$  forms a subgroup of the diffeomorphism group of  $M$ . This set contains the inverse element  $L_g^{-1}$  and the identity element. The identity element is some element  $a \in G$  such that  $L_a: M \rightarrow M$  is the identity mapping  $L_a$ . Suppose  $h \in G$ , then

$$L_h \circ L_g(x) = h(g \cdot x), \quad x \in M,$$

which defines the multiplication operation in the set. Clearly, it satisfies associativity. Therefore,  $\{L_g, g \in G\}$  is a group. Moreover, if the action of  $G$  on  $M$  is effective, then  $G$  is isomorphic to the subgroup  $\{L_g, g \in G\}$  of the diffeomorphism group of  $M$ .

Suppose a Lie group  $G$  acts on a smooth manifold  $M$  on the right, denoted as  $\theta: M \times G \rightarrow M$ , or simply  $(p, g) \mapsto p \cdot g$ . Let  $X$  be an element of the Lie algebra  $\text{Lie}(G)$ , and let  $\exp(tX)$  be the one-parameter subgroup determined by  $X$ . Suppose a point  $p \in M$ . If  $\exp(tX)$  acts on  $M$  on the right, we denote this as

$$(t, p) \mapsto p \cdot \exp(tX).$$

We obtain

$$\tilde{X}_p = \left. \frac{d}{dt} \right|_{t=0} (p \cdot \exp(tX)). \quad (12.6.1)$$

Thus, we obtain a mapping  $\tilde{\theta}: \text{Lie}(G) \rightarrow \chi(M)$ , defined by  $\tilde{\theta}(X) = \tilde{X}$ .

**Definition 12.6.5** The vector field  $\tilde{\theta}(X) = \tilde{X}$  is called the **fundamental vector field** or **infinitesimal generator** on the smooth manifold  $M$  determined by the tangent vector  $X \in \text{Lie}(G)$ .

**Theorem 12.6.1** Let  $G$  be a Lie group, and let  $\theta$  be a smooth right action of  $G$  on  $M$ . For any  $X \in \text{Lie}(G)$  and  $p \in M$ , the vector fields  $X$  and  $\tilde{\theta}(X)$  are  $\theta^{(p)}$ -related.

**Theorem 12.6.2** Let  $G$  be a Lie group, and let  $\theta$  be a smooth right action of  $G$  on  $M$ . Then the mapping  $\tilde{\theta}: \text{Lie}(G) \rightarrow \chi(M)$  defined by equation (12.6.1) is a **homomorphism of Lie algebras**.



Since the mapping  $\tilde{\theta}$  defined by equation (12.6.1) is a homomorphism of Lie algebras, for any  $X, Y \in \text{Lie}(G)$ , the equality  $\tilde{\theta}[X, Y] = [\tilde{\theta}(X), \tilde{\theta}(Y)]$  holds.

If  $\exp(tX)$  is a left action on the smooth manifold  $M$ , denoted as  $\exp(tX) \cdot p$ , i.e.,

$$(t, p) \mapsto \exp(tX) \cdot p.$$

We obtain

$$\tilde{X}_p = \left. \frac{d}{dt} \right|_{t=0} (\exp(tX) \cdot p). \quad (12.6.2)$$

Thus, we also obtain a mapping  $\tilde{\sigma} : \text{Lie}(G) \rightarrow \chi(M)$ , again defined by  $\tilde{\sigma}(X) = \tilde{X}$ .

**Theorem 12.6.3** Let  $G$  be a Lie group, and let  $\sigma$  be a smooth left action of  $G$  on  $M$ . Then the mapping  $\tilde{\sigma} : \text{Lie}(G) \rightarrow \chi(M)$  defined by equation (12.6.2) is an **anti-homomorphism of Lie algebras**, i.e., for any  $X, Y \in \text{Lie}(G)$ , the equality  $\tilde{\sigma}[X, Y] = -[\tilde{\sigma}(X), \tilde{\sigma}(Y)]$  holds.

**Theorem 12.6.4** If the action of an  $r$ -dimensional Lie group  $G$  on a smooth manifold  $M$  has no fixed points, then the fundamental vector fields on  $M$  fall into two categories:  $r$  of them are everywhere linearly independent, while the remaining fundamental vector fields are constant-coefficient linear combinations of these  $r$  fields.

## 2. Homogeneous Spaces

**Definition 12.6.6** Let a Lie group  $G$  act on a smooth manifold  $M$  on the left as a Lie transformation group. If for any two points  $x, y \in M$ , there exists an element  $g \in G$  such that  $y = g \cdot x$ , then the action of  $G$  on  $M$  is said to be **transitive**.

**Definition 12.6.7** If a Lie group  $G$  acts transitively on a smooth manifold  $M$  on the left, then  $M$  is called a **homogeneous space**.

**Theorem 12.6.5** Let  $M$  be an  $m$ -dimensional homogeneous space, and let the Lie transformation group  $G$  act transitively on  $M$  on the left. Fix a point  $x \in M$ , and define

$$H = \{g \in G : g \cdot x = x\}.$$

Then  $H$  is a **closed subgroup** of  $G$ , and the homogeneous space  $M$  is diffeomorphic to the smooth manifold  $G/H$  (the **quotient Lie group**).

The closed subgroup  $H$  is both a subgroup and a closed subset of  $G$ . It is called the **isotropy subgroup** of the Lie transformation group  $G$  with respect to the base point  $x \in M$ .

## 3. Orbits

**Definition 12.6.8** The set of all points that can be reached from a point  $x$  on a smooth manifold  $M$  under the left (or right) action of all elements of a Lie group  $G$  is called the **orbit**  $O_x$  of  $G$  through  $x$ :

$$O_x = Gx = \{gx \mid \forall g \in G\}, \quad O_x \subset M.$$

The isotropy subgroups at different points on the same orbit are mutually isomorphic. If two orbits intersect, they must coincide. Therefore, the set of all orbits forms a **partition** of  $M$ , denoted as  $M/G$ , which itself becomes a smooth manifold called the **orbit space**.

## 4. Inner Automorphisms

A Lie group  $G$  can also act on itself as a Lie transformation group,  $\varphi : G \times G \rightarrow G$ .  $G$  can be regarded as acting on itself either on the right or on the left.

**Definition 12.6.9** Let  $g \in G$ . Define the map  $I_g : G \rightarrow G$  by

$$I_g(h) = ghg^{-1}, \quad \forall h \in G.$$

Then  $I_g$  is called an **inner automorphism** of the Lie group  $G$ .

The map  $I_g$  is composed of the left translation by  $g$  and the right translation by  $g^{-1}$ .  $I_g$  is a diffeomorphism from  $G$  to itself. As shown in Figure 12.6.1, a point  $a$  on the solid curve is first mapped to  $ga$  under the left translation induced by  $g$ , and then mapped to  $gag^{-1}$  under the right



translation induced by  $g^{-1}$ . When  $h = e$ ,  $I_g(e) = geg^{-1} = e$ . Thus,  $I_g$  maps the identity element  $e$  to itself, and maps each curve passing through  $e$  (such as the solid curve through  $e$  in the figure) to another curve that may differ but still passes through  $e$  (such as the dashed curve through  $e$  in the figure).

The tangent map induced by  $I_g$  maps any tangent vector in  $T_e G$  (such as the solid tangent vector  $X_e$  in the figure) to another tangent vector in  $T_e G$  (such as the dashed tangent vector  $X'_e$  in the figure). Denote the pushforward map (tangent map) induced by  $I_g$  at the identity element  $e$  as  $I_{g*}$ , abbreviated as  $I_{g*}$ , or simply  $\text{Ad}_g$ .  $\text{Ad}_g$  is a map from the tangent space  $T_e G$  to itself:

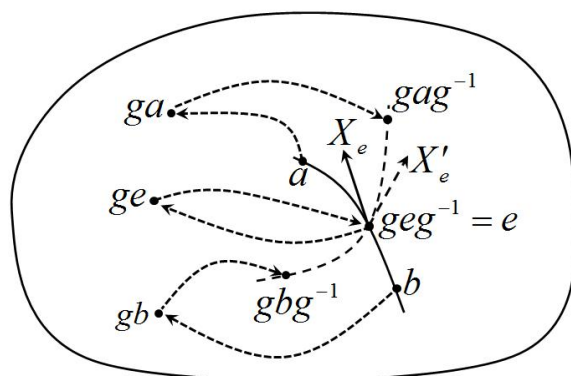


Figure 12.6.1 Inner automorphism

$$\text{Ad}_g : T_e G \rightarrow T_e G. \quad (12.6.3)$$

The map  $\text{Ad}_g$  gives an automorphism of the Lie algebra  $T_e G$ . It is a nondegenerate linear transformation acting on the linear space  $T_e G$ . Since  $T_e G$  is also the Lie algebra of the Lie group  $G$ ,  $\text{Ad}_g : \text{Lie}(G) \rightarrow \text{Lie}(G)$  is a nondegenerate linear transformation on  $\text{Lie}(G)$ .

If the solid curve is the curve of a one-parameter subgroup  $\exp(tX_e)$ , then under the map  $I_g$ , the one-parameter subgroup becomes  $\exp(tX'_e)$ , whose curve is the dashed curve through  $e$  in the figure. Therefore,  $I_g$  fixes  $e$  but generally maps a curve through  $e$  to a different curve through  $e$ , and the tangent vectors of these two curves at  $e$  are different. In general, we have the following theorem.

**Theorem 12.6.6** Let  $\text{Lie}(G)$  be the Lie algebra of the Lie group  $G$ . Then for any  $g \in G$  and  $X_e \in \text{Lie}(G)$ , we have

$$I_g \exp(tX_e) = g \exp(tX_e) g^{-1} = \exp(t \text{Ad}_g(X_e)). \quad (12.6.4)$$

## 5. Adjoint Representation

The general linear group  $GL(n; R)$  is a typical Lie group. Its elements are  $n \times n$  non-degenerate real matrices, the group multiplication is matrix multiplication, and the group identity element is the identity matrix  $E$ . The tangent space  $gl(n; R)$  of the Lie group  $GL(n; R)$  at the identity element  $E$  is an  $n^2$ -dimensional vector space  $R^{n^2}$ , whose elements are  $n \times n$  real matrices.

Since a general (abstract) Lie group  $G$  is homomorphic to the matrix group  $GL(n; R)$ , we can study a general Lie group  $G$  via the matrix group  $GL(n; R)$ . A homomorphism from an  $r$ -dimensional general Lie group  $G$  into the matrix group  $GL(n; R)$  is called an  $n$ -dimensional representation of the general Lie group  $G$ . For example, we say the matrix group  $SO(3)$  is a 3-dimensional representation of some general Lie group, and the matrix group  $SU(2)$  is a 3-dimensional representation of another general Lie group. The same general Lie group can have infinitely many representations. However, if the elements of the general Lie group correspond one-to-one with the elements of the matrix group—that is, the general Lie group is isomorphic to the matrix group—then the representation is called a **faithful representation**. For instance, we say the matrix group  $SO(3)$  is a faithful representation of a certain general Lie group, and the matrix group  $SU(2)$  is a faithful representation of another general Lie group. Precisely because it is faithful, we often identify the matrix group  $SO(3)$  with the general Lie group it faithfully represents; that is, unless otherwise specified, when we speak of the group  $SO(3)$ , we usually refer both to the matrix group  $SO(3)$  and to the general Lie group it faithfully represents. The same

holds for  $SU(2)$ . Furthermore, every  $r$ -dimensional general Lie group  $G$  possesses a natural  $r$ -dimensional representation, called the **adjoint representation**.

The mapping  $\text{Ad}_g$  given by equation (12.6.3) is a non-degenerate linear transformation acting on the vector space  $T_e G$ . Every non-degenerate linear transformation corresponds to an invertible matrix. All such invertible matrices form a group, denoted  $GL(r; R; T_e G)$ . Since each  $g \in G$  corresponds to a transformation  $\text{Ad}_g$ , and  $\text{Ad}_g$  corresponds to an element of  $GL(r; R; T_e G)$ , we obtain a mapping

$$\text{Ad}: G \rightarrow GL(r; R; T_e G).$$

$\text{Ad}$  is a group homomorphism because for any  $g, h \in G$ ,

$$\text{Ad}_{g \cdot h} = (I_{g \cdot h})^* = (I_g \circ I_h)^* = \text{Ad}_g \circ \text{Ad}_h.$$

**Definition 12.6.10** The Lie group homomorphism  $\text{Ad}: G \rightarrow GL(r; R; T_e G)$  given by equation (12.6.3) is called the **adjoint representation** of the  $r$ -dimensional Lie group  $G$ .

The adjoint representation of a Lie group  $G$  is essentially a representation formed by the linear transformations acting on the Lie algebra  $\text{Lie}(G)$  of  $G$ . For example, the adjoint representation group of the general Lie group  $SU(2)$  is the matrix group  $SO(3)$ , and the adjoint representation group of the general Lie group  $SO(3)$  is also the matrix group  $SO(3)$ .

**Definition 12.6.11** The homomorphism from the Lie algebra  $T_e G$  to  $gl(r; R)$  induced by the tangent map  $(\text{Ad})_*$  of the adjoint representation  $\text{Ad}: G \rightarrow GL(r; R; T_e G)$ ,

$$\text{ad} = (\text{Ad})_*: T_e G \rightarrow gl(r; R),$$

is called the **adjoint representation of the Lie algebra**  $T_e G$  of the Lie group  $G$ .

Here  $gl(r; R)$  can be regarded as the set of linear transformations from  $T_e G$  to itself. For any  $X \in T_e G$ ,  $\text{ad}(X)$  is a linear transformation acting on  $T_e G$ .

**Theorem 12.6.7** Let  $\text{Ad}: G \rightarrow GL(r; R; T_e G)$  be the adjoint representation of an  $r$ -dimensional Lie group  $G$ , and let

$$\text{ad} = (\text{Ad})_*: T_e G \rightarrow gl(r; R)$$

be the adjoint representation of the Lie algebra  $T_e G$  of the Lie group  $G$ . Then for any  $X, Y \in T_e G$ ,

$$\text{ad}(X) \cdot Y = -[X, Y].$$

$\text{ad}(X)$  has the following properties:

Let  $X, Y_1, Y_2 \in T_e G$ , let  $\beta_1, \beta_2$  be constants (real or complex). Then

- 1)  $\text{ad}(X) \cdot (\beta_1 Y_1 + \beta_2 Y_2) = \beta_1 \text{ad}(X) \cdot Y_1 + \beta_2 \text{ad}(X) \cdot Y_2$ ;
- 2)  $\text{ad}(\beta_1 X_1 + \beta_2 X_2) \cdot Y = \beta_1 \text{ad}(X_1) \cdot Y + \beta_2 \text{ad}(X_2) \cdot Y$ ;
- 3)  $\text{ad}([X_1, X_2]) \cdot Y = \text{ad}(X_1) \text{ad}(X_2) \cdot Y - \text{ad}(X_2) \text{ad}(X_1) \cdot Y$ .

## §12.7 Orientation of Lie Groups

**Theorem 12.7.1** A connected Lie group is orientable.

**Definition 12.7.1** Let  $G$  be a Lie group. If for a given orientation of  $G$ , every left translation  $L_g$  generated by an element  $g \in G$  preserves this orientation, then the orientation of  $G$  is said to be **left-invariant**.

**Theorem 12.7.2** Every connected Lie group has exactly two left-invariant orientations, corresponding respectively to the two orientations of its Lie algebra.

**Theorem 12.7.3** On an  $r$ -dimensional compact connected Lie group, there exists a nonzero  $r$ -form  $\omega$  that is **bi-invariant**.

A differential form is called bi-invariant if it is both left-invariant and right-invariant. A compact connected Lie group can be oriented using  $\omega$ . Therefore, according to Theorem 12.7.3, the two distinct orientations on a compact connected Lie group are bi-invariant, meaning they

remain unchanged under left (or right) translations.

## §12.8 Bi-invariant Metric Tensors

Since a metric tensor can be defined on a smooth manifold and a Lie group  $G$  is itself a smooth manifold, it is possible to define a metric tensor on  $G$ , making it a Riemannian manifold.

We know that a smooth manifold  $G$  always admits a positive-definite metric tensor. Therefore, we can define a metric tensor on a smooth manifold  $G$ , that is, we can define a positive definite inner product  $\langle \cdot, \cdot \rangle_p$  on the tangent space  $T_p G$  at each point  $p \in G$ , such that for any smooth vector fields  $X$  and  $Y$ ,  $\langle X, Y \rangle$  is a smooth function on  $G$ . For a Lie group  $G$ , left-invariant vector fields are invariant under left translations. Consequently, the metric tensor defined using left-invariant vector fields  $X$  and  $Y$  is also left-invariant, i.e.,

$$\langle X, Y \rangle_p = \langle (L_a)_* X, (L_a)_* Y \rangle_{ap}$$

holds for any  $a, p \in G$  and any left-invariant vector fields  $X, Y \in T_p G$ .

**Theorem 12.8.1** There is a one-to-one correspondence between left-invariant metric tensors  $\langle \cdot, \cdot \rangle$  on a Lie group  $G$  and inner products on the tangent space  $T_e G$  at the identity element  $e$ .

Similarly, we can also define right-invariant metric tensors on a Lie group  $G$ . A metric tensor that is both left-invariant and right-invariant is called a **bi-invariant metric tensor**.

**Definition 12.8.1** Let  $\langle X, Y \rangle$  be a metric tensor on a Lie group  $G$ . If

$$\langle X, Y \rangle_p = \langle (L_a)_* X, (L_a)_* Y \rangle_{ap},$$

$$\langle X, Y \rangle_p = \langle (R_b)_* X, (R_b)_* Y \rangle_{pb},$$

holds for any  $a, b, p \in G$ ,  $X, Y \in T_p G$ , then  $\langle X, Y \rangle$  is called a **bi-invariant metric tensor** on  $G$ .

**Theorem 12.8.2** Every compact connected Lie group admits a bi-invariant metric tensor.

**Theorem 12.8.3** The left-invariant metric tensor  $\langle \cdot, \cdot \rangle$  on a Lie group  $G$  is bi-invariant if and only if the inner product is invariant under the adjoint action, i.e., for all  $g \in G$ ,  $X, Y \in T_p G$ , the adjoint representation  $\text{Ad}_g X = gXg^{-1}$  is an isometry on  $T_p G$ :

$$\langle \text{Ad}_g X, \text{Ad}_g Y \rangle = \langle X, Y \rangle.$$

**Theorem 12.8.4** Let  $G$  be a Lie group with a bi-invariant metric tensor  $g$ , and let  $D$  be the Riemannian connection associated with  $g$ . Then:

1) For any left-invariant vector field  $X$  on  $G$ ,

$$D_X X = 0; \quad (12.8.1)$$

2) For any two left-invariant vector fields  $X$  and  $Y$  on  $G$ ,

$$D_X Y = \frac{1}{2}[X, Y]; \quad (12.8.2)$$

3) For any left-invariant vector fields  $X, Y, Z$  on  $G$ ,

$$R(X, Y)Z = \frac{1}{4}[Z, [X, Y]]; \quad (12.8.3)$$

4) For any left-invariant vector fields  $X, Y, Z, W$  on  $G$ ,

$$R(X, Y, Z, W) = \frac{1}{4}\langle [X, Y], [W, Z] \rangle, \quad (12.8.4)$$

where  $\langle [X, Y], [W, Z] \rangle$  denotes the inner product operation between  $[X, Y]$  and  $[W, Z]$ .

5) If  $X$  and  $Y$  are mutually orthogonal left-invariant unit vector fields, then the sectional curvature of the two-dimensional plane determined by  $X$  and  $Y$  is

$$K(X, Y) = \frac{1}{4}\| [X, Y] \|^2 \geq 0, \quad (12.8.5)$$

with equality if and only if  $[X, Y] = 0$ .

**Theorem 12.8.5** Every compact connected Lie group is a Riemannian symmetric space.

## §12.9 $SO(3)$ Group

### 1. $SO(3)$ Group as the Special Orthogonal Group

The  $SO(3)$  group is a special orthogonal group. Let the matrix  $R$  be an element of  $SO(3)$  group. Then the transpose of  $R$  (The transpose operation is denoted by  $T$ ) is equal to its inverse, i.e.,

$$R^T = R^{-1},$$

and the determinant of  $R$  is equal to 1, i.e.,

$$|R| = 1.$$

### 2. $SO(3)$ Group as the Real Rotation Group

The  $SO(3)$  group is a real three-dimensional pure rotation group with three independent parameters. There are multiple ways to choose parameters for  $SO(3)$ . Consider a three-dimensional Cartesian coordinate system (orthonormal basis)  $O$ -xyz. First rotate by an angle  $\theta_1$  around the  $x$ -axis; the rotation matrix is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix}, \quad (12.9.1)$$

Next rotate by an angle  $\theta_2$  around the  $y$ -axis; the rotation matrix is

$$\begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{pmatrix}, \quad (12.9.2)$$

Finally rotate by an angle  $\theta_3$  around the  $z$ -axis; the rotation matrix is

$$\begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (12.9.3)$$

The composite transformation matrix of these three rotations is

$$\begin{aligned} R &= g(\theta_1, \theta_2, \theta_3) \\ &= \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta_3 \cos \theta_2 & -\sin \theta_3 \cos \theta_1 & \sin \theta_1 \sin \theta_3 \\ & + \sin \theta_1 \cos \theta_3 \sin \theta_2 & + \cos \theta_1 \cos \theta_3 \sin \theta_2 \\ \cos \theta_2 \sin \theta_3 & \cos \theta_3 \cos \theta_1 & -\sin \theta_1 \cos \theta_3 \\ & + \sin \theta_1 \sin \theta_2 \sin \theta_3 & + \cos \theta_1 \sin \theta_2 \sin \theta_3 \\ -\sin \theta_2 & \sin \theta_1 \cos \theta_2 & \cos \theta_1 \cos \theta_2 \end{pmatrix}. \end{aligned} \quad (12.9.4)$$

Now we compute the infinitesimal generators of the  $SO(3)$  group:

$$X_1 = \left. \frac{\partial g}{\partial \theta_1} \right|_{\{\theta_i=0\}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (12.9.5)$$

$$X_2 = \left. \frac{\partial g}{\partial \theta_2} \right|_{\{\theta_i=0\}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad (12.9.6)$$

$$X_3 = \left. \frac{\partial g}{\partial \theta_3} \right|_{\{\theta_i=0\}} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (12.9.7)$$

Therefore, from equation (12.3.6) we obtain

$$R = g(\theta_1, \theta_2, \theta_3) = \exp[\theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3]. \quad (12.9.8)$$

### 3. Euler Angle Parameterization

The  $SO(3)$  group is a rotation group, so it can be parameterized by Euler angles.

As shown in Figure 12.9.1, let the original coordinate system be  $(x, y, z)$ . Rotate the original coordinate system by an angle  $\gamma$  about the  $z$ -axis to obtain a new coordinate system  $(x', y', z')$ . The general transformation relation between the old and new coordinate systems is

$$\begin{cases} x = a_{11}x' + a_{12}y' + a_{13}z', \\ y = a_{21}x' + a_{22}y' + a_{23}z', \\ z = a_{31}x' + a_{32}y' + a_{33}z'. \end{cases} \quad (12.9.9)$$

Since the  $z$ -axis remains unchanged,  $z'=z$ , which gives  $a_{31}=a_{32}=0$  and  $a_{33}=1$ . Moreover, because only the  $xy$ -plane rotates, the changes in  $x$  and  $y$  are independent of  $z'$ , so  $a_{13}=a_{23}=0$ . Thus, equation (12.9.9) simplifies to

$$\begin{cases} x = a_{11}x' + a_{12}y', \\ y = a_{21}x' + a_{22}y', \\ z = z'. \end{cases} \quad (12.9.10)$$

The first and second equations in (12.9.10) are essentially the familiar rotation formulas in the  $xy$ -plane. Therefore, we obtain

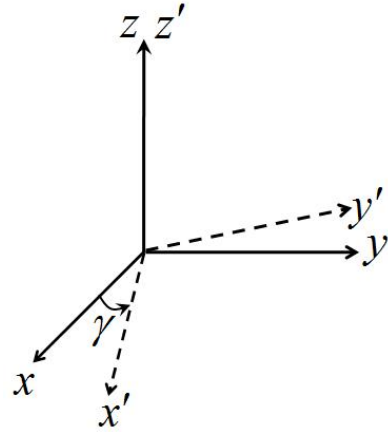
$$\begin{cases} x = x' \cos \gamma - y' \sin \gamma, \\ y = x' \sin \gamma + y' \cos \gamma, \\ z = z'. \end{cases} \quad (12.9.11)$$

In equation (12.9.11), the positive direction of angle  $\gamma$  is determined by the right-hand rule. Denote the transformation matrix of (12.9.11) as

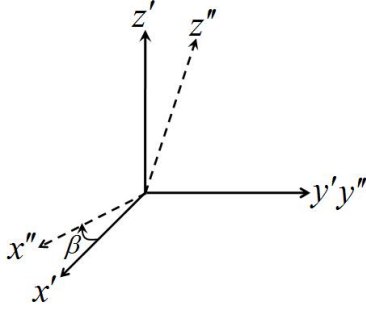
$$C = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (12.9.12)$$

Next, rotate the coordinate system  $(x', y', z')$  by an angle  $\beta$  about the  $y'$ -axis, as shown in Figure 12.9.2, to obtain a new coordinate system  $(x'', y'', z'')$ . The general transformation relation is

$$\begin{cases} x' = a_{11}x'' + a_{12}y'' + a_{13}z'', \\ y' = a_{21}x'' + a_{22}y'' + a_{23}z'', \\ z' = a_{31}x'' + a_{32}y'' + a_{33}z''. \end{cases} \quad (12.9.13)$$



**Figure 12.9.1** Rotation by angle  $\gamma$  about the  $z$ -axis



**Figure 12.9.2** Rotation by angle  $\beta$  about the  $y'$ -axis

Since the  $y'$ -axis remains unchanged,  $y''=y'$ , which gives  $a_{21}=a_{23}=0$ ,  $a_{22}=1$ . Moreover, because only the  $x'z'$ -plane rotates, the changes in  $x', z'$  are independent of  $y''$ , so  $a_{12}=a_{32}=0$ . Thus, equation (12.9.13) simplifies to

$$\begin{cases} x' = a_{11}x'' + a_{13}z'', \\ y' = y'', \\ z' = a_{31}x'' + a_{33}z''. \end{cases} \quad (12.9.14)$$

The second and third equations in (12.9.14) are essentially the rotation formulas in the  $x'z'$ -plane. Therefore, we obtain

$$\begin{cases} x' = x'' \cos(-\beta) - z'' \sin(-\beta), \\ y' = y'', \\ z' = x'' \sin(-\beta) + z'' \cos(-\beta). \end{cases} \quad (12.9.15)$$

Since the rotation direction in the  $x'z'$ -plane is opposite to that determined by the right-hand rule, a negative sign is placed before  $\beta$ . Simplifying (12.9.15) yields

$$\begin{cases} x' = x'' \cos \beta + z'' \sin \beta, \\ y' = y'', \\ z' = -x'' \sin \beta + z'' \cos \beta. \end{cases} \quad (12.9.16)$$

Denote the transformation matrix of (12.9.16) as

$$B = \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix}. \quad (12.9.17)$$

Finally, rotate the coordinate system  $(x'', y'', z'')$  by an angle  $\alpha$  about the  $z''$ -axis to obtain the new coordinate system  $(x''', y''', z''')$ , as shown in Figure 12.9.3. The general transformation relation is

$$\begin{cases} x'' = a_{11}x''' + a_{12}y''' + a_{13}z''', \\ y'' = a_{21}x''' + a_{22}y''' + a_{23}z''', \\ z'' = a_{31}x''' + a_{32}y''' + a_{33}z'''. \end{cases} \quad (12.9.18)$$

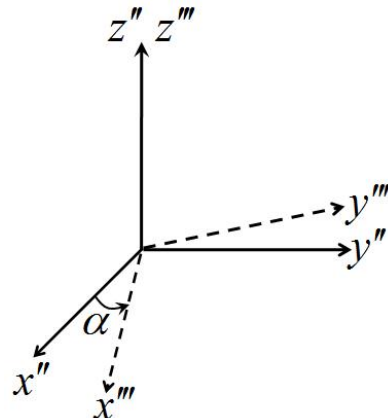
Since the  $z''$ -axis remains unchanged,  $z'''=z''$ , which gives  $a_{31}=a_{32}=0$  and  $a_{33}=1$ . Moreover, because only the  $x''y''$ -plane rotates, the changes in  $x''$  and  $y''$  are independent of  $z''$ , so  $a_{13}=a_{23}=0$ . Thus, equation (12.9.18) simplifies to

$$\begin{cases} x'' = a_{11}x''' + a_{12}y''', \\ y'' = a_{21}x''' + a_{22}y''', \\ z'' = z'''. \end{cases} \quad (12.9.19)$$

The first and second equations in (12.9.19) are the familiar rotation formulas in the  $x''y''$ -plane. Therefore, we obtain

$$\begin{cases} x'' = x''' \cos \alpha - y''' \sin \alpha, \\ y'' = x''' \sin \alpha + y''' \cos \alpha, \\ z'' = z'''. \end{cases} \quad (12.9.20)$$

In equation (12.9.20), the positive direction of angle  $\alpha$  is determined by the right-hand rule. Denote the transformation matrix of (12.9.20) as



**Figure 12.9.3** Rotation by angle  $\alpha$  about the  $z''$ -axis

$$A = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (12.9.21)$$

Combining the three independent transformations into one, the transformation matrix  $R(\alpha, \beta, \gamma)$  of the composite transformation is the product of these three matrices  $A, B, C$ , i.e.,

$$R(\alpha, \beta, \gamma) = ABC = \begin{pmatrix} \cos \alpha \cos \beta \cos \gamma - \sin \alpha \sin \gamma & -\cos \alpha \cos \beta \sin \gamma - \sin \alpha \cos \gamma & \cos \alpha \sin \beta \\ \sin \alpha \cos \beta \cos \gamma + \cos \alpha \sin \gamma & -\sin \alpha \cos \beta \sin \gamma + \cos \alpha \cos \gamma & \sin \alpha \sin \beta \\ -\sin \beta \cos \gamma & \sin \beta \sin \gamma & \cos \beta \end{pmatrix}. \quad (12.9.22)$$

This composite transformation matrix is the rotation matrix in three-dimensional real space, i.e., the  $SO(3)$  group.

## §12.10 $SU(2)$ Group

### 1. General Form of the $SU(2)$ Group

The  $SU(2)$  group is a three-dimensional matrix group whose complex conjugate transpose (also called Hermitian conjugate, denoted by  $+$ ) equals its inverse, and whose determinant is 1. Let the matrix  $U$  be an element of  $SU(2)$ ; then

$$U^+ = U^{-1}, \quad |U| = 1.$$

The  $SU(2)$  group can be written as

$$SU(2) = \{U \mid U \in GL(2, C), U^+ = U^{-1}, |U| = 1\}.$$

Assume the matrix form of an  $SU(2)$  element is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (12.10.1)$$

where  $a, b, c, d$  are complex numbers. Because  $|U| = 1$ , we have

$$ad - bc = 1, \quad (12.10.2)$$

and because ( $*$  denotes taking the complex conjugate of the number)

$$U^+ = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix}, \quad U^+ = U^{-1},$$

the condition  $U^+U = E$  (where  $E$  is the identity matrix) gives

$$\begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

That is,

$$\begin{pmatrix} aa^* + cc^* & ab^* + cd^* \\ ab^* + cd^* & bb^* + dd^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The following relations are obtained:

$$aa^* + cc^* = 1, \quad (12.10.3)$$

$$bb^* + dd^* = 1, \quad (12.10.4)$$

$$a^*b + c^*d = 0, \quad (12.10.5)$$

$$ab^* + cd^* = 0. \quad (12.10.6)$$

From equation (12.10.4), we obtain  $b = \frac{1 - dd^*}{b^*}$ . From equation (12.10.6), we obtain  $c = -\frac{ab^*}{d^*}$ .

So  $bc = \frac{(dd^* - 1)a}{d^*}$ . Substituting the latter into equation (12.10.2) gives  $a = d^*$ , which implies

$d = a^*$ . Consequently, substituting  $a = d^*$  into  $c = -\frac{ab^*}{d^*}$  yields  $c = -b^*$ . Substituting

$c = -b^*$  and  $d = a^*$  into (12.10.1), we obtain the general form of an  $SU(2)$  group element as

$$\begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \quad aa^* + bb^* = 1.$$

Here  $a$  and  $b$  are complex numbers, so there are four real parameters initially. However, the unimodular condition  $aa^* + bb^* = 1$  reduces the number of independent real parameters to three.

### 2 $SU(2)$ Group is Homeomorphic to the Three-Dimensional Unit Sphere

Let  $a = x + iy$ ,  $b = z + iw$ , then the general form of the  $SU(2)$  group becomes

$$\begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}.$$

From the unimodular condition, we have

$$(x + iy)(x - iy) + (z + iw)(z - iw) = 1.$$

Simplifying yields

$$x^2 + y^2 + z^2 + w^2 = 1. \quad (12.10.7)$$

This equation is precisely the equation of the unit sphere of radius 1 in four-dimensional Euclidean space, i.e., the 3-sphere  $S^3$ . Therefore, each element of the  $SU(2)$  group corresponds to a point on  $S^3$ , and the  $SU(2)$  group is homeomorphic to  $S^3$ .

### 3.A Real Parameterization of the $SU(2)$ Group

Introduce three independent real parameters  $\theta_1, \theta_2, \theta_3$ , such that

$$a = \left( \cos \frac{\theta_1}{2} \cos \frac{\theta_2}{2} + i \sin \frac{\theta_1}{2} \sin \frac{\theta_2}{2} \right) e^{i \frac{\theta_3}{2}},$$

$$b = \left( -\cos \frac{\theta_1}{2} \sin \frac{\theta_2}{2} + i \sin \frac{\theta_1}{2} \cos \frac{\theta_2}{2} \right) e^{-i \frac{\theta_3}{2}},$$

Then  $a$  and  $b$  satisfy the unimodular condition, and the element of the  $SU(2)$  group is

$$U = g(\theta_1, \theta_2, \theta_3)$$

$$= \begin{pmatrix} \left( \cos \frac{\theta_1}{2} \cos \frac{\theta_2}{2} + i \sin \frac{\theta_1}{2} \sin \frac{\theta_2}{2} \right) e^{i \frac{\theta_3}{2}} & \left( -\cos \frac{\theta_1}{2} \sin \frac{\theta_2}{2} + i \sin \frac{\theta_1}{2} \cos \frac{\theta_2}{2} \right) e^{-i \frac{\theta_3}{2}} \\ \left( \cos \frac{\theta_1}{2} \sin \frac{\theta_2}{2} + i \sin \frac{\theta_1}{2} \cos \frac{\theta_2}{2} \right) e^{i \frac{\theta_3}{2}} & \left( \cos \frac{\theta_1}{2} \cos \frac{\theta_2}{2} - i \sin \frac{\theta_1}{2} \sin \frac{\theta_2}{2} \right) e^{-i \frac{\theta_3}{2}} \end{pmatrix}$$

$$= \begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix} \begin{pmatrix} \cos \frac{\theta_2}{2} & -\sin \frac{\theta_2}{2} \\ \sin \frac{\theta_2}{2} & \cos \frac{\theta_2}{2} \end{pmatrix} \begin{pmatrix} e^{i \frac{\theta_3}{2}} & 0 \\ 0 & e^{-i \frac{\theta_3}{2}} \end{pmatrix}. \quad (12.10.8)$$

Now compute the infinitesimal generators of the  $SU(2)$  group:

$$X_1 = \left. \frac{\partial g}{\partial \theta_1} \right|_{\{\theta_i=0\}} = \frac{1}{2} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

$$X_2 = \left. \frac{\partial g}{\partial \theta_2} \right|_{\{\theta_i=0\}} = \frac{1}{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad (12.10.9)$$

$$X_3 = \left. \frac{\partial g}{\partial \theta_3} \right|_{\{\theta_i=0\}} = \frac{1}{2} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$



Therefore, from equation (12.3.6) we obtain

$$U = g(\theta_1, \theta_2, \theta_3) = \exp[\theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3]. \quad (12.10.10)$$

The commutation relations among  $X_1, X_2$  and  $X_3$  are

$$[X_1, X_2] = X_3, \quad [X_2, X_3] = X_1, \quad [X_3, X_1] = X_2. \quad (12.10.11)$$

If we perform the following coordinate transformation on the parameters in (12.10.8):

$$\frac{\theta_1}{2} = \theta'_1, \quad \frac{\theta_2}{2} = \theta'_2, \quad \frac{\theta_3}{2} = \theta'_3,$$

i.e.,

$$\theta'_1 = 2\theta_1, \quad \theta'_2 = 2\theta_2, \quad \theta'_3 = 2\theta_3, \quad (12.10.12)$$

then the infinitesimal generators of the  $SU(2)$  group become

$$\begin{aligned} X'_1 &= \left. \frac{\partial g}{\partial \theta'_1} \right|_{\{\theta'_i=0\}} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \\ X'_2 &= \left. \frac{\partial g}{\partial \theta'_2} \right|_{\{\theta'_i=0\}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \\ X'_3 &= \left. \frac{\partial g}{\partial \theta'_3} \right|_{\{\theta'_i=0\}} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \end{aligned} \quad (12.10.13)$$

The commutation relations among  $X'_1, X'_2$ , and  $X'_3$  are

$$[X'_1, X'_2] = 2X'_3, \quad [X'_2, X'_3] = 2X'_1, \quad [X'_3, X'_1] = 2X'_2. \quad (12.10.14)$$

The Jacobian matrix of the coordinate transformation (12.10.12) is

$$J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The transformation relations between  $X'_1, X'_2, X'_3$  and  $X_1, X_2, X_3$  are

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X'_1 \\ X'_2 \\ X'_3 \end{pmatrix}.$$

## §12.11 Homomorphism between the $SO(3)$ Group and the $SU(2)$ Group

Let the components of a position vector  $\mathbf{r}$  in three-dimensional Euclidean space be  $(x, y, z)$ .

Using the three Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and the position vector  $\mathbf{r}$ , we can construct the following matrix:

$$h = x\sigma_1^* + y\sigma_2^* + z\sigma_3^* = \mathbf{r} \cdot \boldsymbol{\sigma}^* = \begin{pmatrix} z & x+iy \\ x-iy & -z \end{pmatrix}.$$

There is a one-to-one correspondence between the matrix  $h$  and the position vector  $\mathbf{r}$ .

Using the  $SU(2)$  group to perform a similarity transformation on  $h$ ,

$$h' = UhU^{-1}, \quad U \in SU(2) \text{ i.e., } U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}. \quad (12.11.1)$$

Let

$$h' = \begin{pmatrix} z' & x'+iy' \\ x'-iy' & -z' \end{pmatrix}. \quad (12.11.2)$$

When the matrix  $h$  transforms to  $h'$ , the position vector also changes from  $\mathbf{r}$  to  $\mathbf{r}'$ . The transformation from  $h$  to  $h'$  is induced by an element  $U$  of the  $SU(2)$  group. The transformation

that takes the position vector from  $\mathbf{r}$  to  $\mathbf{r}'$  can be assumed to correspond to a matrix  $R_U$ , i.e.,

$$\mathbf{r}' = R_U \mathbf{r}, \quad (12.11.3)$$

Therefore, there exists a correspondence between the matrix  $U$  and the matrix  $R_U$ , denoted by  $f$ . That is,

$$f: U \rightarrow R_U,$$

or

$$f(U) = R_U.$$

Next, we determine the matrix  $R_U$ . From equation (12.11.1), we obtain

$$h' = \begin{pmatrix} aa^*z + ab^*(x+iy) & -abz + a^2(x+iy) \\ +a^*b(x-iy) - bb^*z & -b^2(x-iy) - abz \\ -a^*b^*z - b^{*2}(x+iy) & bb^*z - ab^*(x+iy) \\ +a^{*2}(x-iy) - a^*b^*z & -a^*b(x-iy) - aa^*z \end{pmatrix}.$$

After comparing the result with equation (12.11.2), we obtain the following relations:

$$\begin{cases} z' = ab^*(x+iy) + a^*b(x-iy) - bb^*z + aa^*z, \\ x' + iy' = a^2(x+iy) - b^2(x-iy) - 2abz, \\ x' - iy' = a^{*2}(x-iy) - b^{*2}(x+iy) - 2a^*b^*z. \end{cases}$$

Rewriting these in matrix form gives:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(a^2 + a^{*2} - b^2 - b^{*2}) & \frac{i}{2}(a^2 - a^{*2} + b^2 - b^{*2}) & -(ab + a^*b^*) \\ \frac{i}{2}(a^{*2} - a^2 + b^2 - b^{*2}) & \frac{1}{2}(a^2 + a^{*2} + b^2 + b^{*2}) & i(ab - a^*b^*) \\ (ab^* + a^*b) & i(ab^* - a^*b) & (aa^* - bb^*) \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (12.11.4)$$

Equation (12.11.4) is the coordinate transformation formula for the position vector from  $\mathbf{r}$  to  $\mathbf{r}'$ . Since

$$\det h = -(x^2 + y^2 + z^2) = -|\mathbf{r}|^2,$$

$$\det h' = -(x'^2 + y'^2 + z'^2) = -|\mathbf{r}'|^2 = \det U \det h \det U^{-1} = \det h,$$

so

$$|\mathbf{r}'|^2 = |\mathbf{r}|^2.$$

Therefore, the transformation matrix in equation (12.11.4) is a real orthogonal matrix. Let

$$R_U(a, b) = \begin{pmatrix} \frac{1}{2}(a^2 + a^{*2} - b^2 - b^{*2}) & \frac{i}{2}(a^2 - a^{*2} + b^2 - b^{*2}) & -(ab + a^*b^*) \\ \frac{i}{2}(a^{*2} - a^2 + b^2 - b^{*2}) & \frac{1}{2}(a^2 + a^{*2} + b^2 + b^{*2}) & i(ab - a^*b^*) \\ (ab^* + a^*b) & i(ab^* - a^*b) & (aa^* - bb^*) \end{pmatrix}, \quad (12.11.5)$$

then we have

$$R_U(a, b)^T = R_U(a, b)^{-1}.$$

Using the condition  $|a|^2 + |b|^2 = 1$ , we compute the determinant of  $R_U(a, b)$  to be

$$\det R_U(a, b) = 1.$$

Therefore, the matrix  $R_U(a, b)$  belongs to the pure rotation group in three-dimensional Euclidean space, i.e., the  $SO(3)$  group.

Equation (12.11.1) shows that two complex numbers  $a$  and  $b$  determine an element  $U$  of the  $SU(2)$  group, while equation (12.11.5) shows that the same two complex numbers  $a$  and  $b$  also determine an element of the  $SO(3)$  group. Hence, there exists a correspondence  $f$  between the  $SU(2)$  group and the  $SO(3)$  group, i.e.,

$$f : SU(2) \rightarrow SO(3).$$

If there is another matrix  $V \in SU(2)$  such that

$$h'' = Vh'V^{-1}, \quad \mathbf{r}' \rightarrow \mathbf{r}'', \quad (12.11.6)$$

then, analogous to equation (12.11.3), there must be another matrix  $R_V \in SO(3)$  satisfying

$$\mathbf{r}'' = R_V \mathbf{r}'. \quad (12.11.7)$$

Thus, there is also a correspondence between the matrix  $V$  and the matrix  $R_V$ :

$$f(V) = R_V.$$

From equations (12.11.1) and (12.11.6), we have

$$h'' = VUhU^{-1}V^{-1} = (VU)h(VU)^{-1}, \quad \mathbf{r} \rightarrow \mathbf{r}''.$$

Similarly, from equations (12.11.3) and (12.11.7), we obtain

$$\mathbf{r}'' = R_V \mathbf{r}' = R_V R_U \mathbf{r}, \quad \mathbf{r} \rightarrow \mathbf{r}''.$$

Therefore, the mapping satisfies

$$f(VU) = R_V R_U = f(V)f(U), \quad \forall U, V \in SU(2),$$

which shows that the mapping  $f$  is a homomorphism from the  $SU(2)$  group to the  $SO(3)$  group.

This homomorphism is also surjective, meaning that for any element of the  $SO(3)$  group, there exists a corresponding element in the  $SU(2)$  group.

Since the  $SO(3)$  group is a real three-dimensional rotation group with three independent parameters, we previously used the Euler angle parameterization. However, there are multiple ways to parameterize Euler angles. Here, to clearly see that this homomorphism is surjective, we adopt a different Euler angle parameterization: regard the rotation as first rotating by an angle  $\gamma$  about the  $z$ -axis, then rotating by an angle  $\beta$  about the new  $y'$ -axis, and finally rotating by an angle  $\alpha$  about the latest  $z''$ -axis.

Let  $a = e^{i\frac{\gamma}{2}}, b = 0$ , and substitute into equation (12.11.5). The  $SO(3)$  matrix corresponding to a rotation by  $\gamma$  about the  $z$ -axis is

$$R_z(\gamma) = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (12.11.8)$$

From equation (12.11.1), the  $SU(2)$  matrix corresponding to  $R_z(\gamma)$  is

$$U_z(\gamma) = \begin{pmatrix} e^{i\frac{\gamma}{2}} & 0 \\ 0 & e^{-i\frac{\gamma}{2}} \end{pmatrix}. \quad (12.11.9)$$

Let  $a = \cos \frac{\beta}{2}$ ,  $b = -\sin \frac{\beta}{2}$ , and substitute into equation (12.11.5). The  $SO(3)$  matrix corresponding to a rotation by  $\beta$  about the new  $y'$ -axis is

$$R_{y'}(\beta) = \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix}, \quad (12.11.10)$$

From equation (12.11.1), the  $SU(2)$  matrix corresponding to  $R_{y'}(\beta)$  is

$$U_{y'}(\beta) = \begin{pmatrix} \cos \frac{\beta}{2} & -\sin \frac{\beta}{2} \\ \sin \frac{\beta}{2} & \cos \frac{\beta}{2} \end{pmatrix}. \quad (12.11.11)$$

Let  $a = e^{i\frac{\alpha}{2}}, b = 0$ , and substitute into equation (12.11.5). The  $SO(3)$  matrix corresponding to a rotation by  $\alpha$  about the latest  $z''$ -axis is

$$R_{z^*}(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (12.11.12)$$

From equation (12.11.1), the  $SU(2)$  matrix corresponding to  $R_{z^*}(\alpha)$  is

$$U_{z^*}(\alpha) = \begin{pmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}} \end{pmatrix}. \quad (12.11.13)$$

Any rotation in three-dimensional space is equivalent to the following successive rotations:

$$R_S(\alpha, \beta, \gamma) = R_{z^*}(\alpha) R_{y^*}(\beta) R_z(\gamma)$$

$$= \begin{pmatrix} \cos \alpha \cos \beta \cos \gamma - \sin \alpha \sin \gamma & -\cos \alpha \cos \beta \sin \gamma - \sin \alpha \cos \gamma & \cos \alpha \sin \beta \\ \sin \alpha \cos \beta \cos \gamma + \cos \alpha \sin \gamma & -\sin \alpha \cos \beta \sin \gamma + \cos \alpha \cos \gamma & \sin \alpha \sin \beta \\ -\sin \beta \cos \gamma & \sin \beta \sin \gamma & \cos \beta \end{pmatrix}. \quad (12.11.14)$$

$R_S(\alpha, \beta, \gamma)$  corresponds to the successive transformation in  $SU(2)$  is

$$U(\alpha, \beta, \gamma) = U_{z^*}(\alpha) U_{y^*}(\beta) U_z(\gamma) = \begin{pmatrix} e^{\frac{i}{2}(\alpha+\gamma)} \cos \frac{\beta}{2} & -e^{\frac{i}{2}(\alpha-\gamma)} \sin \frac{\beta}{2} \\ e^{-\frac{i}{2}(\alpha-\gamma)} \sin \frac{\beta}{2} & e^{\frac{i}{2}(\alpha+\gamma)} \cos \frac{\beta}{2} \end{pmatrix}. \quad (12.11.15)$$

From the analysis above, it is clear that for every element of the  $SO(3)$  group, there exists a corresponding element in the  $SU(2)$  group.

This homomorphism is also a 2:1 mapping. The identity matrix in  $SO(3)$ ,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

corresponds to the following two  $SU(2)$  matrices:

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad -E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

because

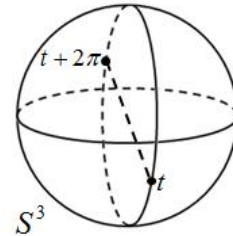
$$E \rightarrow R_U(0,0,0), \quad -E \rightarrow R_U(2\pi, 2\pi, 2\pi).$$

In the  $SO(3)$  group,

$$R_U(0,0,0) = R_U(2\pi, 2\pi, 2\pi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

represent the same element. Therefore, the kernel of the homomorphism  $f$  from  $SU(2)$  to  $SO(3)$  is  $N(E, -E)$ . For any element  $g$  in  $SU(2)$ , there are two elements  $(Eg, -Eg)$  in  $SU(2)$  that correspond to the same element in  $SO(3)$ .

As discussed earlier, the  $SU(2)$  group is homeomorphic to the three-dimensional sphere  $S^3$  of radius 1, and each pair of antipodal points on  $S^3$  corresponds to the two distinct elements  $(Eg, -Eg)$  in  $SU(2)$ . Since these two different elements  $(Eg, -Eg)$  in  $SU(2)$  correspond to the same element in  $SO(3)$ , if we identify all antipodal points on  $S^3$ , we obtain the geometric structure of the  $SO(3)$  group, which is the three-dimensional real projective space  $RP^3$ , as illustrated in Figure 12.11.1.



**Figure 12.11.1** Identification of antipodal points on  $S^3$  yields  $RP^3$

Finally, we establish the correspondence between the  $SO(3)$  group expressed by equation (12.9.4) and the  $SU(2)$  group expressed by equation (12.10.8).

Let  $a = \cos \frac{\theta_1}{2}$ ,  $b = i \sin \frac{\theta_1}{2}$ , and substitute into equation (12.11.1) to obtain the element of  $SU(2)$ :

$$\begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix}, \quad (12.11.16)$$

substituting further into equation (12.11.5) yields the element of  $SO(3)$ :

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix}. \quad (12.11.17)$$

Thus, the matrix (12.11.16) corresponds to the rotation matrix about the  $x$ -axis in three-dimensional space, given by (12.11.17).

Let  $a = \cos \frac{\theta_2}{2}$ ,  $b = -\sin \frac{\theta_2}{2}$ , and substitute into equation (12.11.1) to obtain the element of  $SU(2)$ :

$$\begin{pmatrix} \cos \frac{\theta_2}{2} & -\sin \frac{\theta_2}{2} \\ \sin \frac{\theta_2}{2} & \cos \frac{\theta_2}{2} \end{pmatrix}, \quad (12.11.18)$$

substituting into (12.11.5) yields the element of  $SO(3)$ :

$$\begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{pmatrix}. \quad (12.11.19)$$

Thus, the matrix (12.11.18) corresponds to the rotation matrix about the  $y$ -axis in three-dimensional space, given by (12.11.19).

Let  $a = e^{i\frac{\theta_3}{2}}$ ,  $b = 0$ , and substitute into equation (12.11.1) to obtain the element of  $SU(2)$ :

$$\begin{pmatrix} e^{i\frac{\theta_3}{2}} & 0 \\ 0 & e^{-i\frac{\theta_3}{2}} \end{pmatrix}, \quad (12.11.20)$$

substituting into equation (12.11.5) yields the element of  $SO(3)$ :

$$\begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (12.11.21)$$

Thus, the matrix (12.11.20) corresponds to the rotation matrix about the  $z$ -axis in three-dimensional space, given by (12.11.21).

## §12.12 Direct Product of Lie Groups

### 1. Direct Product Group

The direct product of groups is an extension of the direct product of sets.

Let there be two groups  $G_1$  and  $G_2$ . They may be identical ( $G_1 = G_2$ ), different ( $G_1 \neq G_2$ ), or have their own multiplication rules. However, we can perform a direct product operation on these two groups.

First construct the direct product set  $G$  of the two sets  $G_1$  and  $G_2$ :

$$G = G_1 \times G_2, \quad x_1 \in G_1, \quad x_2 \in G_2, \quad x = (x_1, x_2) \in G.$$

Then define the multiplication of elements in the direct product set  $G$ : let  $x = (x_1, x_2) \in G$ ,  $y = (y_1, y_2) \in G$ , with  $x_1, y_1 \in G_1, x_2, y_2 \in G_2$ . Define the product of  $x$  and  $y$  as

$$xy = (x_1, x_2) (y_1, y_2) = (x_1 y_1, x_2 y_2). \quad (12.12.1)$$

$G$  is called the **direct product group** of  $G_1$  and  $G_2$ .

In the multiplication definition (12.12.1), elements of  $G_1$  are multiplied with elements of  $G_1$ , and elements of  $G_2$  are multiplied with elements of  $G_2$ . Since when  $G_1$  and  $G_2$  are different groups, multiplication between an element of  $G_1$  and an element of  $G_2$  has not been defined, such a product is not considered here.

It can be verified that the multiplication defined by (12.12.1) makes  $G$  a group.

1)**Closure**: Let

$$x = (x_1, x_2) \in G, \quad y = (y_1, y_2) \in G,$$

then from (12.12.1) we have

$$xy = (x_1, x_2) (y_1, y_2) = (x_1 y_1, x_2 y_2),$$

since  $x_1 y_1 \in G_1, x_2 y_2 \in G_2$ , it follows that  $xy \in G$ .

2)**Associativity**: Let

$$x = (x_1, x_2) \in G, \quad y = (y_1, y_2) \in G, \quad z = (z_1, z_2) \in G,$$

then from (12.12.1) we have

$$x(yz) = (x_1(y_1 z_1), x_2(y_2 z_2)), \quad (xy)z = ((x_1 y_1)z_1, (x_2 y_2)z_2).$$

Because

$$x_1(y_1 z_1) = (x_1 y_1)z_1, \quad x_2(y_2 z_2) = (x_2 y_2)z_2,$$

we have  $x(yz) = (xy)z$ .

3)**Identity element**:  $G$  has the identity element  $e = (e_1, e_2)$ , where  $e_1$  is the identity of  $G_1$  and  $e_2$  is the identity of  $G_2$ .

4)**Inverse**: Every element  $x = (x_1, x_2)$  has an inverse

$$x^{-1} = (x_1^{-1}, x_2^{-1}), \quad x_1^{-1} \in G_1, \quad x_2^{-1} \in G_2.$$

## 2.Direct Product of Lie Groups

Let  $G_1$  and  $G_2$  be two Lie groups. Define the group multiplication on the product manifold  $G_1 \times G_2$  as follows:

For  $(a_1, a_2), (b_1, b_2) \in G_1 \times G_2$ , set

$$(a_1, a_2) (b_1, b_2) = (a_1 b_1, a_2 b_2), \quad (12.12.2)$$

Then  $G_1 \times G_2$  equipped with operation (12.12.2) becomes a Lie group, called the **direct product** of the Lie groups  $G_1$  and  $G_2$ .

The multiplication defined in (12.12.2) is consistent with that in (12.12.1).

## 3.Direct Sum of Lie Algebras

The Lie algebra of a Lie group is also a linear space; thus, two or more Lie algebras can be combined via the direct sum operation as in Definition 5.3.2.

The direct sum of Lie algebras possesses the following properties.

Let the Lie algebras of the Lie groups  $G_i$  and  $G_j$  be  $\text{Lie}(G_i)$  and  $\text{Lie}(G_j)$ , respectively.

Then:

$$1) \text{Lie}(G_i) \cap \text{Lie}(G_j) = \emptyset, \quad i \neq j;$$

$$2) [X_i, X_j] = 0 \quad \text{if } X_i \in \text{Lie}(G_i), X_j \in \text{Lie}(G_j), i \neq j.$$

**Theorem 12.12.1** The direct sum of two Lie algebras corresponds to the direct product of the two Lie groups, and conversely, the direct product of two Lie groups corresponds to the direct sum of their Lie algebras.

**Theorem 12.12.2** If  $G$  is a connected abelian (commutative) Lie group, then  $G$  is

isomorphic to the product of a Euclidean space additive group  $(R^k, +)$  and a torus  $T^n = S^1 \times \cdots \times S^1$ . If  $G$  is a compact abelian Lie group, then  $G$  is isomorphic to  $T^n \times H$ , where  $H$  is a finite abelian group.

#### 4. $SO(4, R)$ Group

The  $SO(4, R)$  group is a 6-dimensional real matrix group, which is the transformation group in four-dimensional space. Let  $A$  be an element of  $SO(4, R)$ ; then

$$A^T = A^{-1}, \quad |A| = 1.$$

Let the coordinates of four-dimensional space be  $x, y, z, t$ . The infinitesimal generators of  $SO(4, R)$  are

$$\begin{aligned} X_1 &= z \frac{\partial}{\partial y} - y \frac{\partial}{\partial z}, & X_2 &= x \frac{\partial}{\partial z} - z \frac{\partial}{\partial x}, & X_3 &= y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y}, \\ Y_1 &= x \frac{\partial}{\partial t} - t \frac{\partial}{\partial x}, & Y_2 &= y \frac{\partial}{\partial t} - t \frac{\partial}{\partial y}, & Y_3 &= z \frac{\partial}{\partial t} - t \frac{\partial}{\partial z}. \end{aligned}$$

Define

$$J_i = \frac{1}{2}(X_i + Y_i), \quad K_i = \frac{1}{2}(X_i - Y_i), \quad i = 1, 2, 3.$$

Then the following commutation relations hold:

$$\begin{aligned} [J_1, J_2] &= J_3, & [J_2, J_3] &= J_1, & [J_3, J_1] &= J_2, \\ [K_1, K_2] &= K_3, & [K_2, K_3] &= K_1, & [K_3, K_1] &= K_2, \\ [K_j, J_i] &= 0, & i, j &= 1, 2, 3. \end{aligned}$$

We observe that the commutation relations among  $J_1, J_2, J_3$ , as well as those among  $K_1, K_2, K_3$ , are identical to the commutation relations of the infinitesimal generators of the  $SO(3)$  group or the  $SU(2)$  group. Therefore, the infinitesimal generators of  $SO(4, R)$  can be regarded as composed of two sets of  $SO(3)$  (or  $SU(2)$ ) infinitesimal generators. In other words, the Lie algebra of  $SO(4, R)$  is the direct sum of the Lie algebras of two  $SO(3)$  groups or two  $SU(2)$  groups.

Since the Lie algebra of  $SO(4, R)$  is the sum of the Lie algebras of two  $SO(3)$  groups, it follows from Theorem 12.12.1 that  $SO(4, R)$  is the direct product of two  $SO(3)$  groups, i.e.,

$$SO(4, R) = SO(3) \times SO(3).$$

Similarly, because the Lie algebra of  $SO(4, R)$  is also the sum of the Lie algebras of two  $SU(2)$  groups,  $SO(4, R)$  is likewise the direct product of two  $SU(2)$  groups, i.e.,

$$SO(4, R) = SU(2) \times SU(2).$$

Combining the two relations above yields the important conclusion

$$SO(3) \times SO(3) = SU(2) \times SU(2). \quad (12.12.3)$$

#### 5. Connection of a Direct-Product Lie Group

Let  $G_1$  and  $G_2$  be two compact connected Lie groups, and let their product manifold be  $G_1 \times G_2$ . Consider left-invariant vector fields  $X \in \text{Lie}(G_1)$  and  $Y \in \text{Lie}(G_2)$ , then  $(X + Y)$  is a left-invariant vector field on  $G$ . According to equation (12.8.1), we have

$$D_{X+Y}X + Y = 0.$$

Expanding this gives

$$D_X X + D_X Y + D_Y X + D_Y Y = 0.$$

From equation (12.8.1),  $D_X X = 0$  and  $D_Y Y = 0$ ; hence

$$D_X Y + D_Y X = 0. \quad (12.12.4)$$

Assume the affine connection  $D$  is torsion-free. Then by equation (10.3.14),

$$D_X Y - D_Y X = [X, Y].$$

Substituting (12.12.4) into this yields

$$D_X Y = \frac{1}{2}[X, Y].$$

Since  $X$  and  $Y$  are left-invariant vector fields belonging to different Lie groups, the second property of the direct sum of Lie algebras gives  $[X, Y] = 0$ . Therefore,

$$D_X Y = 0. \quad (12.12.5)$$

Alternatively, Equation (12.12.5) can also be obtained directly from Equation (11.10.6).

### 6. Curvature Tensor of a Direct-Product Lie Group

Let  $X_1, Y_1, Z_1, W_1$  be left-invariant vector fields on the compact connected Lie group  $G_1$ , and let  $X_2, Y_2, Z_2, W_2$  be left-invariant vector fields on the compact connected Lie group  $G_2$ . Then  $X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2, W_1 + W_2$  are left-invariant vector fields on the direct-product Lie group  $G_1 \times G_2$ .

Since  $X_1, Y_1, Z_1, W_1$  and  $X_2, Y_2, Z_2, W_2$  belong to different Lie groups, the second property of the direct sum of Lie algebras gives

$$[X_1, Y_2] = 0, \quad [X_2, Y_1] = 0, \quad [W_1, Z_2] = 0, \quad [W_2, Z_1] = 0.$$

Therefore,

$$\begin{aligned} & [X_1 + X_2, Y_1 + Y_2] \\ &= (X_1 + X_2)(Y_1 + Y_2) - (Y_1 + Y_2)(X_1 + X_2) \\ &= X_1 Y_1 + X_1 Y_2 + X_2 Y_1 + X_2 Y_2 - Y_1 X_1 - Y_1 X_2 - Y_2 X_1 - Y_2 X_2 \\ &= [X_1, Y_1] + [X_1, Y_2] + [X_2, Y_1] + [X_2, Y_2] \\ &= [X_1, Y_1] + [X_2, Y_2]. \end{aligned}$$

That is

$$[X_1 + X_2, Y_1 + Y_2] = [X_1, Y_1] + [X_2, Y_2]. \quad (12.12.6)$$

Similarly, we obtain

$$[W_1 + W_2, Z_1 + Z_2] = [W_1, Z_1] + [W_2, Z_2]. \quad (12.12.7)$$

From equation (12.8.3), we obtain

$$\begin{aligned} R(X_1 + X_2, Y_1 + Y_2)(Z_1 + Z_2) &= \frac{1}{4} [Z_1 + Z_2, [X_1 + X_2, Y_1 + Y_2]] \\ &= \frac{1}{4} \{ (Z_1 + Z_2)[X_1 + X_2, Y_1 + Y_2] - [X_1 + X_2, Y_1 + Y_2](Z_1 + Z_2) \} \\ &= \frac{1}{4} \{ Z_1[X_1 + X_2, Y_1 + Y_2] + Z_2[X_1 + X_2, Y_1 + Y_2] \\ &\quad - [X_1 + X_2, Y_1 + Y_2]Z_1 - [X_1 + X_2, Y_1 + Y_2]Z_2 \} \\ &= \frac{1}{4} \{ Z_1[X_1, Y_1] + Z_1[X_2, Y_2] + Z_2[X_1, Y_1] + Z_2[X_2, Y_2] \\ &\quad - [X_1, Y_1]Z_1 - [X_2, Y_2]Z_1 - [X_1, Y_1]Z_2 - [X_2, Y_2]Z_2 \} \\ &= \frac{1}{4} \{ Z_1[X_1, Y_1] + Z_2[X_2, Y_2] - [X_1, Y_1]Z_1 - [X_2, Y_2]Z_2 \} \\ &= \frac{1}{4} [Z_1, [X_1, Y_1]] + \frac{1}{4} [Z_2, [X_2, Y_2]] \\ &= R(X_1, Y_1)Z_1 + R(X_2, Y_2)Z_2. \end{aligned}$$

That is,

$$R(X_1 + X_2, Y_1 + Y_2)(Z_1 + Z_2) = R(X_1, Y_1)Z_1 + R(X_2, Y_2)Z_2. \quad (12.12.8)$$

From equation (12.8.4), we have

$$\begin{aligned} & R(X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2, W_1 + W_2) \\ &= \frac{1}{4} \langle [X_1 + X_2, Y_1 + Y_2], [W_1 + W_2, Z_1 + Z_2] \rangle \\ &= \frac{1}{4} \langle [X_1 + Y_1] + [X_2 + Y_2], [W_1 + Z_1] + [W_2 + Z_2] \rangle \end{aligned}$$



$$= \frac{1}{4} \langle [X_1 + Y_1], [W_1 + Z_1] \rangle + \frac{1}{4} \langle [X_2 + Y_2], [W_2 + Z_2] \rangle$$

(Note: from equation (11.10.1),  $\langle [X_2 + Y_2], [W_1 + Z_1] \rangle = 0$ ,  $\langle [X_1 + Y_1], [W_2 + Z_2] \rangle = 0$ .)

$$= R(X_1, Y_1, Z_1, W_1) + R(X_2, Y_2, Z_2, W_2).$$

That is

$$R(X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2, W_1 + W_2) = R(X_1, Y_1, Z_1, W_1) + R(X_2, Y_2, Z_2, W_2). \quad (12.12.9)$$

Thus, the curvature tensor field of the direct-product Lie group  $G_1 \times G_2$  equals the sum of the curvature tensor fields of the Lie groups  $G_1$  and  $G_2$ .

Let the local coordinate system of the Lie group  $G_1$  at point  $p \in G_1$  be  $(U; u^i)$ , and let the local coordinate system of the Lie group  $G_2$  at point  $q \in G_2$  be  $(V; v^\alpha)$ . Then the local coordinate system of the direct-product Lie group  $G_1 \times G_2$  at point  $(p, q)$  is  $(U \times V; u^i, v^\alpha)$ . From equation (12.12.8), we obtain the components  $R_{ijk}^h$  of the curvature tensor field  $R(X_1 + X_2, Y_1 + Y_2)(Z_1 + Z_2)$  in the local coordinates  $(U \times V; u^i, v^\alpha)$ . Using equation (10.3.4), we have

$$\begin{aligned} R_{ijk}^h &= \left\langle R \left( \frac{\partial}{\partial u^j} + \frac{\partial}{\partial v^j}, \frac{\partial}{\partial u^k} + \frac{\partial}{\partial v^k} \right) \left( \frac{\partial}{\partial u^i} + \frac{\partial}{\partial v^i} \right), du^h + dv^h \right\rangle \\ &= \left\langle R \left( \frac{\partial}{\partial u^j}, \frac{\partial}{\partial u^k} \right) \frac{\partial}{\partial u^i} + R \left( \frac{\partial}{\partial v^j}, \frac{\partial}{\partial v^k} \right) \frac{\partial}{\partial v^i}, du^h + dv^h \right\rangle \\ &= \left\langle R \left( \frac{\partial}{\partial u^j}, \frac{\partial}{\partial u^k} \right) \frac{\partial}{\partial u^i}, du^h + dv^h \right\rangle + \left\langle R \left( \frac{\partial}{\partial v^j}, \frac{\partial}{\partial v^k} \right) \frac{\partial}{\partial v^i}, du^h + dv^h \right\rangle \\ &= \left\langle R \left( \frac{\partial}{\partial u^j}, \frac{\partial}{\partial u^k} \right) \frac{\partial}{\partial u^i}, du^h \right\rangle + \left\langle R \left( \frac{\partial}{\partial u^j}, \frac{\partial}{\partial u^k} \right) \frac{\partial}{\partial u^i}, dv^h \right\rangle \\ &\quad + \left\langle R \left( \frac{\partial}{\partial v^j}, \frac{\partial}{\partial v^k} \right) \frac{\partial}{\partial v^i}, du^h \right\rangle + \left\langle R \left( \frac{\partial}{\partial v^j}, \frac{\partial}{\partial v^k} \right) \frac{\partial}{\partial v^i}, dv^h \right\rangle \\ &= \left\langle R \left( \frac{\partial}{\partial u^j}, \frac{\partial}{\partial u^k} \right) \frac{\partial}{\partial u^i}, du^h \right\rangle + \left\langle R \left( \frac{\partial}{\partial v^j}, \frac{\partial}{\partial v^k} \right) \frac{\partial}{\partial v^i}, dv^h \right\rangle \\ &= \tilde{R}_{ijk}^h + \hat{R}_{ijk}^h. \end{aligned}$$

That is

$$R_{ijk}^h = \tilde{R}_{ijk}^h + \hat{R}_{ijk}^h. \quad (12.12.10)$$

### 7. Ricci Tensor of the Direct-Product Lie Group

Contracting indices  $h$  and  $j$  of  $R_{ijk}^h$ , the components of the Ricci tensor are

$$R_{ihk}^h = \tilde{R}_{ihk}^h + \hat{R}_{ihk}^h,$$

that is

$$R_{ik} = \tilde{R}_{ik} + \hat{R}_{ik}. \quad (12.12.11)$$

### 8. Scalar Curvature of a Direct-Product Lie Group

Let the orthonormal basis of the tangent space at point  $p \in G_1$  of the compact connected Lie group  $G_1$  be  $\{\tilde{e}_i\}$ , and let the orthonormal basis of the tangent space at point  $q \in G_2$  of the compact connected Lie group  $G_2$  be  $\{\hat{e}_i\}$ . Then the orthonormal basis of the tangent space at point  $(p, q)$  of the direct-product Lie group  $G_1 \times G_2$  is the union of these two bases:  $\{\tilde{e}_i, \hat{e}_i\}$ . Using equation (11.2.9) and equation (12.12.9), the scalar curvature of the direct-product Lie group  $G_1 \times G_2$  is

$$S = \sum_{i,j} R(\tilde{e}_i + \hat{e}_i, \tilde{e}_j + \hat{e}_j, \tilde{e}_j + \hat{e}_j, \tilde{e}_i + \hat{e}_i) = \sum_{i,j} R(\tilde{e}_i, \tilde{e}_j, \tilde{e}_j, \tilde{e}_i) + R(\hat{e}_i, \hat{e}_j, \hat{e}_j, \hat{e}_i) = \tilde{S} + \hat{S},$$

that is,

$$S = \tilde{S} + \hat{S}, \quad (12.12.12)$$

where

$$\tilde{S} = \sum_{i,j} R(\tilde{e}_i, \tilde{e}_j, \tilde{e}_j, \tilde{e}_i), \quad \hat{S} = \sum_{i,j} R(\hat{e}_i, \hat{e}_j, \hat{e}_j, \hat{e}_i)$$

are the scalar curvatures of the Lie groups  $G_1$  and  $G_2$ , respectively.

From equation (11.10.1) together with equations (12.12.8) through (12.12.12), we see that the metric tensor, curvature tensor, Ricci tensor, and scalar curvature of the Lie group  $G_1 \times G_2$  at point  $(p, q)$  equal the sums of the corresponding tensors of the constituent Lie groups  $G_1$  at point  $p$  and  $G_2$  at point  $q$ . This is easy to understand: originally, points  $p$  and  $q$  belonged to different Lie groups  $G_1$  and  $G_2$ , and tensors at point  $p$  and point  $q$  could not be added. However, after taking the direct product of  $G_1$  and  $G_2$ , points  $p$  and  $q$  merge into a single point  $(p, q)$ ; consequently, the tensors at point  $p$  and point  $q$  can now be added.

### 9. Internal Direct Product of Groups

For a group  $G$ , define the set

$$G \otimes G = \{(a_i, a_i) \mid a_i \in G\},$$

and call it the **internal direct product** of the group  $G$ .

$G \otimes G$  is a subset of the direct-product set  $G \times G$ , i.e.,  $G \otimes G \subset G \times G$ .

In  $G \otimes G$ , following the pattern of equation (12.12.1), define multiplication as follows: for  $(a_i, a_i), (b_j, b_j) \in G \otimes G$ , set

$$(a_i, a_i)(b_j, b_j) = (a_i b_j, a_i b_j), \quad (12.12.13)$$

Then  $G \otimes G$  with operation (12.12.13) forms a group, called the **internal direct product group** of  $G$ .

$G \otimes G$  is a subgroup of  $G \times G$ . The element  $(a_i, a_i)$  can be put in one-to-one correspondence with  $a_i$ ; therefore,  $G \otimes G$  is isomorphic to  $G$ , denoted by  $G \otimes G \cong G$ .

If  $G$  is a Lie group, then  $G \otimes G$  is the internal direct product group of the Lie group  $G$ .

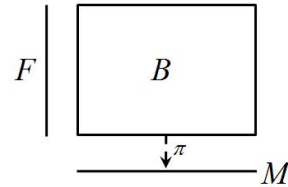
## §12.13 Principal Bundles

### 1. Differentiable Fiber Bundles

**Definition 12.13.1** As shown in Figure 12.13.1, let  $B, M, F$  be smooth manifolds, and let  $G$  be a Lie transformation group acting on the left on the smooth manifold  $F$ . Suppose there exists a smooth map  $\pi: B \rightarrow M$  and the following three conditions hold:

1) There exists an open covering  $\{U_\alpha; \alpha \in I\}$  of  $M$  (where  $I$  is an index set,  $I = \{1, 2, \dots\}$ ), and for each  $\alpha \in I$  there is a diffeomorphism  $\psi_\alpha: U_\alpha \times F \rightarrow \pi^{-1}(U_\alpha)$  such that for any  $p \in U_\alpha$  and any  $f \in F$ ,

$$\pi \circ \psi_\alpha(p, f) = p;$$



**Figure 12.13.1** Relationship among  $B, M$  and  $F$

2) For a fixed point  $p \in U_\alpha$  and any  $f \in F$ , define  $\psi_{\alpha,p}(f) = \psi_\alpha(p, f)$ . Then the map  $\psi_{\alpha,p}: F \rightarrow \pi^{-1}(p)$  is a diffeomorphism. Moreover, when  $p \in U_\alpha \cap U_\beta \neq \emptyset$ , the

diffeomorphism  $\psi_{\alpha,p}^{-1} \circ \psi_{\beta,p} : F \rightarrow F$  is equivalent to the left action of an element  $g_{\alpha\beta}(p) \in G$  on  $F$ , i.e.,

$$\psi_{\alpha,p}^{-1} \circ \psi_{\beta,p}(f) = g_{\alpha\beta}(p) \cdot f, \quad \forall f \in F;$$

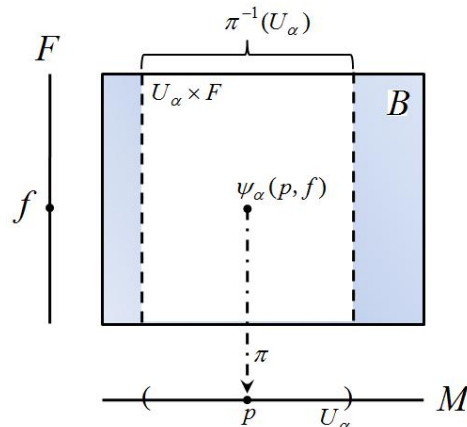
3) When  $U_\alpha \cap U_\beta \neq \emptyset$ , the map  $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow G$  is smooth.

Then  $(B, M, F, \pi, G)$  is called a **differentiable fiber bundle** over the smooth manifold  $M$ . Here  $B$  is the **total space**,  $M$  the **base manifold**,  $F$  the **fiber type**,  $\pi$  the **bundle projection**, and  $G$  the **structure group**. The differentiable fiber bundle is often simply denoted by  $B$ .

The first condition is explained below.

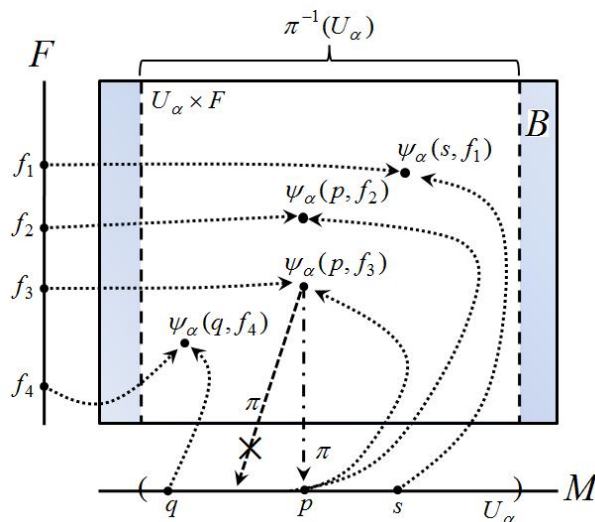
As shown in Figure 12.13.2, the mapping  $\pi : B \rightarrow M$  is the projection from the total space  $B$  to the base manifold  $M$ . Its inverse  $\pi^{-1}$ , when restricted to an open set  $U_\alpha$  of the base manifold  $M$ , gives a local portion  $\pi^{-1}(U_\alpha)$  of the total space  $B$  (the region between the two dashed lines in the figure). This local portion  $\pi^{-1}(U_\alpha)$  is diffeomorphic to the product space  $U_\alpha \times F$ , i.e.,

$$\psi_\alpha : U_\alpha \times F \rightarrow \pi^{-1}(U_\alpha)$$



**Figure 12.13.2** Fiber bundles can be locally trivialized

is a diffeomorphism. Therefore, the region between the two dashed lines in the figure represents both  $U_\alpha \times F$  and  $\pi^{-1}(U_\alpha)$ . The mapping  $\psi_\alpha : U_\alpha \times F \rightarrow \pi^{-1}(U_\alpha)$  is called a **local trivialization** of the fiber bundle  $B$ . In other words, a fiber bundle can locally be regarded as the product of an open set  $U_\alpha$  of the base manifold and the fiber type  $F$ . The property that a fiber bundle can be locally trivialized is the most essential characteristic of a fiber bundle.

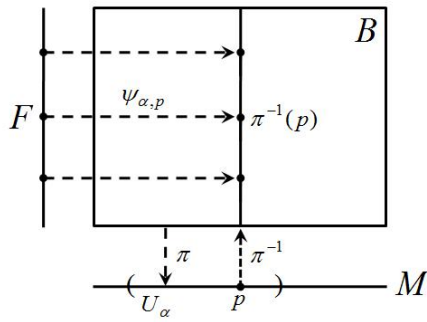


**Figure 12.13.3** Smooth diffeomorphism  $\psi_\alpha : U_\alpha \times F \rightarrow \pi^{-1}(U_\alpha)$

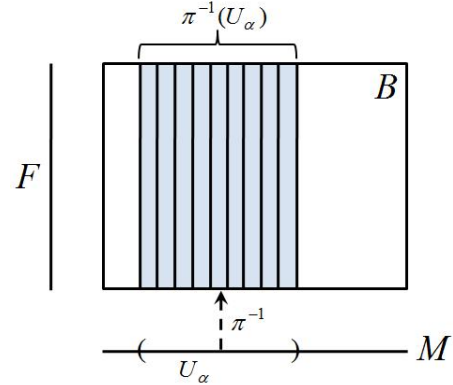
As shown in Figure 12.13.3, the smooth diffeomorphism  $\psi_\alpha(s, f_1)$  maps the point  $s$  of the base manifold  $M$  and the point  $f_1$  of the fiber type  $F$  together to a point  $\psi_\alpha(s, f_1)$  on  $\pi^{-1}(U_\alpha)$ . The mapping  $\psi_\alpha(p, f_2)$  maps the point  $p$  of the base manifold  $M$  and the point  $f_2$  of the fiber type  $F$  together to a point  $\psi_\alpha(p, f_2)$  on  $\pi^{-1}(U_\alpha)$ . The mapping  $\psi_\alpha(p, f_3)$  maps the

point  $p$  of the base manifold  $M$  and the point  $f_3$  of the fiber type  $F$  together to a point  $\psi_\alpha(p, f_3)$  on  $\pi^{-1}(U_\alpha)$ . The mapping  $\psi_\alpha(q, f_4)$  maps the point  $q$  of the base manifold  $M$  and the point  $f_4$  of the fiber type  $F$  together to a point  $\psi_\alpha(q, f_4)$  on  $\pi^{-1}(U_\alpha)$ , and so on. In these mapping processes, the first condition must be satisfied. For example, it is required that the image of  $\psi_\alpha(p, f_3)$  can only project to the point  $p$  on the base manifold  $M$  and not to any other point; i.e., it is required that  $\pi \circ \psi_\alpha(p, f_3) = p$ .

The second condition is explained below.

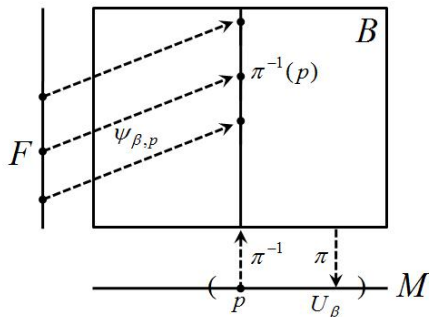


**Figure 12.13.4** Fiber at point  $p$



**Figure 12.13.5** Attaching fibers at every point of  $U_\alpha$

As shown in Figure 12.13.4, if the point  $p \in U_\alpha$  in the mapping  $\psi_\alpha(p, f)$  is fixed, then the mapping depends only on the variable  $f$ , while  $p$  remains constant; denote this as  $\psi_{\alpha,p}(f) = \psi_\alpha(p, f)$ . If the inverse mapping  $\pi^{-1}$  is restricted to the open set  $U_\alpha$  of  $M$  and takes all points of  $U_\alpha$ , we obtain a local region  $\pi^{-1}(U_\alpha)$  of the total space  $B$  (the region between the two dashed lines in Figure 12.13.2). However, if only a single point  $p$  of  $U_\alpha$  is taken, we obtain a fiber of the total space  $B$ , denoted as  $\pi^{-1}(p)$ , represented by a line segment in Figure 12.13.4. The mapping  $\psi_{\alpha,p}(f)$  maps all points of the fiber type  $F$  onto  $\pi^{-1}(p)$ ; that is, there exists a homeomorphism  $\psi_{\alpha,p}: F \rightarrow \pi^{-1}(p)$  between the fiber type  $F$  and the fiber  $\pi^{-1}(p)$ . Figure 12.13.4 illustrates a mapping  $\psi_{\alpha,p}(f)$  acting on the fiber type  $F$ . The manifold  $F$  is like a fiber, and the mapping process  $\psi_{\alpha,p}(f)$  corresponds to attaching this fiber at point  $p$ . The manifold  $F$  is called the **fiber type** because it serves as a template that can be used to replicate infinitely many fibers, and then a fiber can be attached at every point of the open set  $U_\alpha$  (as the point  $p$  varies to other points on  $U_\alpha$ ), as shown in Figure 12.13.5.



**Figure 12.13.6** Mapping  $\psi_{\beta,p}: F \rightarrow \pi^{-1}(p)$

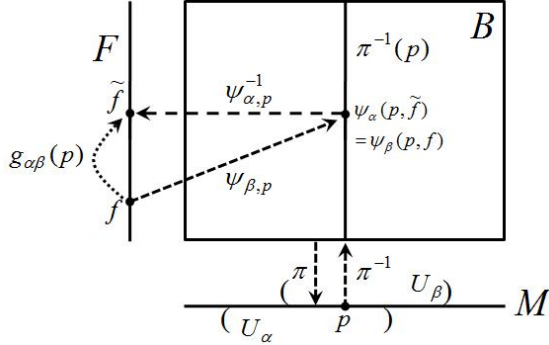
$F$  onto  $\pi^{-1}(p)$ , establishing a diffeomorphism  $\psi_{\beta,p}: F \rightarrow \pi^{-1}(p)$  between the fiber type  $F$  and

However, there is usually not just one open set on the base manifold  $M$ , but multiple ones. As shown in Figure 12.13.6, let  $U_\beta$  be another open set on the base manifold  $M$ . Suppose the point  $p \in U_\beta$  is fixed, and denote  $\psi_{\beta,p}(f) = \psi_\beta(p, f)$ . Since the point  $p \in U_\beta$  is fixed, the inverse mapping  $\pi^{-1}$  acting on the point  $p$  of  $U_\beta$  yields another fiber of the total space  $B$ , denoted as  $\pi^{-1}(p)$ . The mapping  $\psi_{\beta,p}(f)$  thus maps all points of the fiber type

the fiber  $\pi^{-1}(p)$ . Similarly, a fiber can be attached at every point of  $U_\beta$ . Because  $U_\alpha$  and  $U_\beta$  employ different local coordinate systems, in general the mappings  $\psi_{\beta,p}(f)$  and  $\psi_{\alpha,p}(f)$  are two distinct mappings, as illustrated in Figure 12.13.4 and Figure 12.13.6.

As shown in Figure 12.13.7, the open sets  $U_\alpha$  and  $U_\beta$  intersect. Let  $p \in U_\alpha \cap U_\beta \neq \emptyset$ ; then there exists a diffeomorphism

$$\psi_{\alpha,p}^{-1} \circ \psi_{\beta,p} : F \rightarrow F.$$



**Figure 12.13.7** Mapping  $\psi_{\alpha,p}^{-1} \circ \psi_{\beta,p}(f) = \tilde{f}$

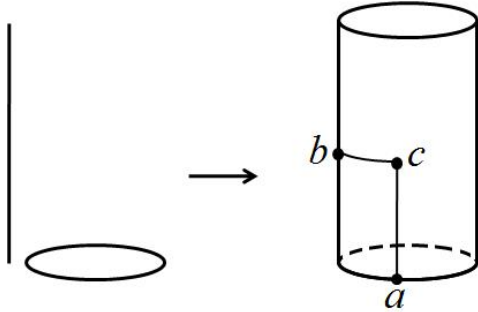
This mapping sends a point  $f$  on the fiber type  $F$  to a point  $\tilde{f}$ , i.e.,  $\psi_{\alpha,p}^{-1} \circ \psi_{\beta,p}(f) = \tilde{f}$ . The points  $f$  and  $\tilde{f}$  may be identical or different. The transition from  $f$  to  $\tilde{f}$  can also be obtained by the left action of an element  $g_{\alpha\beta}(p) \in G$  on  $F$ , i.e.,  $g_{\alpha\beta}(p) \cdot f = \tilde{f}$ . Therefore,

$$\psi_{\alpha,p}^{-1} \circ \psi_{\beta,p}(f) = g_{\alpha\beta}(p) \cdot f, \quad \forall f \in F.$$

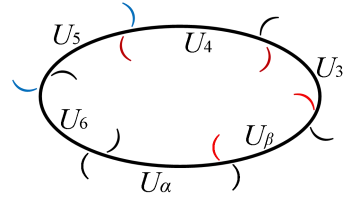
The mapping  $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow G$  is called

a **transition function** of the fiber bundle  $B$ . All transition functions together form a set  $\{g_{\alpha\beta}\}$ , called the **family of transition functions**.

**Theorem 12.13.1** Let  $p \in U_\alpha \cap U_\beta \neq \emptyset$ ,  $f, \tilde{f} \in F$ . Then the points  $\psi_\alpha(p, \tilde{f})$  and  $\psi_\beta(p, f)$  represent the same point in the total space  $B$  if and only if  $\tilde{f} = g_{\alpha\beta}(p) \cdot f$ .



**Figure 12.13.8** A cylinder as a total space



**Figure 12.13.9** The base manifold  $M$  covered by six open sets

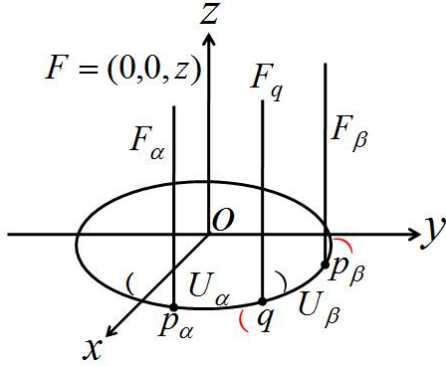
**Example 12.13.1** As shown in Figure 12.13.8, the product of a unit circle  $M$  and a line segment  $F$  forms a cylinder  $B$ . The cylinder can be regarded as a total space, the unit circle as the base manifold, and the line segment as the fiber type. The differentiable fiber bundle is denoted as  $\pi : B \rightarrow M$ . Taking a point  $a$  on the unit circle and a point  $b$  on the line segment, their product corresponds to a point  $c$ , which is equivalent to the mapping  $\psi(a, b) = c$ .

We cover the base manifold  $M$  with six open sets:  $U_\alpha$ ,  $U_\beta$ ,  $U_3$ ,  $U_4$ ,  $U_5$ , and  $U_6$ , as shown in Figure 12.13.9.

We describe the relationship among the total space  $B$ , the base manifold  $M$ , and the fiber type  $F$  in a spatial Cartesian coordinate system, as shown in Figure 12.13.10. The fiber type  $F$  is represented by the upper half of the  $z$ -axis, with  $F = (0, 0, z)$ ,  $z > 0$ . Let the local coordinate of the open set  $U_\alpha$  be  $(\cos \alpha, \sin \alpha)$ , and the local coordinate of the open set  $U_\beta$  be  $(\cos \beta, \sin \beta)$ .

Take an arbitrary point  $p_\alpha \in U_\alpha$  and fix it. Then the mapping  $\psi_\alpha(p_\alpha, f)$  has only one variable  $f$ ,  $f \in F$ , and denote  $\psi_{\alpha, p_\alpha}(f) = \psi_\alpha(p_\alpha, f)$ .  $\psi_{\alpha, p_\alpha}(f)$  can be expressed as

$$\begin{pmatrix} 1 & 0 & \frac{\cos \alpha}{z} \\ 0 & 1 & \frac{\sin \alpha}{z} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} = \begin{pmatrix} \cos \alpha \\ \sin \alpha \\ z \end{pmatrix},$$



**Figure 12.13.10** Relationship among the total space  $B$ , the base manifold  $M$ , and the fiber type  $F$

thus, the fiber  $F_\alpha$  over point  $p_\alpha \in U_\alpha$ , can be represented as  $(\cos \alpha, \sin \alpha, z)$ .

Similarly, take an arbitrary point  $p_\beta \in U_\beta$  and fix it. Then the mapping  $\psi_\beta(p_\beta, f)$  has only one variable  $f$ ,  $f \in F$ , and denote  $\psi_{\beta, p_\beta}(f) = \psi_\beta(p_\beta, f)$ .  $\psi_{\beta, p_\beta}(f)$  can be expressed as

$$\begin{pmatrix} 1 & 0 & \frac{\cos \beta}{z} \\ 0 & 1 & \frac{\sin \beta}{z} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} = \begin{pmatrix} \cos \beta \\ \sin \beta \\ z \end{pmatrix},$$

thus, the fiber  $F_\beta$  over point  $p_\beta \in U_\beta$ , can be represented as  $(\cos \beta, \sin \beta, z)$ .

The inverse mapping of  $\psi_{\alpha, p_\alpha}(f)$ , denoted as  $\psi_{\alpha, p_\alpha}^{-1}(f_\alpha)$ , can be expressed as

$$\begin{pmatrix} 1 & 0 & -\frac{\cos \alpha}{z} \\ 0 & 1 & -\frac{\sin \alpha}{z} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \alpha \\ \sin \alpha \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}.$$

The open sets  $U_\alpha$  and  $U_\beta$  intersect. Let  $q \in U_\alpha \cap U_\beta$ ; then at point  $q$  there exists a smooth diffeomorphism

$$\psi_{\alpha, q}^{-1} \circ \psi_{\beta, q} : F \rightarrow F.$$

The composite mapping  $\psi_{\alpha, q}^{-1} \circ \psi_{\beta, q}$  can be represented in matrix form as

$$\begin{pmatrix} 1 & 0 & -\frac{\cos \alpha}{z} \\ 0 & 1 & -\frac{\sin \alpha}{z} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{\cos \beta}{z} \\ 0 & 1 & \frac{\sin \beta}{z} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{\cos \beta}{z} - \frac{\cos \alpha}{z} \\ 0 & 1 & \frac{\sin \beta}{z} - \frac{\sin \alpha}{z} \\ 0 & 0 & 1 \end{pmatrix}.$$

Since the point  $q \in U_\alpha \cap U_\beta$ , we can set  $\alpha = \beta$ , and then the composite mapping  $\psi_{\alpha, q}^{-1} \circ \psi_{\beta, q}$  can be represented by the matrix

$$g_{\alpha\beta}(q) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E$$

Here  $g_{\alpha\beta}(q)$  is the identity matrix. Clearly, this mapping sends a point  $f$  on the fiber manifold  $F$

to the same point  $f$ .

Because we cover the base manifold  $M$  with six open sets  $U_\alpha, U_\beta, U_3, U_4, U_5$ , and  $U_6$ , there are altogether six transition functions, all of which can be represented by the identity matrix  $E$ . They form the family of transition functions  $\{E, E, E, E, E, E\}$ . Since every member of this family is the identity matrix, we can simplify the family to a set  $\{E\}$  consisting of a single identity matrix  $E$ . This set  $\{E\}$  is the structure group  $G$ .

## 2. Vector Bundles and Tangent Bundles

If the fiber type  $F$  is a  $q$ -dimensional vector space  $R^q$ , then the differentiable fiber bundle  $(B, M, R^q, \pi, G)$  is called a **vector bundle**. The dimension  $q$  of the vector space  $R^q$  may be equal to or different from the dimension  $m$  of the base manifold  $M$ . A vector bundle is a special kind of differentiable fiber bundle, where a vector space  $R^q$  is attached to each point of the base manifold  $M$ .

As is well known, at each point  $p \in M$  of an  $m$ -dimensional base manifold  $M$ , there exists an  $m$ -dimensional tangent space  $T_p M$ , and the tangent space  $T_p M$  is also a vector space. Therefore, the **tangent bundle**  $TM = \bigcup_{p \in M} T_p M$  formed by the base manifold  $M$  and its tangent spaces  $T_p M$  is also a vector bundle. Suppose the smooth structure of the base manifold  $M$  is given by an atlas  $\{(U_\alpha, \varphi_\alpha); \alpha \in I\}$ . Then the family of transition functions of the tangent bundle  $TM$  is  $\{g_{\alpha\beta}: U_\alpha \cap U_\beta \rightarrow GL(m)\}$ , where  $g_{\alpha\beta}$  is the Jacobian matrix of the coordinate transformation

$$\varphi_\alpha \circ \varphi_\beta^{-1}: \varphi_\beta(U_\alpha \cap U_\beta) \rightarrow \varphi_\alpha(U_\alpha \cap U_\beta),$$

i.e.,

$$g_{\alpha\beta} = \left( \frac{\partial x_\alpha^i}{\partial x_\beta^j} \right).$$

## 3. Principal Bundles

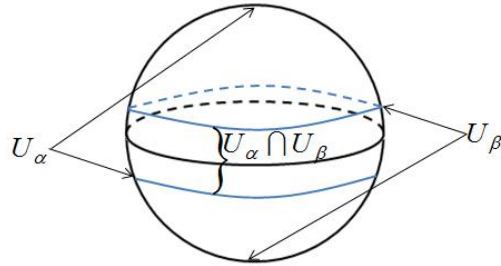
**Definition 12.13.2** If, in a differentiable fiber bundle  $(B, M, F, \pi, G)$ , the fiber type  $F$  is also the structure group  $G$ , then the bundle is called a  $G$ -**principal bundle**, denoted as  $(B, M, \pi, G)$ , or simply  $B$ .

**Example 12.13.2** Consider a sphere  $S^2$  of radius  $r$  in  $R^3$ , whose coordinates are given by

$$\begin{cases} x = r \cos \theta \cos \varphi, \\ y = r \cos \theta \sin \varphi, \\ z = r \sin \theta. \end{cases}$$

$$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \quad 0 \leq \varphi < 2\pi.$$

The sphere  $S^2$  can also be considered independently of  $R^3$ . Now suppose  $S^2$  is a unit sphere of radius 1, independent of  $R^3$ . As shown in Figure 12.13.11, we construct two local coordinate systems on  $S^2$ :



**Figure 12.13.11** Open cover of  $S^2$

$$U_\alpha = \{(\cos \theta \cos \varphi, \cos \theta \sin \varphi, \sin \theta); -\varepsilon < \theta \leq \frac{\pi}{2}, 0 \leq \varphi < 2\pi\},$$

$$U_\beta = \{(\cos \theta \cos \varphi, \cos \theta \sin \varphi, \sin \theta); -\frac{\pi}{2} \leq \theta < \varepsilon, 0 \leq \varphi < 2\pi\}.$$



The collection  $\{U_1, U_2\}$  is an open cover of  $S^2$ , and the intersection

$$U_1 \cap U_2 = \{(\cos \theta \cos \varphi, \cos \theta \sin \varphi, \sin \theta); -\varepsilon < \theta < \varepsilon, 0 \leq \varphi < 2\pi\}$$

lies near the equator.

The unitary group  $U(1) = \{z \in \mathbb{C} : |z| = 1\} = S^1$  together with  $S^2$  can form a principal bundle  $(B, S^2, \pi, U(1))$ , where the base space is  $S^2$ , and both the fiber type and the structure group are  $U(1)$ . An element of the structure group  $U(1)$  can be expressed as  $e^{i\varphi}$ . The local product spaces are  $U_\alpha \times S^1$  and  $U_\beta \times S^1$ . The mapping  $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow S^1$  serves as the transition function of the principal bundle  $B$  and can be written as  $g_{\alpha\beta}(\theta, \varphi) = e^{in\varphi}$ , where  $n$  is an integer. Let

$$p(\theta, \varphi) = (\cos \theta \cos \varphi, \cos \theta \sin \varphi, \sin \theta) \in U_1 \cap U_2,$$

$$f = e^{i\alpha} \in S^1, \quad \tilde{f} = e^{i\beta} \in S^1,$$

then we have

$$(p(\theta, \varphi), e^{i\alpha}) \in U_\alpha \times S^1, \quad (p(\theta, \varphi), e^{i\beta}) \in U_\beta \times S^1,$$

and

$$e^{i\alpha} = e^{in\varphi} e^{i\beta} = e^{i(n\varphi + \beta)}.$$

**Theorem 12.13.2** Let  $(B, M, \pi, G)$  be a principal  $G$ -bundle. Then the structure group  $G$  acts freely and smoothly on the total space  $B$  as a Lie transformation group on the right, and for any  $p \in U$  and  $g, h \in G$ , the following holds:

$$\psi(p, g) \cdot h = \psi(p, g \cdot h).$$

**Theorem 12.13.3** Every product space  $M \times F$  is a fiber bundle  $(M \times F, M, \pi, F)$ , where the projection map is the first projection  $\pi : M \times F \rightarrow M$ . Such a fiber bundle is called a **trivial fiber bundle** (or **product fiber bundle**).

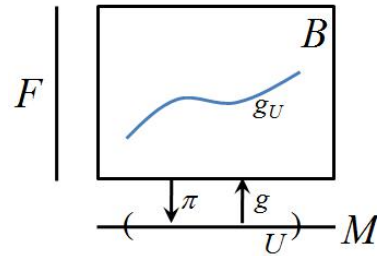
**Theorem 12.13.4** Let  $G$  be a Lie group and  $H$  a closed subgroup of  $G$ . Then  $G$  is a principal  $H$ -bundle over the quotient space  $G/H$ , i.e.,  $\pi : G \rightarrow G/H$  is a principal bundle.

Here  $G/H$  is the quotient space, a smooth manifold, and also the set of left cosets of  $H$  in  $G$ , i.e.,  $G/H = \{g \cdot H : g \in G\}$ .

#### 4. Connection and Curvature of the Base Manifold

**Definition 12.13.3** Let  $(B, M, F, \pi, G)$  be a differentiable fiber bundle, and let  $U$  be an open subset of  $M$ . If there exists a smooth map  $g : U \rightarrow B$  such that  $\pi \circ g = id : U \rightarrow U$ , then  $g$  is called a **section** of the fiber bundle  $B$  over the open subset  $U$ .

As shown in Figure 12.13.12, the curve  $g_U$  represents the image of the open set  $U$  of the base manifold  $M$  under the smooth map  $g$ .



**Figure 12.13.12** A section  $g_U$  over  $U$

Let the Lie algebra of the  $r$ -dimensional Lie group  $G$  be  $\text{Lie}(G)$ , with structure constants  $c_{jk}^i$ , and let  $\{e_1, \dots, e_r\}$  be a basis of  $\text{Lie}(G)$ . Then the commutation relations are

$$[e_j, e_k] = \sum_{i=1}^r c_{jk}^i e_i.$$

The Lie algebra  $\text{Lie}(G)$  is an  $r$ -dimensional vector space. Suppose  $f^1, \dots, f^r$  are smooth functions on the Lie group  $G$ . Define

$$f = f^1 e_1 + \dots + f^r e_r.$$

Clearly,  $f \in \text{Lie}(G)$ ,  $f$  takes values in  $\text{Lie}(G)$ . Moreover, since each  $f^1, \dots, f^r$  is a 0-form, we call  $f$  a **0-form on the Lie group  $G$  with values in the Lie algebra  $\text{Lie}(G)$** .



Analogously, replace  $f^1, \dots, f^r$  with 1-forms  $\varphi^1, \dots, \varphi^r$  on the Lie group  $G$  and define

$$\varphi = \varphi^1 e_1 + \dots + \varphi^r e_r = \varphi^i e_i.$$

Since the value  $\varphi|_p$  of  $\varphi$  at any point  $p$  on the Lie group  $G$  maps a tangent vector  $v$  at  $p$  to an element of  $\text{Lie}(G)$ :

$$\varphi|_p(v) = e_i \varphi^i|_p(v) \in \text{Lie}(G),$$

where each  $\varphi^i|_p(v)$  is a real- or complex-valued function, we call  $\varphi$  a **1-form on the Lie group  $G$  with values in the Lie algebra  $\text{Lie}(G)$** .

To facilitate understanding, we now assume that the Lie group  $G$  is a matrix group.

Let the open cover of the base manifold  $M$  be  $\{(U_\alpha, \psi_\alpha); \alpha \in I\}$ , and the family of transition functions be  $\{g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow G\}$ . On an open set  $U_\alpha$  of  $M$ , define a 1-form  $\omega_\alpha$  with values in the Lie algebra  $\text{Lie}(G)$ ,

$$\omega_\alpha = A_\mu(x) dx^\mu = A_\mu^\lambda(x) dx^\mu e_\lambda. \quad (12.13.1)$$

On the overlap  $U_\alpha \cap U_\beta$ , the transformation relation between the 1-forms  $\omega_\alpha$  and  $\omega_\beta$  defined on  $U_\alpha$  and  $U_\beta$ , respectively, is

$$\omega_\beta = g_{\alpha\beta}^{-1} \omega_\alpha g_{\alpha\beta} + g_{\alpha\beta}^{-1} dg_{\alpha\beta}. \quad (12.13.2)$$

$\omega_\alpha$  is called the **connection 1-form** on the open set  $U_\alpha$  of the base manifold  $M$ .

On the open set  $U_\alpha$  of  $M$ , define a 2-form  $\Omega_\alpha$  with values in  $\text{Lie}(G)$ ,

$$\Omega_\alpha = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu = \frac{1}{2} F_{\mu\nu}^r e_r dx^\mu \wedge dx^\nu, \quad (12.13.3)$$

where

$$F_{\mu\nu}^r = \partial_\mu A_\nu^r - \partial_\nu A_\mu^r + c_{st}^r A_\mu^s A_\nu^t. \quad (12.13.4)$$

On the overlap  $U_\alpha \cap U_\beta$ , the transformation relation between the curvature 2-forms  $\Omega_\alpha$  and  $\Omega_\beta$  defined on  $U_\alpha$  and  $U_\beta$ , respectively, is

$$\Omega_\beta = g_{\alpha\beta}^{-1} \Omega_\alpha g_{\alpha\beta}. \quad (12.13.5)$$

$\Omega_\alpha$  is called the **curvature 2-form** on the open set  $U_\alpha$  of the base manifold  $M$ .

## 5. Connection and Curvature of a Principal Bundle

Given the open cover  $\{(U_\alpha, \psi_\alpha); \alpha \in I\}$  of the base manifold  $M$ , for each open set  $U_\alpha$  of  $M$ , the principal bundle  $B$  has a smooth section

$$\begin{aligned} g_\alpha : U_\alpha &\rightarrow B, \\ x \in U_\alpha &\rightarrow g_\alpha(x). \end{aligned} \quad (12.13.6)$$

When  $U_\alpha \cap U_\beta \neq \emptyset$ , on the overlap we have

$$g_\alpha(x) = g_{\alpha\beta}(x) g_\beta(x), \quad (12.13.7)$$

where  $g_{\alpha\beta}(x) \in G$  is the transition function.

It can be considered that the connection 1-form  $\omega_\alpha$  and the curvature 2-form  $\Omega_\alpha$  on the base manifold  $M$  are the results of pulling back the forms  $\tilde{\omega}$  and  $\tilde{\Omega}$  on the principal bundle  $B$  to the base manifold via the mapping (12.13.6), i.e.,

$$g_\alpha^* \tilde{\omega} = \omega_\alpha, \quad g_\alpha^* \tilde{\Omega} = \Omega_\alpha. \quad (12.13.8)$$

$\tilde{\omega}$  is related to  $\omega_\alpha$  and  $\omega_\beta$  as follows:

$$\tilde{\omega} = g_\alpha^{-1} \omega_\alpha g_\alpha + g_\alpha^{-1} dg_\alpha = g_\beta^{-1} \omega_\beta g_\beta + g_\beta^{-1} dg_\beta. \quad (12.13.9)$$

$\tilde{\omega}$  is called the **connection form** on the principal bundle  $B$ .  $\tilde{\Omega}$  is related to  $\Omega_\alpha$  and  $\Omega_\beta$  as follows:

$$\tilde{\Omega} = g_\alpha^{-1} \Omega_\alpha g_\alpha = g_\beta^{-1} \Omega_\beta g_\beta. \quad (12.13.10)$$

$\tilde{\Omega}$  is called the **curvature form** on the principal bundle  $B$ .

**Theorem 12.13.5** Let  $\pi : B \rightarrow M$  be a principal  $G$ -bundle over a smooth manifold  $M$ , with an open cover  $\{(U_\alpha, \psi_\alpha); \alpha \in I\}$  of  $M$  and the corresponding family of transition functions  $\{g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow G\}$ . For each  $\alpha \in I$ , define a smooth section of  $B$  over  $U_\alpha$  via the map (12.13.6). If, for every  $\alpha \in I$ , a 1-form  $\omega_\alpha$  with values in the Lie algebra  $\text{Lie}(G)$  is defined on  $U_\alpha$ , such that on overlaps  $U_\alpha \cap U_\beta \neq \emptyset$  they satisfy the transformation formula (12.13.2), then one can use formula (12.13.9) to define a connection form  $\tilde{\omega}$  on the principal bundle  $B$ , and this  $\tilde{\omega}$  determines a connection on  $B$ .

## Part Two Physical Theories

### Contents

<b>Chapter 13 Two Fundamental Assumptions.....</b>	<b>205</b>
§13.1 The First Assumption.....	205
§13.2 The Second Assumption.....	207
<b>Chapter 14 Distinguishing Particles and Antiparticles by Orientation.....</b>	<b>208</b>
§14.1 Distinguishing Particles and Antiparticles Using the Orientation of Lie Groups.....	208
§14.2 Particles and Antiparticles Can Mutually Transform.....	209
<b>Chapter 15 Particles in Minkowski Spacetime.....</b>	<b>211</b>
§15.1 Particles Are All Three-Dimensional.....	211
1. $SU(2) \otimes SU(2)'$ .....	211
2. $SO(3) \otimes SO(3)'$ .....	212
3. $SO(3) \otimes SU(2)'$ .....	212
§15.2 Particles as Submanifolds.....	214
1. Submanifolds in Four-Dimensional Euclidean Space.....	214
2. Submanifolds in Minkowski Space.....	214
3. The Speed of Light is the Maximum Speed.....	215
§15.3 Why the Speed of Light is Constant.....	216
§15.4 Lorentz Transformations.....	217
1. Lorentz Transformations.....	217
2. Length Contraction.....	219
3. Velocity Transformation Formulas.....	220
§15.5 Particles Expand.....	221
1. Expansion Speed.....	221
2. Expansion Acceleration.....	224
3. The Expansion Process is a Conformal Transformation.....	227
4. No Singularity Exists.....	228
<b>Chapter 16 Lie-Group Interpretation of Quantum Mechanics Principles.....</b>	<b>229</b>
§16.1 Structure of Particles.....	229
§16.2 Point-Like Particles Undergo Random Motion.....	230
1. Point-Like Particles Undergo Random Motion.....	230
2. Why Particles Exhibit Wave-Like Behavior.....	232
3. Regarding the Local Lie Group as a Wave Packet.....	232
4. Plane-Wave Function.....	233
§16.3 Probability Interpretation of the Wave Function.....	233
§16.4 Dynamical Variables Are Represented by Linear Hermitian Operators.....	235
1. Dynamical Variables Represented by Linear Operators.....	235
2. Condition for a Dynamical Variable to Have a Definite Value.....	237
3. Dynamical Variables Are Represented by Linear Hermitian Operators.....	240
§16.5 Measurement Results of Dynamical Variables.....	243

1.An Example.....	243
2.Orthonormality.....	244
3.Completeness.....	245
4.Meaning of the Expansion Coefficients.....	246
5.Geometric Interpretation of the Expectation-Value Formula.....	247
§16.6 Conditions for Two Different Dynamical Variables to Have Simultaneous Definite Values.....	249
§16.7 Uncertainty Principle.....	249
§16.8 Identical Particles.....	252
1.Homomorphism from the $SU(2)$ Group to the $SO(3)$ Group.....	252
2.Principle of Identity.....	253
§16.9 The Three Major Equations.....	255
§16.10 Spin Operators of the $SO(3)$ Group.....	256
1.Matrix Form of the Spin Operators of the $SO(3)$ Group.....	256
2.Differential Form of the Spin Operators of the $SO(3)$ Group.....	258
§16.11 Spin Operators of the $SU(2)$ Group.....	259
1.Matrix Form of the Spin Operators of the $SU(2)$ Group.....	259
2.Differential Form of the Spin Operators of the $SU(2)$ Group.....	260
§16.12 Orbits of Point-Like Particles and Quantum Entanglement.....	262
1.Orbits of Point-Like Particles.....	262
2.Quantum Entanglement.....	263
§16.13 Forming the Projective Space $RP^3$ .....	264
1.Antiparallel-Oriented Particles Form $RP^3$ .....	264
2.Projective Space $RP^3$ Formed by Particles with the Same Orientation.....	269
<b>Chapter 17 Particles as Principal Bundles.....</b>	<b>271</b>
§17.1 The Relationship Between Inertial Force and Rest Mass.....	271
1.Expressing Inertial Force Using Divergence and Curl.....	271
2.Relationship Between the Inertial Force of a Free Particle and Its Rest Mass.....	272
§17.2 Transformations with Only a One-Parameter Transformation Group.....	273
§17.3 Physical Significance of Complex Function Integration.....	277
§17.4 Photon as a Principal Bundle.....	278
1.Circulation and Flux of the Spin Operators of the $SO(3)$ Group.....	278
2.Treating Photons as Principal Bundles.....	278
§17.5 Spin-1/2 Elementary Particles as Principal Bundles.....	280
1.Circulation and Flux of the Spin Operator $\hat{S}_x$ .....	280
2.Circulation and Flux of the Spin Operator $\hat{S}_y$ .....	280
3.Circulation and Flux of the Spin Operator $\hat{S}_z$ .....	281
4.Reasons for Unequal Rest Masses.....	281
5.Distinguishing Spin-1/2 Elementary Particles via Principal Bundles.....	282
<b>Chapter 18 The Origin of Electromagnetic Fields.....</b>	<b>285</b>
§18.1 The Reason Why Electrons/Positrons and Protons/Antiprotons Carry Charge.....	285

1. Gaussian Curvature of $S^2$ .....	285
2. Concentric Spherical Structure of Electrons/Positrons and Protons/Antiprotons.....	285
3. Electric Field Strength.....	287
4. Electric Flux.....	288
5. Gauss's Theorem for a Single Charge.....	288
6. The Origin of Positive and Negative Charges.....	289
7. The Reason for Charge Conservation.....	291
8. Gauss's Theorem for Multiple Charges.....	292
§18.2 The Reason Why Neutrinos and Antineutrinos Carry No Charge.....	294
1. Gaussian Curvature of the Torus.....	294
2. The Reason Why Neutrinos and Antineutrinos Carry No Charge.....	295
3. Free Neutrinos and Antineutrinos Move at the Speed of Light.....	299
4. Neutrinos and Antineutrinos that Appear and Disappear Intermittently.....	301
5. Three States of Neutrinos.....	302
6. Types of Neutrinos and Antineutrinos.....	303
7. Oscillations of Neutrinos and Antineutrinos.....	306
§18.3 The Reason Why Photons Carry No Charge.....	308
1. First Explanation.....	308
2. Second Explanation.....	309
3. Why Light Waves Are Transverse Waves.....	310
§18.4 Derivation of Maxwell's Equations.....	310
1. The First Equation.....	311
2. The Second Equation.....	313
3. The Third Equation.....	313
4. The Fourth Equation.....	314
5. Tensor Form of the Equations.....	315
§18.5 Coulomb Force.....	321
§18.6 Lorentz Force.....	321
§18.7 Like Charges Repel and Opposite Charges Attract.....	323
1. Repulsive Force Between Positively Charged Particles.....	323
2. Repulsive Force Between Negatively Charged Particles.....	324
3. Attractive Force Between Positive and Negative Charged Particles.....	326
<b>Chapter 19 Gravitational Field.....</b>	<b>328</b>
§19.1 Characteristics of the Gravitational Field.....	328
1. The Gravitational Field as a Generalized Riemannian Manifold.....	328
2. Bi-invariant Metric Tensor.....	328
3. Sectional Curvature.....	329
4. Gravitational Fields of Particles and Antiparticles Are Identical.....	330
5. Conditions for Spatial Symmetry.....	331
§19.2 Equations of Motion in a Gravitational Field.....	331
§19.3 Curvature of Composite Particles.....	333
§19.4 Time as a One-Parameter Transformation Group.....	334
§19.5 Describing the Gravitational Field in Minkowski Spacetime.....	336

§19.6	Einstein's Field Equations of Gravitation.....	337
§19.7	Determination of the Coefficient $\kappa$ .....	338
1.	Linear Approximation of the Metric Tensor.....	338
2.	Linear Approximation of the Gravitational Field Equations.....	339
3.	Determining $h_{00}$ .....	340
4.	Determining $\kappa$ by Comparison with the Poisson Equation.....	341
<b>Chapter 20</b>	<b>Spin Magnetic Moments of Electrons, Protons, and Neutrons.....</b>	<b>343</b>
§20.1	Integral Curves of the $SU(2)$ Group Spin Operators.....	343
1.	Integral Curves of the Spin Operator $\hat{S}_x$ for Protons and Antiprotons.....	343
2.	Integral Curves of the Spin Operator $\hat{S}_y$ for Neutrinos and Antineutrinos.....	346
3.	Integral Curves of the Spin Operator $\hat{S}_z$ for Electrons and Positrons.....	347
3.	Integral Curves of the Spin Operator $\hat{S}_z$ for Electrons and Positrons.....	348
4.	The Relationship between $\hat{S}_x$ and $\hat{S}_z$ .....	350
§20.2	Integral Curves of Photon Spin Operators.....	350
§20.3	Orbital and Spin Magnetic Moments of Electrons.....	352
1.	Magnetic Moment.....	352
2.	Orbital Magnetic Moment of Electrons.....	353
3.	Magnetic Moment as a Curve.....	354
4.	Spin Magnetic Moment of the Electron.....	355
5.	The Reason Why Electrons Possess a Spin Magnetic Moment.....	357
§20.4	Transformation from Electron to Proton.....	358
1.	General Transformation Formulas.....	358
2.	Spherical Right Triangle Formula.....	359
§20.5	Spin Magnetic Moment of the Proton.....	359
§20.6	Spin Magnetic Moment of the Neutron.....	363
1.	The Proton's Spin Magnetic Moment within the Neutron is Halved.....	363
2.	Angles Between the $z$ -Axis Components of Spin for the Proton, Electron, and Antineutrino within the Neutron.....	364
3.	The Spin Magnetic Moment of the Neutron is Negative.....	366
4.	Spherical Law of Cosines.....	366
5.	The Neutron as a Curved Manifold.....	366
6.	Units for the Spin Magnetic Moments of the Proton and Electron within the Neutron.....	367
7.	Two States of the Neutron's Spin Magnetic Moment.....	367
8.	Average Value of the Neutron's Spin Magnetic Moment.....	370
§20.7	Weak and Strong Interactions.....	376
1.	Attractive Forces Near point-like Neutrinos, Antineutrinos, and Photons.....	376
2.	Repulsive Forces Near Point-like Protons and Antiprotons.....	379
3.	Repulsive Forces Near Point-like Electrons and Positrons.....	380
§20.8	Definition of Rest Mass and Its Applications.....	382

1.A New Definition of Rest Mass.....	382
2.Reasons for Different Rest Masses.....	383
3.Reasons Why the Proton-to-Electron Rest Mass Ratio is Approximately 1836.....	384
4.Equal Rest Masses of Particles and Antiparticles.....	397
<b>Chapter 21 Transformation of Particles.....</b>	<b>398</b>
§21.1 Influence of Photons on the Energy of Other Particles.....	398
§21.2 Mutual Transformation of Particles.....	399
§21.3 Analysis of Examples.....	400
§21.4 Examination of the Solar Nuclear Reaction Model.....	401
§21.5 The Internal Mechanism of Electron-to-Proton Transformation.....	405
1.Transformation Formula.....	405
2.Transformation of the Electron's Spin Operators.....	405
3.Implementation Method.....	410
4.Elastic Collision.....	412
§21.6 Internal Mechanism of Electron-to-Neutrino Transformation.....	414
1.Transformation Formula.....	414
2.Transformation of the Electron's Spin Operators.....	416
§21.7 Internal Mechanism of Proton-to-Neutrino Transformation.....	418
1.Transformation Formula.....	418
2.Transformation of the Proton's Spin Operators.....	420
§21.8 Internal Mechanism of Photon Transformation into Other elementary Particles.....	422
1.Transformation of the photon's spin operator $\hat{L}_x$ .....	423
2.Transformation of the photon's spin operator $\hat{L}_y$ .....	424
3.Transformation of the photon's spin operator $\hat{L}_z$ .....	425
4.Summary.....	426
§21.9 The Structure and Composition of the Universe.....	426
1.The Structure of the Universe.....	426
2.Types of Elementary Particles.....	427
§21.10 The Origin of Matter–Antimatter Asymmetry.....	429
1.Orientation of Minkowski Spacetime.....	429
2.The Reason for the Scarcity of Antimatter.....	431
§21.11 Properties of Dark Particles and the Birth of Particles.....	432
1.Topological Structure of Elementary Dark Particles.....	432
2.Dark Particles Have Antiparticles.....	432
3.The Spin of an Elementary Dark Particle is 1.....	432
4.Dark Particles Possess a Gravitational Field.....	435
5.Classification of Elementary Dark Particles.....	436
6.Rest Mass and Short-Range Forces of Elementary Dark Particles.....	437
7.Dark Particles Do Not Generate Electromagnetic Fields.....	438
8.Similar Properties of Dark Particles and Their Antiparticles.....	438
9.Conception and Birth of Particles.....	438

10.Obeying Conservation Laws.....	441
11.The Abundance of Dark Matter in the Universe.....	443
12.Reasons Why the Lorentz Transformation is More Accurate.....	446
<b>References.....</b>	<b>449</b>



## Chapter 13 Two Fundamental Assumptions

### §13.1 The First Assumption

**Assumption One:** Photons and antiphotons, electrons and positrons, protons and antiprotons, neutrinos and antineutrinos, dark matter particles  $\chi_0$  and anti-dark matter particles  $\bar{\chi}_0$ , dark matter particles  $\chi_1$  and anti-dark matter particles  $\bar{\chi}_1$ , dark matter particles  $\chi_2$  and anti-dark matter particles  $\bar{\chi}_2$  that do not decay are all elementary particles; particles that do decay are composed of combinations of the particles produced from their decay.

We refer to the dark matter particles  $\chi_0$  and  $\bar{\chi}_0$ ,  $\chi_1$  and  $\bar{\chi}_1$ ,  $\chi_2$  and  $\bar{\chi}_2$  as elementary dark matter particles. Their detailed descriptions will be provided later in this book.

The reason for assuming that the observed photons, antiphotons, electrons, positrons, protons, antiprotons, neutrinos, antineutrinos, and elementary dark matter particles are elementary is that all these particles are stable and do not decay, whereas all other observed particles decay, and their final decay products are these elementary particles.

Based on this assumption, we assert: **All discovered or yet-to-be-discovered decaying particles ultimately decay into these elementary particles.**

A decaying particle is composed of the particles produced in its decay. For example, the neutron  $n$  decays into a proton  $p$ , an electron  $e^-$ , and an electron antineutrino  $\bar{\nu}_e$ :

$$n \rightarrow p + e^- + \bar{\nu}_e.$$

Hence, we say that the neutron  $n$  is composed of a proton  $p$ , an electron  $e^-$ , and an electron antineutrino  $\bar{\nu}_e$ . The negatively charged muon  $\mu^-$  decays into an electron  $e^-$ , an electron antineutrino  $\bar{\nu}_e$ , and a muon neutrino  $\nu_\mu$ :

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu.$$

Thus, the negatively charged muon  $\mu^-$  is composed of an electron  $e^-$ , an electron antineutrino  $\bar{\nu}_e$ , and a muon neutrino  $\nu_\mu$ . The positively charged muon  $\mu^+$  decays into a positron  $e^+$ , an electron neutrino  $\nu_e$ , and a muon antineutrino  $\bar{\nu}_\mu$ :

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu.$$

Thus, the positively charged muon  $\mu^+$  is composed of a positron  $e^+$ , an electron neutrino  $\nu_e$ , and a muon antineutrino  $\bar{\nu}_\mu$ . The neutral pion  $\pi^0$  decays into two photons:

$$\pi^0 \rightarrow \gamma + \gamma.$$

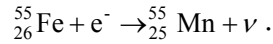
Hence, the neutral pion  $\pi^0$  is composed of two photons  $\gamma$ . Another neutral pion (often denoted  $\pi^0$ ) decays into a positron  $e^+$ , an electron  $e^-$ , and a photon  $\gamma$ :

$$\pi^0 \rightarrow e^+ + e^- + \gamma.$$

Thus, this neutral pion  $\pi^0$  is composed of a positron  $e^+$ , an electron  $e^-$ , and a photon  $\gamma$ . And so on. Although the decay products of these two neutral pions are distinctly different, they are conventionally regarded as the same particle. In contrast, the present theory holds the opposite view: because their decay products are clearly different, the theory considers them to be different particles. In other words, we should primarily distinguish particle types based on the particles produced when a decaying particle decays, rather than solely on properties such as mass, charge, etc., of the decaying particle.

The decay of a neutron into a proton, an electron, and an electron antineutrino is a deterministic event. Regardless of when or where it occurs, a neutron always decays into a proton, an electron, and an electron antineutrino. Therefore, we have reason to believe that a neutron is

composed of a proton, an electron, and an electron antineutrino. One might ask: if a neutron is indeed composed of a proton, an electron, and an electron antineutrino, then combining a proton, an electron, and an electron antineutrino should yield a neutron, correct? Indeed, a proton, an electron, and an electron antineutrino can combine to form a neutron. This is confirmed by the "orbital electron capture" experiment:



The nucleus  ${}_{26}^{55}\text{Fe}$  captures an orbital electron (primarily an electron from the  $K$ -shell because it is closest to the nucleus) and transforms into  ${}_{25}^{55}\text{Mn}$ . "Note: In this process, after an orbital electron is captured, the atomic number  $Z$  of the parent nucleus decreases by 1. This can be regarded as a proton combining with an electron to form a neutron, while simultaneously emitting a neutrino. The  $\beta$  decay energy  $E_{\beta m}$  is entirely carried away as the kinetic energy of the neutrino; in this example,  $E_{\beta m} = 0.231\text{MeV}$ . This neutrino cannot be detected directly in experiments. However, the Chinese physicist Wang Ganchang pointed out that when the neutrino is emitted, the nucleus (in the example above,  ${}_{25}^{55}\text{Mn}$ ) experiences a recoil, which can be measured experimentally. Later experiments indeed confirmed the existence of this recoil, thereby proving both the existence of the neutrino and the reality of the electron-capture process."<sup>1</sup> The neutrino produced here and another antineutrino are generated together; this antineutrino then becomes a constituent part of the neutron.

There are many objections to this hypothesis. Opponents argue that if a neutron were composed of a proton, an electron, and an electron antineutrino, wouldn't the neutron resemble a hydrogen atom, with the electron orbiting the proton inside the neutron? Those who hold this opposing view unconsciously assume that the way the proton and electron combine inside the neutron is the same as the way they combine in a hydrogen atom, so the neutron would behave like a hydrogen atom. This understanding is mechanical. The manner in which particles are composed is not limited to the way electrons and protons form a hydrogen atom; there should be other modes of composition. Just as the Earth is composed of animals, plants, mountains, subterranean rocks, etc., these constituents are clearly arranged in different ways: animals, especially birds, fly above the Earth—this is analogous to an electron orbiting a proton in a hydrogen atom; whereas mountains are firmly fixed on the Earth's surface and can only rotate with the Earth's rotation. The way that non-decaying particles combine to form decaying particles is similar to how mountains constitute the Earth.

A significant objection to regarding the neutron as composed of a proton, an electron, and an electron antineutrino is that this view violates the uncertainty principle. Opponents argue that if a neutron contains an electron, then atomic nuclei would also contain electrons. Since the radius of an atomic nucleus is about  $<10^{-12}\text{ cm}$ , the uncertainty in the position of an electron inside the nucleus would be on the order of  $10^{-12}\text{ cm}$ ,  $\Delta x \leq 10^{-12}\text{ cm}$ . According to the uncertainty relation, the uncertainty in the electron's momentum would be approximately  $\Delta p \approx \hbar / \Delta x \approx 10^{-15}\text{ g} \cdot \text{cm} \cdot \text{s}^{-1}$ . Thus, from an order-of-magnitude estimate, the momentum of an electron inside the nucleus would be  $p \approx \Delta p$ . From this, the energy of the electron inside the nucleus can be estimated as

$$E = \sqrt{p^2 c^2 + m^2 c^4} \approx pc \approx c\Delta p \approx \frac{\hbar c}{\Delta x} \approx 20\text{MeV}.$$

However, experimental observations show that the electron energies emitted in nuclear  $\beta$ -decay are on the order of a few  $1\text{MeV}$ , which is much smaller than the electron energy estimated from the uncertainty principle (roughly  $20\text{MeV}$ ).

Using the uncertainty relation to estimate the energy of an electron inside the nucleus rests on the premise that the electron in the nucleus is in motion, just as the electron in a hydrogen atom orbits the proton. If, on the contrary, one assumes that the electron in the neutron is not in motion but is tightly bound to the proton inside the neutron, then such an electron does not possess physical quantities like momentum  $\Delta p$  and position uncertainty  $\Delta x$ , and the uncertainty relation cannot be applied to estimate its energy. The present theory holds precisely this view. One

<sup>1</sup>Ni Guangjiong, Li Hongfang, *Modern Physics*[M]. Shanghai: Shanghai Science and Technology Press, 1979: 238.

might ask: Is it possible for a proton and an electron inside a neutron to be tightly bound together without orbiting each other? Does such a binding mode exist? And how can such a binding mode be described? To answer these questions, the following second assumption is required.

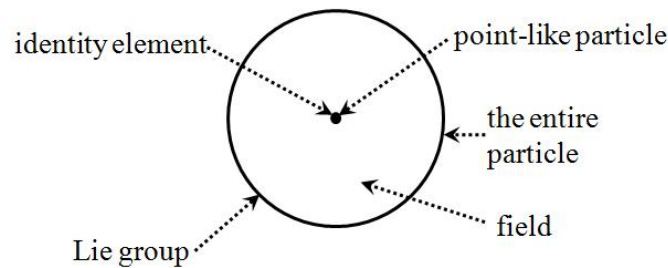
### §13.2 The Second Assumption

**Assumption Two:** Photons and antiphotons are each an  $SO(3)$  group; electrons, positrons, protons, antiprotons, neutrinos, and antineutrinos are each an  $SU(2)$  group; dark matter particles  $\chi_0, \chi_1, \chi_2$  and their antiparticles are each a  $L_+^\uparrow$  group.

The reason for assuming that photons and antiphotons are each an  $SO(3)$  group is that they are three-dimensional particles that generate electromagnetic fields and their spins are 1. The reason for assuming that electrons, positrons, protons, antiprotons, neutrinos, and antineutrinos are each an  $SU(2)$  group is that they are also three-dimensional particles and their spins are 1/2. The rationale for postulating the existence of dark matter particles and antiparticles and identifying them each as a  $L_+^\uparrow$  group (the proper orthochronous Lorentz group) will be presented in the final part of this book.

Both  $SO(3)$  and  $SU(2)$  are Lie groups, which are composites of group structures and smooth manifolds. In this book, the terms “ $SO(3)$  group” and “ $SU(2)$  group” sometimes refer to the matrix groups isomorphic to  $SO(3)$  and  $SU(2)$ , respectively, and sometimes refer to the abstract Lie groups isomorphic to  $SO(3)$  and  $SU(2)$ .

The term “particle” can be understood in both a broad and a narrow sense. In the narrow sense, a particle refers to its extremely tiny core portion; however, this core is surrounded by a field that is inseparable from the particle. In this book, the term “**point-like particle**” is used to denote the extremely tiny core portion of a particle—the part that participates in collisions with other particles. The term “**particle**,” on the other hand, refers to the entity as a whole, encompassing both the point-like core and the surrounding field space.



**Figure 13.2.1** Structure of a particle

Assuming that photons and antiphotons are each an  $SO(3)$  group means that the entire entity of each photon is an  $SO(3)$  group. The point-like photon is represented by the identity element  $e$  of the  $SO(3)$  group, while the rest of the  $SO(3)$  group—other than the identity element—constitutes the field-space part of the photon, as shown in Figure 13.2.1. Similarly, assuming that electrons, positrons, protons, antiprotons, neutrinos, and antineutrinos are each an  $SU(2)$  group means that the entire entity of each such particle is an  $SU(2)$  group. The point-like particle is represented by the identity element  $e$  of the  $SU(2)$  group. The portion of the  $SU(2)$  group other than the identity element constitutes the field-space part of these particles. Likewise, assuming that elementary dark matter particles and their antiparticles are each a  $L_+^\uparrow$  group means that the entire entity of each such particle is a  $L_+^\uparrow$  group. The point-like particle is represented by the identity element  $e$  of the  $L_+^\uparrow$  group. The portion of the  $L_+^\uparrow$  group other than the identity element constitutes the field-space part of these particles.

# Chapter 14 Distinguishing Particles and Antiparticles by Orientation

## §14.1 Distinguishing Particles and Antiparticles Using the Orientation of Lie Groups

According to Theorem 12.7.1, a connected Lie group is orientable. Since the Lie groups  $SO(3)$  and  $SU(2)$  are both connected, they are orientable. By Theorems 12.7.2 and 12.7.3, a connected Lie group possesses exactly two orientations. Consequently, as connected groups,  $SO(3)$  and  $SU(2)$  each have two orientations, corresponding to two distinct types of particles—particles and antiparticles.

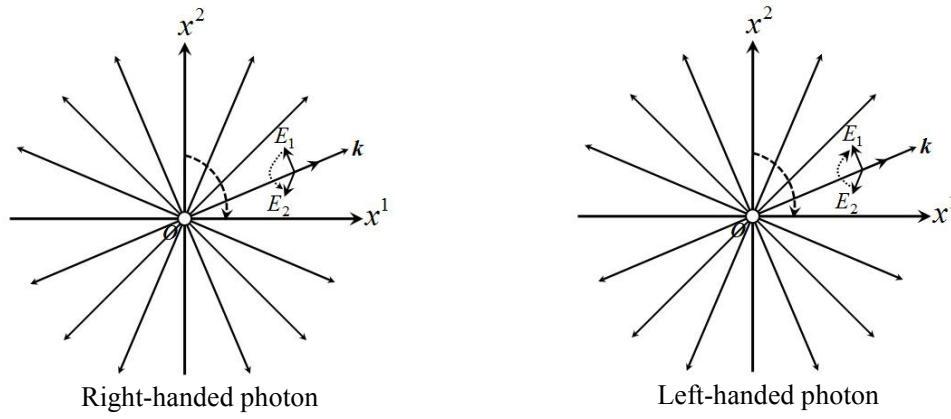


Figure 14.1.1 Positively oriented photon

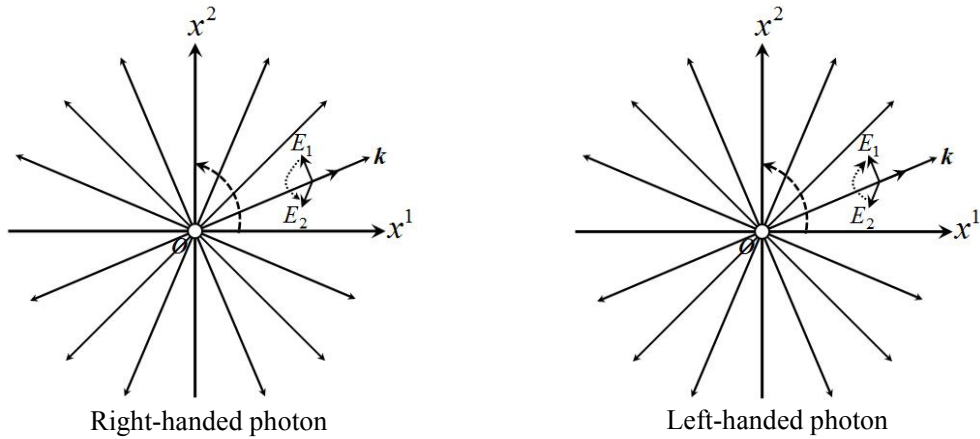


Figure 14.1.2 Negatively oriented photon

A photon, being an  $SO(3)$  group, therefore has two orientations and thus two kinds: one with a fixed left-handed orientation (positive orientation), as shown in Figure 14.1.1; the other with a fixed right-handed orientation (negative orientation), as shown in Figure 14.1.2. In the real world, photons with left-handed orientation outnumber those with right-handed orientation; the reason for this will be discussed in §21.10.

As shown in Figure 14.1.3, for a photon, if its spin direction  $\mathbf{S}$  is the same as its propagation direction  $\mathbf{k}$ , it is conventionally called a **right-handed photon**. If its spin direction  $\mathbf{S}$  is opposite to

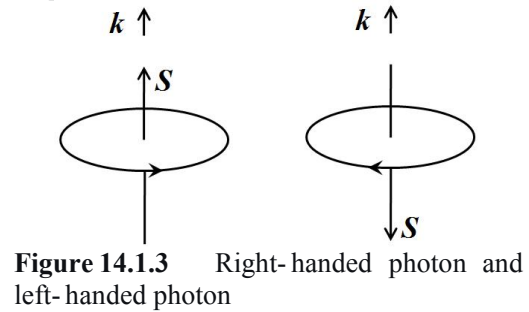
its propagation direction  $\mathbf{k}$ , it is conventionally called a **left-handed photon**. Both photons and antiphotons can be further divided into right-handed photons and left-handed photons, as illustrated in Figures 14.1.1 and 14.1.2. Therefore, a positively oriented photon can be either right-handed or left-handed; when measuring its spin angular momentum, it is possible to obtain either  $+\hbar$  or  $-\hbar$ . Similarly, a negatively oriented photon can also be either right-handed or left-handed; when measuring its spin angular momentum, it is also possible to obtain either  $+\hbar$  or  $-\hbar$ .

The  $SU(2)$  group also has two orientations, thereby giving rise to protons and antiprotons, electrons and positrons, and neutrinos and antineutrinos.

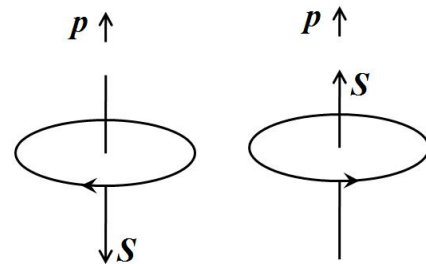
For electrons and positrons, it is stipulated that the electron adopts the orientation consistent with the left-handed system; this orientation is called the positive orientation. The positron adopts the orientation consistent with the right-handed system; this orientation is called the negative orientation. Because the orientations of electrons and positrons are opposite, their electric charges are also opposite; see Chapter 18 for details.

For protons and antiprotons, it is stipulated that the proton adopts the orientation consistent with the left-handed system (positive orientation). The antiproton adopts the orientation consistent with the right-handed system (negative orientation). Since the orientations of protons and antiprotons are opposite, their electric charges are also opposite; the reason will be explained in Chapter 18. As for why electrons and protons, having the same orientation, exhibit opposite charges, the explanation will be given in Chapters 18 and 20.

For neutrinos and antineutrinos, it is stipulated that the neutrino adopts the left-handed orientation (positive orientation); its spin along the  $z$ -axis is opposite to its direction of motion, making it left-handed with spin  $-\hbar/2$ , as shown in the left panel of Figure 14.1.4. The antineutrino adopts the right-handed orientation (negative orientation); its spin along the  $z$ -axis coincides with its direction of motion, making it right-handed with spin  $+\hbar/2$ , as shown in the right panel of Figure 14.1.4.



**Figure 14.1.3** Right-handed photon and left-handed photon



**Figure 14.1.4** Left-handed neutrino and right-handed antineutrino

## §14.2 Particles and Antiparticles Can Mutually Transform

Particles and antiparticles are not fixed entities; under certain conditions they can transform into each other. Changing the orientation of a particle can turn a particle into its antiparticle, or an antiparticle into its particle. For example, let  $p$  denote a proton and  $M$  denote the transformed particle, which could be either a proton or an antiproton  $\bar{p}$ . Suppose there exists an isomorphism  $f$  between  $p$  and  $M$ :

$$f : p \rightarrow M.$$

For the mapping  $f$  to change a proton into an antiproton, the crucial step is to reverse the orientation of the proton. Assume that at any point  $q \in p$  on the proton, the basis  $\{e_i\}$  of the tangent space  $T_q p$  is compatible with the orientation of the proton. Then the mapping  $f$  induces a tangent map  $f_*$ , whose image  $\{f_*(e_i)\}$  becomes a basis of the tangent space  $T_{f(q)} M$ , determining the orientation of  $T_{f(q)} M$ . According to Theorem 9.4.1, if the Jacobian matrix  $J$  of the mapping  $f$  has determinant  $|J| > 0$ , then the orientation of  $T_{f(q)} M$  determined by  $\{f_*(e_i)\}$  is the same as the original orientation (the orientation of the basis  $\{e_i\}$ ); the mapping is said to be **orientation-preserving**, the proton remains a proton, and  $M$  is still the proton  $p$ . However, if the

Jacobian matrix  $J$  has determinant  $|J| < 0$ , then the orientation of  $T_{f(q)}M$  determined by  $\{f_*(e_i)\}$  is opposite to the original orientation (the orientation of the basis  $\{e_i\}$ ), and consequently the proton is transformed into an antiproton; that is,  $M$  becomes the antiproton  $\bar{p}$ .

However, simply converting a proton into an antiproton would violate the law of charge conservation. Therefore, another particle with opposite charge must simultaneously participate in the transformation. For example, when the mapping  $f$  converts a proton  $p$  into an antiproton  $\bar{p}$ , another mapping  $g$  simultaneously converts an electron  $e^-$  into a positron  $e^+$ :

Mapping  $f : p \rightarrow \bar{p}$  causes the proton  $p$  to transform into an antiproton  $\bar{p}$ ;

Mapping  $g : e^- \rightarrow e^+$  causes the electron  $e^-$  to transform into a positron  $e^+$ .

In this way, the total charge remains unchanged before and after the mappings.

The conversion between particles and antiparticles can also be described from another perspective. For example, let  $p$  denote a proton and  $M$  denote either a proton or an antiproton  $\bar{p}$ , and suppose there exists an isomorphism  $f$  between  $p$  and  $M$ :

$$f : p \rightarrow M.$$

Assume the orientation of particle  $M$  is determined by a 3-form  $\omega$ . The mapping  $f$  also induces a cotangent map  $f^*$ . The image of  $\omega$  under  $f^*$ , i.e.,  $f^*\omega$ , is a 3-form on the proton. According to Theorem 9.4.2,  $f^*\omega$  will redefine the orientation of the proton; this is called the **orientation pulled back** by the mapping  $f$ . If the Jacobian matrix of  $f$  has determinant  $|J| > 0$ , then the orientation of the proton redefined by  $f^*\omega$  coincides with the orientation determined by  $\omega$ ; the proton does not change into its antiparticle, and  $M$  remains the proton  $p$ .

However, if the Jacobian matrix of  $f$  has determinant  $|J| < 0$ , then the orientation of the proton redefined by  $f^*\omega$  is opposite to that determined by  $\omega$ . Therefore, since  $p$  is a proton,  $M$  becomes the antiproton  $\bar{p}$ . Again, simply turning a proton into an antiproton would violate charge conservation, so another particle with opposite charge must simultaneously participate in the transformation.

Nevertheless, the conversion from a right-handed-oriented particle to a left-handed-oriented particle is relatively easier than the conversion from a left-handed-oriented particle to a right-handed-oriented particle. The reason for this will be discussed in §21.10.



## Chapter 15 Particles in Minkowski Spacetime

We can study the properties of particles from three different perspectives: **First**, we examine the properties of a particle as a Lie group. Since a Lie group is also a smooth manifold, the particle itself is a smooth manifold, and we can investigate the intrinsic properties that the particle possesses as a smooth manifold. **Second**, we embed the particle into four-dimensional Euclidean space  $R^4$ , where it becomes a hypersurface in  $R^4$ , and study the properties it exhibits as a hypersurface within  $R^4$ . **Third**, let the coordinates of  $R^4$  be  $(x^0, x^1, x^2, x^3)$ . We perform a coordinate transformation:

$$x^0 = ict, \quad x^1 = x^1, \quad x^2 = x^2, \quad x^3 = x^3,$$

where  $t$  is time,  $i$  is the imaginary unit, and  $c$  is the speed of light. This transforms four-dimensional Euclidean space  $R^4$  into four-dimensional Minkowski spacetime (hereafter abbreviated as **Minkowski spacetime** or **Minkowski space**)  $M^4$ . We then study the properties that the particle possesses as a submanifold embedded in  $M^4$ . This chapter adopts the third approach to discuss several important properties of particles when they are embedded in Minkowski spacetime.

The foundation of special relativity rests on two fundamental postulates: first, the principle of the constancy of the speed of light; second, the principle of relativity. The principle of relativity is generalized in general relativity to the principle of general covariance. These two principles require that physical laws be described by tensors. In this chapter, we can deduce the constancy of the speed of light from the topological properties of photons. On the other hand, we have already regarded particles as Lie groups, and in Lie group theory, tensors are used to study the properties of Lie groups. Therefore, the requirements of the principle of relativity and the principle of general covariance are naturally fulfilled within the framework of Lie group theory. Consequently, we need not separately postulate the validity of the principle of relativity and the principle of general covariance.

### §15.1 Particles Are All Three-Dimensional

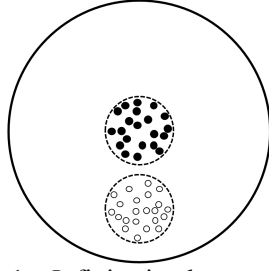
Two or more Lie groups can be combined by the **internal direct product** operation. The definition of an internal direct product group is given in §12.12. An internal direct product group can describe the linear superposition of two or more particles; in other words, when two or more particles combine, they exist in the form of an internal direct product group.

#### 1. $SU(2) \otimes SU(2)'$

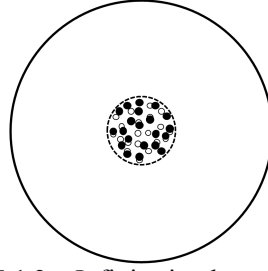
Every elementary particle with spin-1/2 is an  $SU(2)$  group. Consequently, the point elements of these particles are in one-to-one correspondence. Consider two elementary particles with spin-1/2, whose Lie groups are an  $SU(2)$  group and an  $SU(2)'$  group. Let the corresponding group elements of the two groups be paired: the identity element  $e$  of the  $SU(2)$  group and the identity element  $e'$  of the  $SU(2)'$  group form a pair  $(e, e')$ , which becomes the identity element of the internal direct product group  $SU(2) \otimes SU(2)'$ . Let  $a_i \in SU(2)$ . The element  $a_i$  is the same with the element  $a'_i$  in  $SU(2)'$  group, forms a pair of elements  $(a_i, a'_i)$ . The inverse of  $(a_i, a'_i)$  must then be the pair  $(a_i^{-1}, a'^{-1}_i)$  so that the set of pairs constitutes a group.

The two identity elements form a pair  $(e, e')$ , which does **not** require that the points represented by these two identity elements coincide spatially. Similarly, when an element  $a_i$  of the  $SU(2)$  group and an element  $a'_i$  of the  $SU(2)'$  group form a pair  $(a_i, a'_i)$ , there is no requirement that the points represented by these two elements overlap; and so on. Two particles can combine in two ways. **First**, their infinitesimal groups (or local Lie groups) may be disjoint, as illustrated in Figure 15.1.1. **Second**, their infinitesimal groups may overlap partially or completely,

as shown in Figure 15.1.2.



**Figure 15.1.1** Infinitesimal groups are disjoint



**Figure 15.1.2** Infinitesimal groups overlap

Define the multiplication operation as follows:

$$(a_i, a'_i)(b_j, b'_j) = (a_i b_j, a'_i b'_j), \quad (15.1.1)$$

Then the set  $SU(2) \otimes SU(2)'$  of pairs with operation (15.1.1) forms a group, called the **internal direct product group** of the  $SU(2)$  group and the  $SU(2)'$  group.

The internal direct product group  $SU(2) \otimes SU(2)'$  is isomorphic to either the  $SU(2)$  group or the  $SU(2)'$  group. Consequently, all three groups have the same dimension; they are all three-dimensional Lie groups.

### 2. $SO(3) \otimes SO(3)'$

Similarly, a photon (including both photons and antiphotons) is an  $SO(3)$  group. Two or more photons can also undergo an internal direct product operation. The internal direct product group  $SO(3) \otimes SO(3)'$  is isomorphic to either the  $SO(3)$  group or the  $SO(3)'$  group. Hence, all three groups share the same dimension—they are all three-dimensional Lie groups.

### 3. $SO(3) \otimes SU(2)'$

Let  $u_i$  be any element of the  $SU(2)$  group,  $u_i \in SU(2)$ , and let  $E$  be the identity element of the  $SU(2)$  group,  $E \in SU(2)$ . Then both  $Eu_i$  and  $-Eu_i$  are elements of  $SU(2)$ . If we identify  $Eu_i$  and  $-Eu_i$  as the same element, then the group  $SU(2)$  reduces to the  $SO(3)$  group. Alternatively, we may regard any element of the  $SO(3)$  group as consisting of two elements  $u_i$  and  $-u_i$  of the  $SU(2)$  group, with  $u_i = -u_i$ . Therefore, there is no one-to-one correspondence between the  $SO(3)$  group and the  $SU(2)$  group; instead, there is a **one-to-two** relation: one element of the  $SO(3)$  group corresponds to two elements of the  $SU(2)$  group.

A photon is an  $SO(3)$  group, whereas the Lie group representing an elementary particle with spin-1/2 is an  $SU(2)$  group. Since there is no one-to-one correspondence between the  $SO(3)$  group and the  $SU(2)$  group but rather a one-to-two relation, how can these two types of groups be combined via an internal direct product operation?

Let  $o_i$  be an element of the  $SO(3)$  group,  $o_i \in SO(3)$ . This  $o_i$  corresponds to two elements  $(Eu_i, -Eu_i)$  of the  $SU(2)$  group. We pair  $o_i$  with  $(Eu_i, -Eu_i)$  to form the element  $(o_i, (Eu_i, -Eu_i))$ , i.e.,  $o_i$  is paired with  $Eu_i$  and also with  $-Eu_i$  to yield the elements  $(o_i, Eu_i)$  and  $(o_i, -Eu_i)$ . Here,  $(Eu_i, -Eu_i)$  represent **two distinct elements**  $Eu_i$  and  $-Eu_i$ , not that these two elements are identified as a single element. Similarly,  $(o_i, (Eu_i, -Eu_i))$  denote two independent elements  $(o_i, Eu_i)$  and  $(o_i, -Eu_i)$ , not that  $o_i$  and  $(Eu_i, -Eu_i)$  are merged into one element.

Construct the direct product set of the two sets  $SO(3)$  and  $SU(2)'$ :

$$SO(3) \otimes SU(2) = \{(o_i, (Eu_i, -Eu_i)) \mid o_i \in SO(3); Eu_i, -Eu_i \in SU(2)\}.$$

Let  $o_j \in SO(3)$ .  $o_j$  corresponds to two elements  $(Eu_j, -Eu_j) \in SU(2)$ . Pair  $o_j$  with  $(Eu_j, -Eu_j)$  to form the element  $(o_j, (Eu_j, -Eu_j))$ ; then

$$(o_j, (Eu_j, -Eu_j)) \in SO(3) \otimes SU(2).$$



Define the **internal** direct product by formula (15.1.1):

$$(o_i, (Eu_i; -Eu_i))(o_j, (Eu_j; -Eu_j)) = (o_i o_j, (Eu_i u_j; -Eu_i u_j)). \quad (15.1.2)$$

It can be verified that  $SO(3) \otimes SU(2)$  with operation (15.1.2) forms a group, called the **internal direct product group** of  $SO(3)$  and  $SU(2)$ . The verification proceeds as follows:

**1)Closure:** Let  $(o_i, (Eu_i; -Eu_i)) \in SO(3) \otimes SU(2)$ ,  $(o_j, (Eu_j; -Eu_j)) \in SO(3) \otimes SU(2)$ . Then

$$(o_i, (Eu_i; -Eu_i))(o_j, (Eu_j; -Eu_j)) = (o_i o_j, (Eu_i u_j; -Eu_i u_j)).$$

Since  $o_i o_j \in SO(3)$ ,  $(Eu_i u_j; -Eu_i u_j) \in SU(2)$ , we have

$$(o_i o_j, (Eu_i u_j; -Eu_i u_j)) \in SO(3) \otimes SU(2),$$

Thus, multiplication (15.1.2) is closed.

**2)Associativity:** Let  $(o_k, (Eu_k; -Eu_k)) \in SO(3) \otimes SU(2)$ . Then

$$\begin{aligned} & ((o_i, (Eu_i; -Eu_i))(o_j, (Eu_j; -Eu_j)))(o_k, (Eu_k; -Eu_k)) \\ &= (o_i o_j, (Eu_i u_j; -Eu_i u_j))(o_k, (Eu_k; -Eu_k)) \\ &= (o_i o_j o_k, (Eu_i u_j u_k; -Eu_i u_j u_k)), \\ & (o_i, (Eu_i; -Eu_i))((o_j, (Eu_j; -Eu_j))(o_k, (Eu_k; -Eu_k))) \\ &= (o_i, (Eu_i; -Eu_i))(o_j o_k, (Eu_j u_k; -Eu_j u_k)) \\ &= (o_i o_j o_k, (Eu_i u_j u_k; -Eu_i u_j u_k)), \end{aligned}$$

so associativity holds.

**3)Identity element:** Let  $e \in SO(3)$  be the identity element of  $SO(3)$ ;  $e$  corresponds to the two elements  $(EE; -EE) \in SU(2)$ . Therefore, the identity element of  $SO(3) \otimes SU(2)$  is  $(e, (EE; -EE)) = (e, (E; -E))$ .

**4)Inverse element:** Let  $o_i^{-1} \in SO(3)$  and let  $o_i^{-1}$  be the inverse of  $o_i$  in  $SO(3)$ ; then  $o_i^{-1}$  corresponds to two elements  $(Eu_i^{-1}; -Eu_i^{-1}) \in SU(2)$ . Since

$$(o_i, (Eu_i; -Eu_i))(o_i^{-1}, (Eu_i^{-1}; -Eu_i^{-1})) = (o_i o_i^{-1}, (Eu_i u_i^{-1}; -Eu_i u_i^{-1})) = (e, (EE; -EE)) = (e, (E; -E)),$$

every element  $(o_i, (Eu_i; -Eu_i)) \in SO(3) \otimes SU(2)$  has an inverse  $(o_i^{-1}, (Eu_i^{-1}; -Eu_i^{-1}))$ .

Thus, the set  $SO(3) \otimes SU(2)$  together with the multiplication operation (15.1.2) constitutes a group.

$(o_i, (Eu_i; -Eu_i))$  represent two independent elements  $(o_i, Eu_i)$  and  $(o_i, -Eu_i)$ . One element  $o_i$  of the  $SO(3)$  group corresponds to two distinct elements  $(o_i, Eu_i)$  and  $(o_i, -Eu_i)$  of the  $SO(3) \otimes SU(2)$  group, so the  $SO(3)$  group and the  $SO(3) \otimes SU(2)$  group are **not isomorphic**. However, the  $SU(2)$  group is isomorphic to the  $SO(3) \otimes SU(2)$  group, because there is a one-to-one correspondence between an element  $(o_i, Eu_i)$  of  $SO(3) \otimes SU(2)$  and the element  $Eu_i$  of  $SU(2)$ . Consequently, the dimension of the group  $SO(3) \otimes SU(2)$  is the same as that of the  $SU(2)$  group—it remains three-dimensional. Moreover, when a photon (an  $SO(3)$  group) merges with an elementary particle of spin-1/2 (an  $SU(2)$  group), the result is still an elementary particle of spin-1/2. This explains why an electron remains an electron after absorbing a photon. Of course, according to the law of energy conservation, the electron gains energy upon absorbing the photon, but it is still an electron.

The above conclusion can be extended to new particles formed by combining more than two elementary particles. All such new particles are three-dimensional; by extension, **all particles are three-dimensional**, that is, all physical objects are three-dimensional.

## §15.2 Particles as Submanifolds

### 1. Submanifolds in Four-Dimensional Euclidean Space

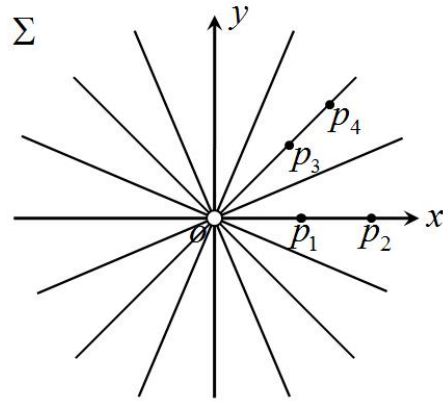
An elementary particle with spin-1/2 is diffeomorphic to a three-dimensional sphere  $S^3$ , which can only be embedded in four-dimensional Euclidean space  $R^4$  as a submanifold. Similarly, a photon is a three-dimensional real projective space  $RP^3$ , which also can only be embedded in  $R^4$  as a submanifold.

In  $R^4$ , there are various models for  $RP^3$ . One of them regards  $RP^3$  as a smooth manifold consisting of infinitely many straight lines passing through the origin, but the origin itself must be removed. From our perspective in Euclidean space, each such line appears as a straight line; however, for  $RP^3$ , the entire line is regarded as a single point—that is, all points on the same line are identified as one point. As shown in Figure 15.2.1 (illustrated in two dimensions for clarity), in the Cartesian coordinate system  $\Sigma$ , to describe  $RP^3$  (which represents a photon in this simplified view), we must regard the points  $p_1$  and  $p_2$  on the same radial line as the same point. Therefore, the Euclidean squared distance  $\Delta s^2$  between these two points is zero:

$$\Delta s^2 = 0. \quad (15.2.1)$$

However, for four-dimensional Euclidean space  $R^4$ , the points  $p_1$  and  $p_2$  are two distinct points, so

$$\Delta s^2 > 0. \quad (15.2.2)$$



**Figure 15.2.1** A photon as a real projective space

### 2. Submanifolds in Minkowski Space

Let us briefly review the history of humanity's understanding of numbers. Ancient humans, driven by the need for counting—such as tallying the number of animals hunted in a day—developed the concept of enumeration, giving rise to the natural numbers. For the distribution of resources like fruits and meat, fractions were invented. Later, basic arithmetic operations such as addition and subtraction emerged. To ensure that subtraction could always be performed, negative numbers and the number zero were introduced. Integers and fractions together are called rational numbers. Subsequently, when calculating numbers like  $\sqrt{2}$ , it became impossible to express their exact values with existing rational numbers, leading to the introduction of irrational numbers. Rational and irrational numbers are collectively known as real numbers. Later, while solving equations, the problem of taking square roots of negative numbers was encountered, giving rise to the imaginary unit  $i$ , with  $\sqrt{-1} = i$ . Real and imaginary numbers together are called complex numbers. Thus, only within the domain of complex numbers can operations such as addition, subtraction, multiplication, division, squaring, and taking square roots be carried out without restriction. Consequently, to describe the properties of particles, we must also employ complex numbers; otherwise, we might face situations where taking the square root of a negative number is unavoidable—that is, we cannot guarantee that describing the properties of particles will never require taking the square root of a negative number.

As we know, a complex number is expressed as  $a + bi$ , where  $a$  and  $b$  are real numbers. Complex numbers can be represented on a complex plane, with the real part on the real axis and the imaginary part on the imaginary axis. On the complex plane, the distance between two points  $p_1(a, ib)$  and  $p_2(c, id)$  should be defined as

$$d(p_1, p_2) = \sqrt{(b-a)^2 + (di-bi)^2} = \sqrt{(b-a)^2 - (d-b)^2}.$$

Four-dimensional Euclidean space  $R^4$  has four spatial coordinates, whereas the actual

Euclidean space is three-dimensional, with only three spatial coordinates. We can combine Euclidean space  $R^4$  with the complex plane by setting the imaginary axis of the complex plane as  $ict$ , where  $t$  denotes time and  $c$  is the speed of light, and taking the real axis of the complex plane as the radial coordinate  $r$ , with  $r = (x, y, z)$ . That is, we establish a Cartesian coordinate system  $(x, y, z, ict)$  in  $R^4$ . The space  $R^4$  endowed with such a coordinate system is also called **Minkowski spacetime**  $M^4$ . The spaces  $R^4$  and  $M^4$  are diffeomorphic; they merely differ in the choice of coordinates.

### 3.The Speed of Light is the Maximum Speed

As shown in Figure 15.2.1, let the coordinates of two points  $p_1$  and  $p_2$  in  $M^4$  be  $(x^1, y^1, z^1, ict^1)$  and  $(x^2, y^2, z^2, ict^2)$ , respectively. Then from equation (15.2.1) we obtain

$$\Delta s^2 = (ct^2 - ct^1)^2 - (x^2 - x^1)^2 - (y^2 - y^1)^2 - (z^2 - z^1)^2 = 0. \quad (15.2.3)$$

For humans living in three-dimensional Euclidean space, the points  $p_1$  and  $p_2$  are two distinct points; moving from  $p_1$  to  $p_2$  would require a finite amount of time. However, for  $RP^3$  (representing the photon) no time is needed, because  $p_1$  and  $p_2$  are the same point. That is, a point-like photon can move from  $p_1$  to  $p_2$  without expending any effort—this manifests as the photon having zero rest mass. Although for  $RP^3$  the points  $p_1$  and  $p_2$  are identical and no time is required to go from one to the other, for observers living in Euclidean space  $R^3$  it still appears to take time, because we perceive  $p_1$  and  $p_2$  as two different points. According to equation (15.2.3), under the condition of the speed of light  $c$ , we would say that a point-like photon moving from  $p_1$  to  $p_2$  requires a time interval  $t^2 - t^1$ . Neutrinos and antineutrinos possess similar properties to photons; see Chapter 18 for details.

From equation (15.2.2) we obtain

$$\Delta s^2 = (ct^2 - ct^1)^2 - (x^2 - x^1)^2 - (y^2 - y^1)^2 - (z^2 - z^1)^2 > 0, \quad (15.2.4)$$

i.e.,

$$(ct^2 - ct^1)^2 > (x^2 - x^1)^2 + (y^2 - y^1)^2 + (z^2 - z^1)^2.$$

Electrons, positrons, protons, and antiprotons are all homeomorphic to the three-dimensional unit sphere  $S^3$ . Like Euclidean space, any two points in  $S^3$ , such as  $p_1$  and  $p_2$ , are distinct points. Since  $S^3$  is embedded in  $M^4$ , the points  $p_1$  and  $p_2$  exist both in  $S^3$  and in  $M^4$ . For point-like particles of these types, moving from  $p_1$  to  $p_2$  requires a finite amount of time, which manifests as these particles having nonzero rest mass. Suppose the speed of such a point-like particle is  $v$ . Then the squared distance traveled from time  $t^1$  to  $t^2$  is

$$(x^2 - x^1)^2 + (y^2 - y^1)^2 + (z^2 - z^1)^2 = v^2(t^2 - t^1)^2 < c^2(t^2 - t^1)^2.$$

Therefore, requiring inequality (15.2.4) to hold is equivalent to demanding that point-like particles with nonzero rest mass cannot move faster than the speed of light.

For the space  $R^4$ , the metric form is

$$ds^2 = dx^2 + dy^2 + dz^2 + dw^2, \quad (15.2.5)$$

where the coordinates  $(x, y, z, w)$  are all spatial coordinates. If we set

$$x = x, \quad y = y, \quad z = z, \quad w = ict, \quad (15.2.6)$$

with  $t$  representing time and  $i$  the imaginary unit, then equation (15.2.5) becomes

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2. \quad (15.2.7)$$

Equation (15.2.7) is the metric form of the four-dimensional spacetime  $M^4$ . The coordinate transformation (15.2.6) converts the space  $R^4$  into the spacetime  $M^4$ .

In the space  $R^4$ ,  $ds^2 \geq 0$ , so  $R^4$  is a **positive-definite Riemannian manifold**. However, in the spacetime  $M^4$ ,  $ds^2$  is indefinite; hence  $M^4$  is called a **pseudo-Riemannian** (or

**generalized Riemannian**) manifold.

The same smooth manifold can be endowed with different metric tensor fields.  $R^4$  and  $M^4$  provide an excellent example: as smooth manifolds they are diffeomorphic and identical, yet we can equip this smooth manifold with two distinct metric tensor fields, making it either  $R^4$  or  $M^4$ . Because  $R^4$  is positive-definite, it is sometimes more convenient to analyze problems within  $R^4$ . Therefore, for convenience, we may place the objects of analysis in  $R^4$ —for instance, analyzing photons ( $RP^3$ ) and electrons ( $S^3$ ) in  $R^4$ —and finally apply the coordinate transformation (15.2.6) to translate the results into conclusions valid in  $M^4$ .

Since  $SU(2)$  and  $S^3$  are homeomorphic, and  $SO(3)$  and  $RP^3$  are homeomorphic, and both  $S^3$  and  $RP^3$  can be embedded in either  $R^4$  or  $M^4$ , it follows that **all elementary particles and the particles composed of them can be embedded in  $R^4$  or  $M^4$**  as submanifolds. Consequently, they inherit an induced metric tensor from  $R^4$  or  $M^4$  and thus become Riemannian or pseudo-Riemannian manifolds.

### §15.3 Why the Speed of Light is Constant

A photon is a real projective space  $RP^3$ . As shown in Figure 15.2.1, if all rays in the coordinate system  $\Sigma$  are shifted to the right, the entire photon appears to move to the right. In this coordinate system, suppose the point-like photon is located at point  $p_1 (x^1, y^1, z^1, ict^1)$  at time  $t^1$  and moves toward point  $p_2 (x^2, y^2, z^2, ict^2)$ , reaching  $p_2$  at time  $t^2$ . Then

$$\Delta s^2 = (ct^2 - ct^1)^2 - (x^2 - x^1)^2 - (y^2 - y^1)^2 - (z^2 - z^1)^2 = 0. \quad (15.3.1)$$

This is equivalent to having an observer stationary at point  $p_2$  who, at time  $t^2$ , receives a light signal emitted by a source at point  $p_1$  at time  $t^1$  and measures the propagation speed of this light signal to be  $c$ .

If at time  $t^2$  the point  $p_2$  moves from right to left along the horizontal axis and arrives at point  $p_1$  at time  $t^1$ , we again obtain equation (15.3.1). This is equivalent to an observer who, at time  $t^2$ , starts from point  $p_2$  and moves from right to left toward the light source, arriving at point  $p_1$  at time  $t^1$ , and measuring the propagation speed of this light signal to be  $c$ .

If at time  $t^1$  the point  $p_2$  coincides with point  $p_1$  and moves to the right, arriving at point  $p_2$  at time  $t^2$ , we again obtain equation (15.3.1). This is equivalent to an observer who, at time  $t^1$ , starts from point  $p_1$  on the horizontal axis and moves left-to-right away from the light source, reaching point  $p_2$  at time  $t^2$ , and measuring the propagation speed of this light signal to be  $c$ .

Therefore, regardless of whether the observer is stationary, moving toward the photon, or moving away from the photon, the measured propagation speed of the photon is always equal to  $c$ . Since motion is relative, if the observer moves toward the photon, this is equivalent to the observer being stationary and the light source moving toward the observer. Similarly, if the observer moves away from the photon, it is equivalent to the observer being stationary and the light source moving away from the observer.

If we instead analyze the situation at points  $p_3$  and  $p_4$ , the same conclusion can be drawn. This demonstrates that the propagation speed of the photon is the same along all radial directions—it is equal to  $c$ —and indicates that the speed of light is isotropic.

If the points  $p_1$  and  $p_2$ , the points  $p_3$  and  $p_4$  are sufficiently close, with distance  $\Delta s^2 \rightarrow ds^2$ , the same conclusion holds. Therefore, integrating the above analysis, we can derive the **principle of the constancy of the speed of light** proposed by Einstein:

Regardless of whether the light source is stationary or in motion, light propagates in a vacuum with a definite speed  $c$  in a "stationary" coordinate system, and the magnitude of this speed is the same in all directions.

## §15.4 Lorentz Transformations

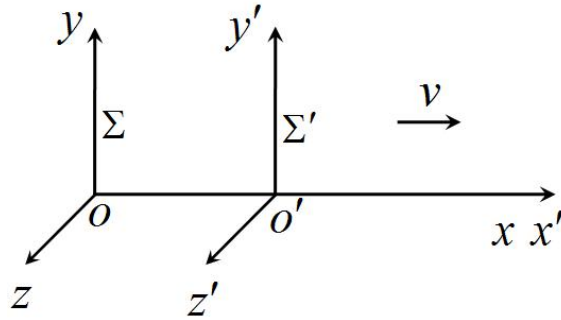
### 1. Lorentz Transformations

Let us first derive the Lorentz transformation. Minkowski spacetime  $M^4$  is flat. We set up a Cartesian coordinate system  $\Sigma$  in this spacetime, with four coordinates  $(x, y, z, ict)$ ; another Cartesian system  $\Sigma'$  has coordinates  $(x', y', z', ict')$ . The coordinates  $x, y, z, t$  and  $x', y', z', t'$  are all real numbers. Since  $M^4$  is flat,  $x, y, z, t$  and  $x', y', z', t'$  are all measurable physical quantities.

If the two coordinate systems are in uniform relative motion—i.e., they are related by a translation—then both are inertial frames. Because the transformation is a translation, it must be linear. Assume the translation takes the system  $\Sigma$  to the system  $\Sigma'$ ; then we have

$$\begin{cases} x' = a_{11}x + a_{12}y + a_{13}z + a_{14}t, \\ y' = a_{21}x + a_{22}y + a_{23}z + a_{24}t, \\ z' = a_{31}x + a_{32}y + a_{33}z + a_{34}t, \\ t' = a_{41}x + a_{42}y + a_{43}z + a_{44}t. \end{cases} \quad (15.4.1)$$

We consider a special translation case to derive the explicit Lorentz transformation. Suppose initially the two systems are aligned: the  $x$ -axis coincides with the  $x'$ -axis, the  $y$ -axis with the  $y'$ -axis, the  $z$ -axis with the  $z'$ -axis, and their origins  $O$  and  $O'$  coincide. The directions of the three axes are the same, as shown in Figure 15.4.1. When the origins of the two coordinate systems coincide, their coordinates are both  $(0,0,0,0)$ . Therefore, equation (15.4.1) becomes



**Figure 15.4.1** The  $\Sigma'$ -system moves with uniform velocity  $v$  relative to the  $\Sigma$ -system

$$\begin{cases} x' = a_{11}x + a_{14}t, \\ y' = y, \\ z' = z, \\ t' = a_{41}x + a_{44}t. \end{cases} \quad (15.4.2)$$

A translation transformation must preserve orientation; therefore, its Jacobian determinant must be greater than zero (see Definition 9.1.1), i.e.,

$$\begin{vmatrix} a_{11} & 0 & 0 & a_{14} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a_{41} & 0 & 0 & a_{44} \end{vmatrix} = a_{11}a_{44} > 0.$$

Suppose the transformation (15.4.2) is one-dimensional:

$$x' = a_{11}x,$$

then we obtain  $|a_{11}| = a_{11} > 0$ . Thus, to preserve orientation, we must have  $a_{11} > 0, a_{44} > 0$ .

Assume that when the origins of the two coordinate systems coincide, a light signal is emitted from the origin  $O$  and timing begins simultaneously. After some time, let an object receive the light signal with coordinates  $(x, y, z, ict)$  in the  $\Sigma$  system and  $(x', y', z', ict')$  in the  $\Sigma'$  system.

Since the speed of light is the same in both systems, we have

$$x^2 + y^2 + z^2 = c^2 t^2, \quad x'^2 + y'^2 + z'^2 = c^2 t'^2,$$

i.e.,

$$x^2 + y^2 + z^2 - c^2 t^2 = 0, \quad (15.4.3)$$

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0.$$

The real number 0 is a scalar, so  $x^2 + y^2 + z^2 - c^2 t^2$  is also a scalar. Scalars remain invariant under coordinate transformations; hence

$$x^2 + y^2 + z^2 - c^2 t^2 = x'^2 + y'^2 + z'^2 - c^2 t'^2. \quad (15.4.4)$$

Substituting (15.4.2) into (15.4.4) gives

$$x^2 + y^2 + z^2 - c^2 t^2 = (a_{11}x + a_{14}t)^2 + y^2 + z^2 - c^2 (a_{41}x + a_{44}t)^2,$$

or

$$\begin{aligned} x^2 + y^2 + z^2 - c^2 t^2 &= (a_{11}^2 - c^2 a_{41}^2)x^2 + y^2 + z^2 \\ &\quad + (a_{14}^2 - c^2 a_{44}^2)t^2 + (2a_{11}a_{14} - 2c^2 a_{41}a_{44})xt. \end{aligned}$$

This is an identity. For it to hold, the coefficients of corresponding terms on both sides must be equal. Therefore,

$$a_{11}^2 - c^2 a_{41}^2 = 1, \quad (15.4.5)$$

$$a_{14}^2 - c^2 a_{44}^2 = -c^2, \quad (15.4.6)$$

$$2a_{11}a_{14} - 2c^2 a_{41}a_{44} = 0. \quad (15.4.7)$$

Assume the point  $O'$  of the  $\Sigma'$  system moves along the positive  $x$ -axis of the  $\Sigma$  system with constant velocity  $v$ . In the  $\Sigma$  system,  $O'$  is moving, so its  $x$ -coordinate is  $x = vt$ . In the  $\Sigma'$  system, however,  $O'$  is stationary, so its  $x'$ -coordinate is  $x' = 0$ . From (15.4.2) we obtain

$$0 = a_{11}vt + a_{14}t.$$

Solving gives

$$a_{14} = -a_{11}v. \quad (15.4.8)$$

From (15.4.5)–(15.4.7) we get

$$\begin{aligned} a_{11}^2 a_{14}^2 &= c^4 a_{41}^2 a_{44}^2, & (1 + c^2 a_{41}^2) a_{14}^2 &= c^2 a_{41}^2 (c^2 + a_{14}^2), \\ a_{14}^2 + c^2 a_{41}^2 a_{14}^2 &= c^4 a_{41}^2 + c^2 a_{41}^2 a_{14}^2, & a_{14}^2 &= c^4 a_{41}^2 = c^2 c^2 a_{41}^2, \\ a_{11}^2 &= 1 + c^2 a_{41}^2 = 1 + \frac{1}{c^2} a_{14}^2 = 1 + \frac{1}{c^2} a_{11}^2 v^2, & \left(1 - \frac{v^2}{c^2}\right) a_{11}^2 &= 1, \\ a_{11} &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \end{aligned} \quad (15.4.9)$$

From (15.4.8) we obtain

$$a_{14} = \frac{-v}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (15.4.10)$$

From (15.4.6) we get

$$a_{44} = \frac{1}{c} \sqrt{c^2 + a_{14}^2} = \frac{1}{c} \sqrt{c^2 + \frac{v^2}{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (15.4.11)$$

From (15.4.7) we have

$$a_{41} = \frac{a_{11}a_{14}}{c^2 a_{44}} = \frac{1}{c^2} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{-v}{\sqrt{1 - \frac{v^2}{c^2}}} \sqrt{1 - \frac{v^2}{c^2}} = \frac{-\frac{v}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (15.4.12)$$

Substituting (15.4.9)–(15.4.12) into (15.4.2) yields

$$x' = a_{11}x + a_{14}t = \frac{x}{\sqrt{1-\frac{v^2}{c^2}}} + \frac{-v}{\sqrt{1-\frac{v^2}{c^2}}}t = \frac{x-vt}{\sqrt{1-\frac{v^2}{c^2}}},$$

$$t' = a_{41}x + a_{44}t = \frac{-\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}} + \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}t = \frac{t-\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Finally, we obtain the Lorentz transformation formulas:

$$\begin{cases} x' = \frac{x-vt}{\sqrt{1-\frac{v^2}{c^2}}}, \\ y' = y, \\ z' = z, \\ t' = \frac{t-\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}}. \end{cases} \quad (15.4.13)$$

The inverse transformation formulas of (15.4.13) are

$$\begin{cases} x = \frac{x'+vt}{\sqrt{1-\frac{v^2}{c^2}}}, \\ y = y', \\ z = z', \\ t = \frac{t'+\frac{v}{c^2}x'}{\sqrt{1-\frac{v^2}{c^2}}}. \end{cases} \quad (15.4.14)$$

From either (15.4.13) or (15.4.14) we find

$$1 - \frac{v^2}{c^2} > 0,$$

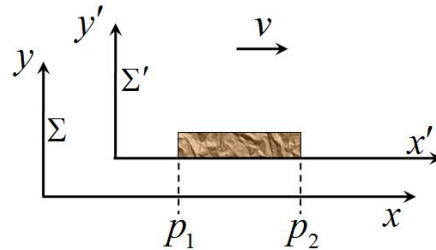
i.e.,  $c > v$  (since  $c > 0$  and  $v \geq 0$ ). Because the speed of light is the maximum speed,  $c$  is precisely the speed of light. This is why we previously chose  $c$  to represent the speed of light.

## 2.Length Contraction

According to the Lorentz transformation formulas, the length of a moving object appears shortened. As shown in Figure 15.4.2, suppose a ruler is fixed on the  $x'$ -axis of the  $\Sigma'$  system and moves along with it. Observed from  $\Sigma'$ , the ruler is stationary. Let the coordinates of the front end and the rear end of the ruler measured in  $\Sigma'$  be  $x'^2$  and  $x'^1$ , respectively. Then the length measured in  $\Sigma'$  is the **rest length**  $l_0$ :

$$l_0 = x'^2 - x'^1.$$

In the  $\Sigma$  system, the ruler is moving. Assume that in  $\Sigma$  the rear end of the ruler through point  $p_1$  and the front end of the ruler through point  $p_2$  are simultaneous events, i.e.,  $t^1 = t^2$ . Let the coordinates of point  $p_1$  be  $x^1$  and those of point  $p_2$  be  $x^2$ . Then the length of the



**Figure 15.4.2** In the moving  $\Sigma'$  system, a ruler appears contracted

ruler measured in  $\Sigma$  is defined as

$$l = x^2 - x^1.$$

From the Lorentz transformation formulas we have

$$x'^1 = \frac{x^1 - vt^1}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x'^2 = \frac{x^2 - vt^2}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Since  $t^1 = t^2$ , subtracting the two equations yields

$$x'^2 - x'^1 = \frac{x^2 - x^1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Hence,

$$l = l_0 \sqrt{1 - \frac{v^2}{c^2}}.$$

According to this formula, an observer in the  $\Sigma$  system finds that the length of a moving ruler is shortened in the direction of motion (the  $x$ -axis direction).

### 3.Velocity Transformation Formulas

Suppose an object moves with speed  $u'$  in the  $\Sigma'$  system. Then

$$u' = \frac{x'}{t'}.$$

Substituting the Lorentz transformation formulas (15.4.13) gives

$$u' = \frac{x - vt}{t - \frac{v}{c^2}x} = \frac{\frac{x}{t} - v}{1 - \frac{v}{c^2} \frac{x}{t}} = \frac{u - v}{1 - \frac{uv}{c^2}}, \quad (15.4.15)$$

where

$$u = \frac{x}{t}$$

is the speed of the object observed in the  $\Sigma$  system.

From (15.4.15) we can solve for the inverse transformation:

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}}. \quad (15.4.16)$$

Equations (15.4.15) and (15.4.16) are the **speed transformation formulas**.

**Case 1:** Assume the  $\Sigma'$  system moves relative to the  $\Sigma$  system with speed  $c$ , and an object in  $\Sigma'$  moves with speed  $u'$ . The speed of this object observed in the  $\Sigma$  system, from (15.4.16), is

$$u = \frac{u' + c}{1 + \frac{u'c}{c^2}} = \frac{u' + c}{\frac{c + u'}{c}} = c.$$

**Case 2:** Assume the  $\Sigma'$  system moves relative to the  $\Sigma$  system with speed  $v$  (with  $v < c$ ), and an object in  $\Sigma'$  moves with speed  $c$ . Then from (15.4.16),

$$u = \frac{c + v}{1 + \frac{cv}{c^2}} = \frac{c + v}{\frac{c + v}{c}} = c.$$

**Case 3:** Assume the  $\Sigma'$  system moves relative to the  $\Sigma$  system with speed  $c$ , and an object in  $\Sigma'$  also moves with speed  $c$ . From (15.4.16),

$$u = \frac{c + c}{1 + \frac{cc}{c^2}} = \frac{2c}{1 + 1} = c.$$



These three cases demonstrate that **in any inertial frame, the speed of an object (including photons and particles with nonzero rest mass) cannot exceed the speed of light  $c$** ; in other words, the speed of light  $c$  is the ultimate limit for the speed of material motion.

## §15.5 Particles Expand

We have assumed that electrons, protons, neutrinos, and their antiparticles are each an  $SU(2)$  group. From equation (12.10.7) we have

$$x^2 + y^2 + z^2 + w^2 = 1. \quad (15.5.1)$$

In four-dimensional Euclidean space  $R^4$ , if we take its coordinate system as the Cartesian system  $(x, y, z, w)$ , then equation (15.5.1) can be regarded as the equation of the three-dimensional unit sphere  $S^3$  in  $R^4$ . The  $SU(2)$  group is homeomorphic to  $S^3$ .

### 1. Expansion Speed

Transform the space  $R^4$  into Minkowski spacetime  $M^4$ . Take the coordinates  $x, y, z$  in the coordinate system  $(x, y, z, w)$  as spatial coordinates, and replace  $w$  with  $ict$ ; then the coordinate system becomes  $(x, y, z, ict)$ , i.e., the coordinate system of  $M^4$  is taken as  $(x, y, z, ict)$ . In this coordinate system, the equation for  $S^3$  becomes

$$x^2 + y^2 + z^2 + (ict)^2 = 1,$$

i.e.,

$$x^2 + y^2 + z^2 - c^2 t^2 = 1,$$

or

$$x^2 + y^2 + z^2 = 1 + c^2 t^2. \quad (15.5.2)$$

Since time  $t$  continuously increases, this equation implies that the spatial extent of the particle should also continuously expand; thus, the particle is **expanding**. Let

$$x^2 + y^2 + z^2 = r^2, \quad (15.5.3)$$

then equation (15.5.2) becomes

$$r^2 = 1 + c^2 t^2, \quad (15.5.4)$$

or

$$r = \sqrt{1 + c^2 t^2}. \quad (15.5.5)$$

Equation (15.5.3) is the equation of a two-dimensional sphere  $S^2$  of radius  $r$ , while equation (15.5.5) tells us that the radius  $r$  of the sphere  $S^2$  increases with time  $t$ ; that is, the particle is expanding. As shown in Figure 15.5.1, the

empty central region in the figure belongs to a part of  $R^3$  and is flat; this region is also called a **black hole**. Therefore, the sphere  $S^3$  is homeomorphic to a three-dimensional spherical shell in  $R^3$  from which a central ball of radius  $r$  has been removed, and this shell expands as time passes.

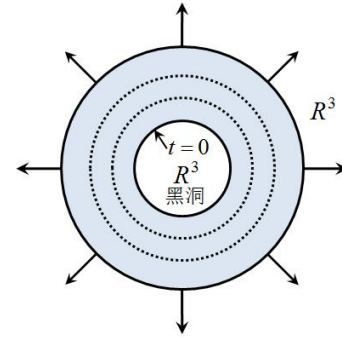
Differentiating equation (15.5.5) with respect to time  $t$  gives the expansion speed of the particle:

$$\frac{dr}{dt} = \frac{2c^2 t}{2\sqrt{1 + c^2 t^2}} = \frac{c^2 t}{\sqrt{1 + c^2 t^2}}. \quad (15.5.6)$$

When time  $t$  tends to infinity, the limiting expansion speed is

$$\lim_{t \rightarrow \infty} \frac{dr}{dt} = \lim_{t \rightarrow \infty} \frac{c^2 t}{\sqrt{\left(\frac{1}{t^2} + c^2\right) t^2}} = \lim_{t \rightarrow \infty} \frac{c^2}{\sqrt{\frac{1}{t^2} + c^2}} = c. \quad (15.5.7)$$

Thus, when the time  $t$  is sufficiently long, the particle expands at a speed close to the speed of



**Figure 15.5.1** The center of the particle is flat

light  $c$ .

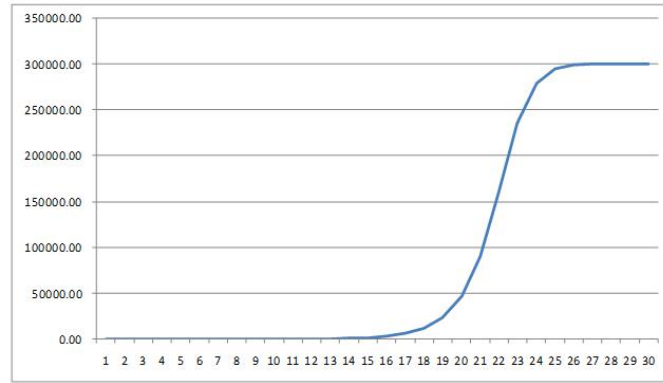
To gain an intuitive understanding of the particle expansion process, we plot the curve  $dr/dt$ . The unit of time  $t$  is seconds, and the speed of light  $c = 300000$  km/s. First, we calculate 30 specific points of the curve, as shown in Table 15.5.1, and then plot the curve based on Table 15.5.1, as illustrated in Figure 15.5.2.

**Table 15.5.1** Particle Expansion Speed as a Function of Time (1)

No.	Time (s)	Speed (km/s)
1	0.000000000001	0.09
2	0.000000000002	0.18
3	0.000000000004	0.36
4	0.000000000008	0.72
5	0.000000000016	1.44
6	0.000000000032	2.88
7	0.000000000064	5.76
8	0.000000000128	11.52
9	0.000000000256	23.04
10	0.000000000512	46.08
11	0.000000001024	92.16
12	0.000000002048	184.32
13	0.000000004096	368.64
14	0.000000008192	737.28
15	0.000000016384	1474.54
16	0.000000032768	2948.98
17	0.000000065536	5897.10
18	0.000000131072	11787.37
19	0.000000262144	23520.34
20	0.000000524288	46612.86
21	0.000001048576	90022.75
22	0.000002097152	159755.96
23	0.000004194304	234863.36
24	0.000008388608	278795.67
25	0.000016777216	294248.54
26	0.000033554432	298530.57
27	0.000067108864	299630.61
28	0.000134217728	299907.52
29	0.000268435456	299976.87
30	0.000536870912	299994.22

**Table 15.5.1** Particle Expansion Speed as a Function of Time (2)

Time (s)	Speed (km/s)
1	300000.00
2	300000.00
4	300000.00
8	300000.00
16	300000.00
32	300000.00
64	300000.00
128	300000.00
256	300000.00
512	300000.00
1024	300000.00
2048	300000.00
4096	300000.00
8192	300000.00
16384	300000.00
32768	300000.00
65536	300000.00
131072	300000.00
262144	300000.00
524288	300000.00
1048576	300000.00
2097152	300000.00
4194304	300000.00
8388608	300000.00
16777216	300000.00
33554432	300000.00
67108864	300000.00
134217728	300000.00
268435456	300000.00
536870912	300000.00



**Figure 15.5.2** Particle expansion speed as a function of time.

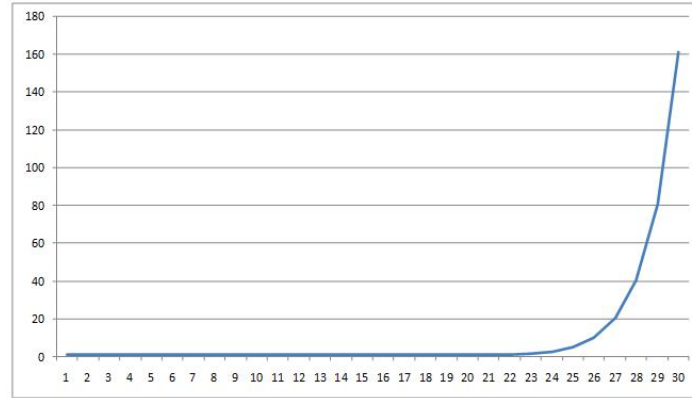
From Table 15.5.1 and Figure 15.5.2, we can see that the expansion speed undergoes three distinct phases. From 0.000000000001 seconds to 0.000000016384 seconds, the speed increases relatively slowly. However, after 0.000000016384 seconds, the speed begins to rise rapidly. By 0.000016777216 seconds, the expansion speed is already close to the speed of light. After that, the increase in expansion speed becomes very small, essentially expanding at the speed of light. If the plot were drawn starting from 1 second, these three phases would not be discernible, as shown in Table 15.5.2.

We calculate several characteristic values of the expansion radius as a function of time, as presented in Table 15.5.3, then plot the curve based on Table 15.5.3, as illustrated in Figure 15.5.3.

**Table 15.5.3** Particle Expansion Radius as a Function of Time

No.	Time (s)	Radius (km)
1	0.000000000001	1
2	0.000000000002	1
3	0.000000000004	1
4	0.000000000008	1
5	0.000000000016	1
6	0.000000000032	1
7	0.000000000064	1
8	0.000000000128	1.000000001
9	0.000000000256	1.000000003
10	0.000000000512	1.000000012
11	0.000000001024	1.000000047
12	0.000000002048	1.000000189
13	0.000000004096	1.000000755
14	0.000000008192	1.00000302
15	0.000000016384	1.00001208
16	0.000000032768	1.000048317
17	0.000000065536	1.000193255
18	0.000000131072	1.000772796
19	0.000000262144	1.00308761
20	0.000000524288	1.012293935
21	0.000001048576	1.048311045
22	0.000002097152	1.181450035
23	0.000004194304	1.607263744
24	0.000008388608	2.707985778

25	0.000016777216	5.131544398
26	0.000033554432	10.11587819
27	0.000067108864	20.15747917
28	0.000134217728	40.27773412
29	0.000268435456	80.53684538
30	0.000536870912	161.064378



**Table 15.5.3** Particle Expansion Radius as a Function of Time

From Table 15.5.3 and Figure 15.5.3, it can be seen that the particle's radius increases only slightly from the initial time of 0.000000000001 seconds to 0.000004194304 seconds, but thereafter it grows rapidly.

## 2.Expansion Acceleration

Let us compute the expansion acceleration  $A$  of the particle:

$$A = \frac{d^2 r}{dt^2} = \frac{d}{dt} \left( \frac{c^2 t}{\sqrt{1 + c^2 t^2}} \right) = \frac{c^2}{\sqrt{1 + c^2 t^2}} \left( 1 - \frac{c^2 t^2}{1 + c^2 t^2} \right). \quad (15.5.8)$$

Since

$$1 > \frac{c^2 t^2}{1 + c^2 t^2},$$

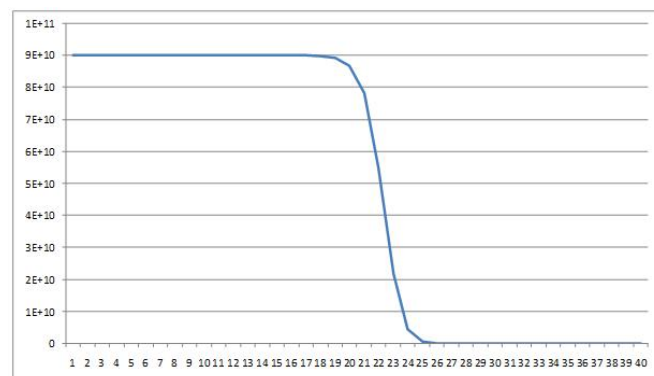
the expansion acceleration  $A$  is greater than zero; that is, the particle expands with **accelerated expansion**. However, as time  $t$  tends to infinity, the acceleration approaches zero, because

$$\lim_{t \rightarrow \infty} A = \lim_{t \rightarrow \infty} \frac{c^2}{\sqrt{1 + c^2 t^2}} \left( 1 - \frac{c^2 t^2}{1 + c^2 t^2} \right) = \lim_{t \rightarrow \infty} \frac{c^2}{\sqrt{1 + c^2 t^2}} \left( 1 - \frac{c^2}{\frac{1}{t^2} + c^2} \right) = 0. \quad (15.5.9)$$

**Table 15.5.4** Particle Expansion Acceleration as a Function of Time

No.	Time (s)	Acceleration (km/s <sup>2</sup> )
1	0.000000000001	90000000000
2	0.000000000002	90000000000
3	0.000000000004	90000000000
4	0.000000000008	89999999999
5	0.000000000016	89999999997
6	0.000000000032	89999999988
7	0.000000000064	89999999950
8	0.000000000128	89999999801

9	0.000000000256	89999999204
10	0.000000000512	89999996815
11	0.000000001024	89999987260
12	0.000000002048	89999949039
13	0.000000004096	89999796157
14	0.000000008192	89999184633
15	0.000000016384	89996738608
16	0.000000032768	89986955613
17	0.000000065536	89947841350
18	0.000000131072	89791667293
19	0.000000262144	89171466989
20	0.000000524288	86760611515
21	0.000001048576	78121761119
22	0.000002097152	54575337557
23	0.000004194304	21676095941
24	0.000008388608	4532140572
25	0.000016777216	666036818
26	0.000033554432	86942422.24
27	0.000067108864	10988384.57
28	0.000134217728	1377359.844
29	0.000268435456	172289.4428
30	0.000536870912	21539.91625
31	0.001073741824	2692.606299
32	0.002147483648	336.5794365
33	0.004294967296	42.07254360
34	0.008589934592	5.259071512
35	0.017179869184	0.657384051
36	0.034359738368	0.082173010
37	0.068719476736	0.010271627
38	0.137438953472	0.001283953
39	0.274877906944	0.000160494
40	0.549755813888	2.00618E-05



**Figure 15.5.4** Particle expansion acceleration as a function of time

From Table 15.5.4 and Figure 15.5.4, it can be seen that the particle expansion acceleration undergoes three distinct phases: the first phase exhibits the highest acceleration; the second phase shows a rapid decline; and the third phase approaches zero. The magnitude of the expansion acceleration reflects the intensity of the particle's expansion: in the first phase, the acceleration is greatest, indicating the most violent and forceful expansion; in the second phase, the intensity of

expansion diminishes rapidly; and in the final phase, the particle expands nearly uniformly. This expansion process resembles an explosion.

Not only electrons, protons, neutrinos, and their antiparticles are expanding—photons are also expanding. First, because a photon is a three-dimensional real projective space  $RP^3$ , which can be obtained by identifying antipodal points of the unit sphere  $S^3$ . The spatial and temporal coordinates of the unit sphere  $S^3$  are continuously increasing, so the photon is also continuously expanding. On the other hand, if we set

$$x^1 = y^1 = z^1 = t^1 = 0, \quad x^2 = x, \quad y^2 = y, \quad z^2 = z, \quad t^2 = t,$$

in equation (15.2.3), we obtain

$$x^2 + y^2 + z^2 = c^2 t^2, \quad (15.5.10)$$

which shows that the photon (including its entire spatial part) expands as time elapses.

In Figure 15.4.1, assume that a particle with rest mass and spin-1/2 (such as an electron) is located at the origin of the  $\Sigma'$  coordinate system and moves uniformly to the right with speed  $v$  relative to the  $\Sigma$  coordinate system. In the  $\Sigma'$  system, the equation of the particle is

$$x'^2 + y'^2 + z'^2 = 1 + c^2 t'^2, \quad (15.5.11)$$

so the expansion process of the particle is as shown in Figures 15.5.2, 15.5.3, and 15.5.4. The question then arises: when observed in the  $\Sigma$  system, does the expansion process of the particle also appear as depicted in Figures 15.5.2, 15.5.3, and 15.5.4?

To answer this question, we need to transform equation (15.5.11) into the  $\Sigma$  coordinate system. Substituting the Lorentz transformation formulas (15.4.13) into (15.5.11), we obtain:

$$\begin{aligned} & \left( \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \right)^2 + y^2 + z^2 = 1 + c^2 \left( \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}} \right)^2, \\ & \frac{x^2 - 2xvt + v^2 t^2}{1 - \frac{v^2}{c^2}} + y^2 + z^2 = 1 + \frac{c^2 t^2 - 2tx + \frac{v^2}{c^2} x^2}{1 - \frac{v^2}{c^2}}, \\ & \frac{x^2 + v^2 t^2 - c^2 t^2 - \frac{v^2}{c^2} x^2}{1 - \frac{v^2}{c^2}} + y^2 + z^2 = 1, \\ & \frac{x^2(1 - \frac{v^2}{c^2}) + c^2 t^2(\frac{v^2}{c^2} - 1)}{1 - \frac{v^2}{c^2}} + y^2 + z^2 = 1, \\ & x^2 + y^2 + z^2 - c^2 t^2 = 1. \end{aligned}$$

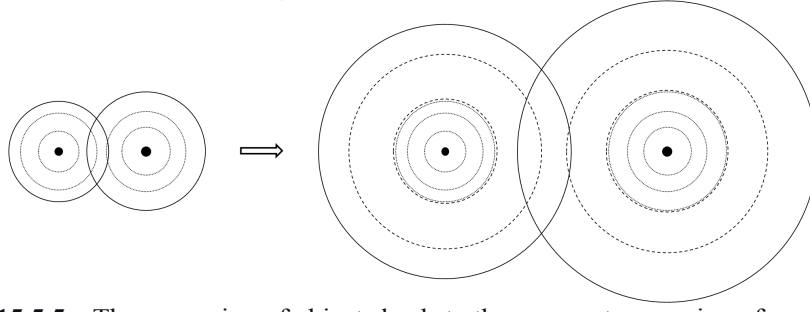
Finally we obtain

$$x^2 + y^2 + z^2 = 1 + c^2 t^2. \quad (15.5.12)$$

Therefore, the expansion process of a spin-1/2 elementary particle located in the  $\Sigma'$  system, as observed in the  $\Sigma$  system, is identical to the expansion process of that particle observed in the  $\Sigma'$  system. Similarly, the expansion process of a photon located in the  $\Sigma'$  system, as observed in the  $\Sigma$  system, is also the same as the expansion process of that photon observed in the  $\Sigma'$  system.

Elementary particles are expanding; consequently, particles composed of elementary particles, as well as all macroscopic objects, are also expanding. Since objects are three-dimensional and the physical space is also three-dimensional, a three-dimensional object cannot be contained in a three-dimensional space alone. Instead, it must reside in the four-dimensional spacetime  $M^4$ , which combines three spatial dimensions with one temporal dimension. This causes objects

(including the spatial extent of their associated fields) to enlarge as time elapses—this is the reason why objects expand. Because objects expand, the entire observable universe appears to be expanding as well, as illustrated in Figure 15.5.5.

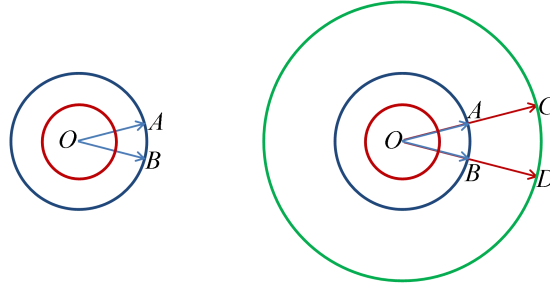


**Figure 15.5.5** The expansion of objects leads to the apparent expansion of cosmic space

Astronomer Hubble (1929), through observations of the observable universe, discovered that the observable universe is expanding. In reality, it is not the observable universe itself that is expanding, but rather the matter within the observable universe. The expansion of the observable universe is the collective manifestation of the expansion of all matter within it.

Traditional cosmological theories suggest that the universe expands due to a repulsive force that causes matter in the universe to move apart. For such a repulsive force to exist, a form of energy is required, termed **dark energy**, leading to the dark energy hypothesis. However, according to the analysis in this chapter, **cosmic expansion is not a mechanical phenomenon but a geometric phenomenon. Therefore, our theory does not require the dark energy hypothesis.**

### 3.The Expansion Process is a Conformal Transformation



**Figure 15.5.6** During the expansion process,  $\angle AOB = \angle COD$

As shown in Figure 15.5.6, during the expansion process the outermost boundary of the space occupied by a particle or object continuously expands outward, causing the spatial extent to grow, while  $\angle AOB = \angle COD$ . Since a particle is a Lie group, both particles and objects composed of particles can be regarded as smooth manifolds. A positive-definite Riemannian metric can be defined on a smooth manifold, making it a Riemannian manifold. Consequently, the spatial expansion process can be viewed as a **conformal transformation** (angle-preserving transformation) of the Riemannian manifold. Suppose the metric tensor of the particle or object before expansion is  $g_{ij}$  ( $i, j = 1, 2, 3$ ), and after expansion it is  $\tilde{g}_{ij}$ . Then, according to Definition 11.4.5 of a conformal transformation, we have

$$\tilde{g} = \lambda^2 g ,$$

where  $\lambda$  is a positive smooth function. Before expansion, the distance between points A and B is

$$s = \int_A^B \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt ,$$

and after expansion it becomes

$$\tilde{s} = \int_A^B \sqrt{\tilde{g}_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt = \int_A^B \sqrt{\lambda^2 g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt .$$

In a local region at a given moment,  $\lambda$  can be approximated as a positive constant. Hence,

$$\tilde{s} = \int_A^B \sqrt{\tilde{g}_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt = \lambda \int_A^B \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt = \lambda s ,$$

so the distance between any two points in that local region is scaled by a factor of  $\lambda$  compared to the distance  $\tilde{s}$  before expansion, i.e.,  $\tilde{s} = \lambda s, \lambda > 1$ .

#### 4.No Singularity Exists

Astronomers, by extrapolating the expansion speed of the observable universe backward to its starting point, infer that the expansion originated from a **singularity**—a point where the Riemann curvature becomes infinite. We contend that such a singularity does **not** exist. First, it is not the observable universe that is expanding, but rather the objects within it. Second, according to equation (15.5.2), at time  $t = 0$ , the radius of the two-dimensional sphere  $S^2$  is not zero but equal to 1; hence, no singularity is present. Refer to Figure 15.5.1: the region within the radius  $r < 1$  of the sphere  $S^2$  is part of the flat space  $R^3$  and is flat. Similarly, for a photon at time  $t = 0$ , equation (15.5.10) becomes

$$x^2 + y^2 + z^2 = c^2 t^2 = 0 , \quad (15.5.13)$$

which yields

$$x = 0 , \quad y = 0 , \quad z = 0 , \quad t = 0 ,$$

However, the point  $(x, y, z, t) = (0, 0, 0, 0)$  does not belong to the photon, because this point is precisely the origin shown in Figure 15.2.1, which is excluded from  $(x, y, z, t) = (0, 0, 0, 0)$ . Therefore, at time  $t = 0$ , no singularity exists. Since the origin  $(0, 0, 0, 0)$  is a point of the flat spacetime  $M^4$ , the Riemann curvature there is zero.



## Chapter 16 Lie-Group Interpretation of Quantum Mechanics

### Principles

If photons and antiphotons are regarded as an  $SO(3)$  group, electrons, protons, neutrinos, and their antiparticles are all regarded as an  $SU(2)$  group, and the composite particles or matter formed by these particles are regarded as a Lie group, then the basic principles of quantum mechanics become very intuitive and very reasonable.

#### §16.1 Structure of Particles

Among the elementary particles, spin-1/2 particles—electrons, protons, neutrinos, and their antiparticles—are each an  $SU(2)$  group. Spin-1 photons and antiphotons are each an  $SO(3)$  group. Particles composed of these elementary particles (called **composite particles**) are internal direct products of their respective  $SU(2)$  or  $SO(3)$  groups. An internal direct product group is also a Lie group. Therefore, **both elementary particles and composite particles are Lie groups**.

In Chapter 13 it was already stipulated: the term “**point-like particle**” refers to the extremely tiny core of a particle; whereas the term “**particle**” denotes the entity as a whole, including both the point-like core and the field surrounding it. The entire particle is a Lie group. As shown in Figure 16.1.1, a Lie group consists of two parts: the point-like particle and the field. We represent the point-like particle by the identity element  $e$ , while the remainder of the Lie group—other than the identity element—constitutes the field. The neighborhood (or vicinity) of the identity element  $e$  is the **local Lie group** (often called the infinitesimal group). Since the identity element always lies within the local Lie group, we can regard the local Lie group as the **domain of activity** of the point-like particle. There is no absolute distinction between the full connected Lie group and the local Lie group, because except for the identity element  $e$ , the interior and exterior of the local Lie group are the same. Moreover, the extent of a local Lie group can be arbitrarily small—reducing to just a neighborhood of  $e$ —or it can be as large as the entire Lie group, which can also be viewed as a local Lie group. This is because through the interaction of group elements, every group element can potentially become the identity element; that is, the location of any group element could be where the point-like particle appears. Hence, the whole Lie group may also be considered a local Lie group.

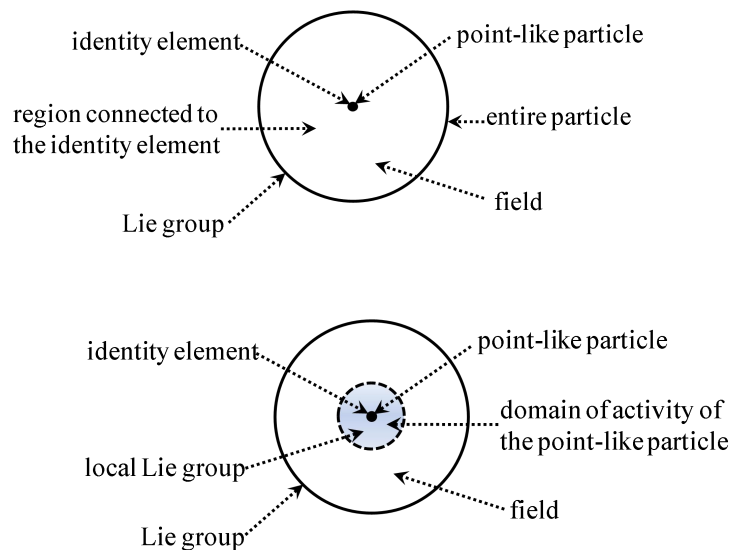
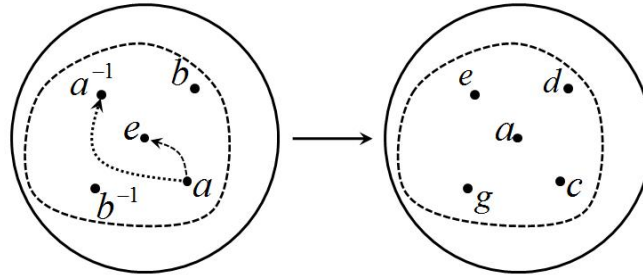


Figure 16.1.1 Structure of a particle

## §16.2 Point-Like Particles Undergo Random Motion

### 1. Point-Like Particles Undergo Random Motion

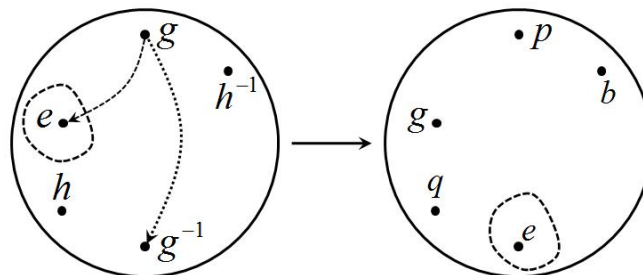
Since a Lie group is a continuous group with infinitely many elements, every group element can act on other elements by left or right multiplication. Consequently, every group element also acts on the identity element  $e$ —the point-like particle—by left or right multiplication, causing the point-like particle to undergo left or right translation.



**Figure 16.2.1** Random action of elements of the local Lie group on the point-like particle

This characteristic of interaction among group elements implies that the trajectory of the identity element  $e$  (or the point-like particle) is not continuous but **discontinuous**. As shown in Figure 16.2.1, within the local Lie group, the point-like particle originally appears at point  $e$ . Under the action of some group element  $a$ , it becomes  $ae$ , which is not the identity element; hence the point-like particle vanishes at  $e$ . However, when the group element  $a^{-1}$  is acted upon by  $a$ , it becomes the identity element:  $aa^{-1} = e$ . This means the point-like particle reappears at the original location of  $a^{-1}$ .

During the transition from the left panel to the right panel in Figure 16.2.1, does the identity element vanish at  $e$  and simultaneously appear at  $a^{-1}$ , or do these two events occur sequentially? They should be **simultaneous**. If they occurred one after the other, there would be an interval during which no identity element exists, and the set would cease to be a group. This is not difficult to understand: the local Lie group can be regarded as a whole, subject simultaneously to a certain transformation that causes all its elements to change at the same time—analogous to a rigid rotation of a triangle. When a triangle rotates, every point on the triangle changes position simultaneously. In a Lie group, every group element can act on the group itself by left or right multiplication; both left and right actions are smooth diffeomorphisms of the group onto itself.



**Figure 16.2.2** Random action of Lie group elements on the point-like particle

Not only the elements of the local Lie group can act (left or right) on the identity element  $e$ ; **all** elements of the full Lie group can act on the identity element  $e$ . As illustrated in Figure 16.2.2, within the entire Lie group, the point-like particle originally resides at point  $e$ . Under the action of some group element  $g$ , it becomes  $ge$ , which is not the identity element; consequently, the point-like particle vanishes at  $e$ . However, when the group element  $g^{-1}$  is acted upon by  $g$ , it becomes the identity element:  $gg^{-1} = e$ . This means the point-like particle reappears at the

original location of  $g^{-1}$ .

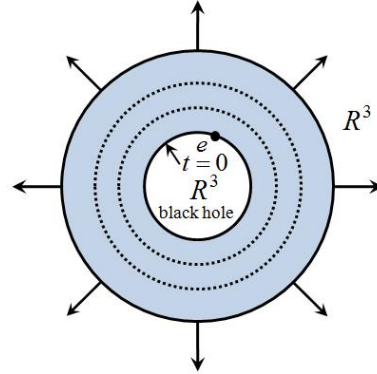
Similarly, during the transition from the left panel to the right panel in Figure 16.2.2, the disappearance of the identity element at  $e$  and its reappearance at  $g^{-1}$  occur **simultaneously**. The entire Lie group can be viewed as a whole, subject simultaneously to a transformation that causes **all** elements of the Lie group to change at the same time.

Frequent interactions among group elements cause **all** elements of the entire group to change. The point-like particle vanishes at one location within the group and simultaneously appears at another location. We call this kind of motion of the point-like particle **appearance-disappearance motion** (or emergence-vanishing motion). The trajectory of this motion is **discontinuous** rather than continuous. Because the group contains infinitely many elements, it is impossible to predict which specific group element will become the identity element  $e$  at a given moment. Consequently, we cannot predict **when** or **where** the point-like particle will appear. Therefore, appearance-disappearance motion is a **random motion**.

Every non-identity group element, when multiplied by its own inverse, yields the identity element  $e$ . However, by the definition of a group, a group has **one and only one** identity element  $e$ . Therefore, the point-like particle can appear at **only one** location; it cannot appear simultaneously at two or more positions.

Since the local Lie group is defined in the neighborhood of the identity element, the action of its elements on the identity element is more frequent or convenient. However, by the definition of a group, the interactions among elements of the local Lie group will still keep the identity element **within** the local Lie group. Hence, we say the local Lie group is the **domain of activity** of the point-like particle. It should be noted that the local Lie group has no fixed boundary; moreover, **both** the elements inside the local Lie group and those outside it can act on the identity element by left or right multiplication, thereby also acting on the point-like particle and causing it to shift left or right.

In Chapter 15, the structure of an elementary particle with rest mass and spin-1/2 (such as an electron) in Minkowski spacetime  $M^4$  was already described: its inner central region is an empty zone that belongs to a part of  $R^3$  and is flat—this part is also called a **black hole**—while its outermost part continuously expands, as shown in Figure 16.2.3. The innermost two-dimensional sphere  $S_0^2$  is the sphere at the moment the particle is created, i.e., at time  $t=0$ . As time elapses, successive layers of two-dimensional spheres are generated outside this sphere, with outer spheres larger than inner ones, forming a **foliated structure**.



**Figure 16.2.3** The identity element  $e$  in  $R^3$  lies on  $S_0^2$

An element of the  $SU(2)$  group is represented as

$$\begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}.$$

From the unimodular condition we obtain

$$(x + iy)(x - iy) + (z + iw)(z - iw) = 1,$$

i.e.,

$$x^2 + y^2 + z^2 + w^2 = 1.$$

Let  $w = ict$ . Then

$$x^2 + y^2 + z^2 = 1 + c^2 t^2.$$

Therefore, the identity element  $e$  of the  $SU(2)$  group corresponds to the point  $(1,0,0,0)$ , i.e.,  $(x, y, z, ict) = (1,0,0,0)$ . This point lies precisely on the two-dimensional sphere  $S_0^2$  at time  $t=0$ .

In Minkowski spacetime  $M^4$ ,  $S_0^2$  is sometimes regarded as the point-like particle because

$S_0^2$  is sufficiently small and close to the identity element  $e$ . If  $S_0^2$  is taken as the point-like particle, then when the identity element  $e$  is acted upon by other group elements but remains on  $S_0^2$ , the spatial position of the point-like particle does not change.

In  $M^4$ , if  $S_0^2$  is regarded as the point-like particle, then identifying antipodal points of  $S_0^2$  yields the two-dimensional projective plane  $RP_0^2$ , which can be taken as the point-like particle of a photon. The structure of the point-like particle for neutrinos and antineutrinos will be discussed in Chapter 18.

## 2. Why Particles Exhibit Wave-Like Behavior

In Figure 16.2.2, the group element at point  $g^{-1}$  in the Lie group  $G$  is  $g^{-1}$ . However, due to interactions among group elements, the group element at point  $g^{-1}$  can change into other elements. For example, as shown in Figure 16.2.4, when a group element  $g$  acts on  $g^{-1}$  from the left, the group element  $g^{-1}$  at point  $g^{-1}$  becomes the identity element  $e$  ( $gg^{-1} = e$ ), and at that moment the point-like particle appears at point  $g^{-1}$ . When a group element  $a$  then acts on  $e$  from the left, the group element at the original point  $g^{-1}$  changes to  $a$ , and the point-like particle disappears from that location  $g^{-1}$ . When another group element  $b$  subsequently acts on  $a$  from the left, the group element  $a$  becomes  $ba$ . Finally, when the group element  $g^{-1}a^{-1}b^{-1}$  acts on  $ba$  from the left, the group element  $ba$  reverts to its original form  $g^{-1}$ , completing one cycle of change for the group element  $g^{-1}$  at the original point  $g^{-1}$ .

$$g^{-1} \xrightarrow{g} e \xrightarrow{a} a \xrightarrow{b} ba \xrightarrow{g^{-1}a^{-1}b^{-1}} g^{-1}$$

**Figure 16.2.4** Periodic variation of group elements.

Other elements of the Lie group  $G$  undergo similar periodic variations. This periodic oscillation of all elements throughout the Lie group  $G$  makes the entire group resemble a **standing wave**. If the whole Lie group  $G$  undergoes a uniform translation, then it becomes a **traveling wave**. This explains why the entire particle (including its field space) exhibits wave-like properties. Equations (12.9.4) and (12.10.8) can both be regarded as the **wave functions** of the particle.

If a rope were infinitely rigid, shaking it would never produce a wave. However, if the rope is flexible, shaking it generates a wave because the rope can be regarded as composed of infinitely many point, each independent yet interacting with the others; thus, shaking the rope causes it to undulate. Similarly, a particle is a smooth manifold consisting of infinitely many point. Each point is both independent and interactive (through left and right actions), so the particle possesses wave-like properties. Since the point-like particle can appear at any point of the smooth manifold, it exhibits both wave-like and particle-like characteristics—that is, it possesses **wave-particle duality**.

## 3. Regarding the Local Lie Group as a Wave Packet

A local Lie group is also a group; the elements inside it undergo periodic variations similar to those shown in Figure 16.2.4. Therefore, the local Lie group can be viewed as a “**wave packet**.” This is not a wave packet in the classical sense but a special, non-dispersing wave packet whose extent can be defined as the neighborhood surrounding the point-like particle.

The group velocity of the wave packet is

$$v_g = \frac{d\omega}{dk}, \quad (16.2.1)$$

where  $\omega = 2\pi\nu$ ,  $\nu$  is the frequency and  $k = 2\pi/\lambda$ ,  $\lambda$  is the wave number.

From special relativity, the energy of a free point-like particle is

$$E = \sqrt{p^2c^2 + m_0^2c^4}.$$

From this we can compute the velocity of the free point-like particle:

$$v = \frac{dE}{dp}, \quad (16.2.2)$$

because

$$\frac{dE}{dp} = \frac{d}{dp} \sqrt{p^2 c^2 + m_0^2 c^4} = \frac{pc^2}{\sqrt{p^2 c^2 + m_0^2 c^4}} = \frac{pc^2}{E} = \frac{mvc^2}{mc^2} = v.$$

The identity element  $e$  lies at the center of the local Lie group, so the position of the point-like particle is also at the center of the wave packet. Hence, the group velocity  $v_g$  of the wave packet equals the velocity  $v$  of the point-like particle:  $v_g = v$ . From equations (16.2.1) and (16.2.2) we obtain

$$\frac{d\omega}{dk} = \frac{dE}{dp}. \quad (16.2.3)$$

Comparing the two sides of (16.2.3) yields

$$E = \hbar\omega + a, \quad p = \hbar k + b,$$

where  $\hbar$  is the proportionality coefficient,  $a$  and  $b$  are constants. With appropriate choice of units we can set  $a = 0$  and  $b = 0$ . Therefore,

$$E = \hbar\omega, \quad p = \hbar k.$$

Substituting  $\omega = 2\pi\nu$  and  $k = 2\pi/\lambda$  gives

$$E = h\nu, \quad (16.2.4)$$

$$p = \frac{h}{\lambda}, \quad (16.2.5)$$

where  $h$  is the Planck constant, and

$$h = 2\pi\hbar. \quad (16.2.6)$$

#### 4. Plane-Wave Function

In classical mechanics, a plane wave of wavelength  $\lambda$ , frequency  $\nu$ , propagating along the unit vector  $\mathbf{n}$  can be expressed as

$$\psi_\lambda = a \cos \left[ 2\pi \left( \frac{\mathbf{r} \cdot \mathbf{n}}{\lambda} - \nu t \right) - \delta \right],$$

where  $a$  is the amplitude,  $\delta$  is the initial phase, and  $\mathbf{r}$  is the position vector from the origin to any point  $p(x, y, z)$  on the wavefront. Using Euler's formula,  $\psi_\lambda$  can be written in complex form (taking only the real part):

$$\psi_\lambda = A e^{2\pi i \left( \frac{\mathbf{r} \cdot \mathbf{n}}{\lambda} - \nu t \right)}, \quad (16.2.7)$$

where  $A = a e^{-i\delta}$ .

Substituting equations (16.2.4) and (16.2.5) into (16.2.7) yields

$$\psi_p = A e^{-\frac{i}{\hbar}(Et - \mathbf{r} \cdot \mathbf{p})}, \quad (16.2.8)$$

where

$$\mathbf{r} \cdot \mathbf{n} p = \mathbf{r} \cdot \mathbf{n} m v = \mathbf{r} \cdot \mathbf{v} m = \mathbf{r} \cdot \mathbf{p}.$$

The function  $\psi_p$  can serve as the **wave function** for a free particle with momentum  $\mathbf{p}$  and energy  $E$ ; it is defined on the smooth manifold of the particle.

### §16.3 Probability Interpretation of the Wave Function

We have stipulated that the identity element  $e$  represents the point-like particle (i.e., the location where the point-like particle appears). Due to interactions among group elements, the position of the identity element is not fixed; it can appear at different points of the Lie group  $G$ . That is, every element of the Lie group may transform into the identity element. For example, let

$e, g, h^{-1} \in G$  and let  $e$  be the identity element. When the element  $g^{-1}$  acts on  $G$  from the left,  $g$  becomes the identity element  $e$ , while the original identity element becomes the element  $g(g = ge)$ . Similarly, when the element  $h$  acts on  $G$  from the right,  $h^{-1}$  becomes the identity element ( $h^{-1}h = e$ ), and the original identity element  $g^{-1}g (= e)$  becomes the element  $h(g^{-1}gh = eh = h)$ .

Because of these interactions, within a given period the identity element appears with different frequencies at different points of the Lie group. For instance, suppose there are three points  $p_e, p_g, p_{h^{-1}}$  on the Lie group whose corresponding group elements are  $e, g, h^{-1} \in G$ , with  $e$  being the identity element. In a certain time interval, the following left actions occur successively:

	Left-acting element	Point $p_e$	Point $p_g$	Point $p_{h^{-1}}$
		$e$	$g$	$h^{-1}$
↓ ↓ ↓ ↓ ↓	$g^{-1}$	$g^{-1}e = g^{-1}$	$g^{-1}g = e$	$g^{-1}h^{-1} = g^{-1}h^{-1}$
	$h$	$hg^{-1} = hg^{-1}$	$he = h$	$hg^{-1}h^{-1} = hg^{-1}h^{-1}$
	$h^{-1}$	$h^{-1}hg^{-1} = g^{-1}$	$h^{-1}h = e$	$h^{-1}hg^{-1}h^{-1} = g^{-1}h^{-1}$
	$a$	$ag^{-1} = ag^{-1}$	$ae = a$	$ag^{-1}h^{-1} = ag^{-1}h^{-1}$
	$a^{-1}$	$a^{-1}ag^{-1} = g^{-1}$	$a^{-1}a = e$	$a^{-1}ag^{-1}h^{-1} = g^{-1}h^{-1}$

Although the probability of the identity element appearing at each point is equal, during this time interval the number of occurrences of the identity element  $e$  at point  $p_g$  is greater than at the other two points: point  $p_g$  appears three times, point  $p_{h^{-1}}$  not at all. This is similar to flipping a coin: the probabilities of heads and tails are equal, yet over a short period the actual counts of heads and tails are usually not equal.

Because interactions among group elements are random, the frequency with which the identity element appears at different points of the Lie group over a given time interval is also random. Therefore, we need a function to describe the random occurrence counts of the identity element at various points of the Lie group within that time interval; this function is the **probability density function**. Since the Lie group is a continuous group, this probability density function must be a **continuous function** and, like all probability density functions, it must also be **single-valued** and **finite**.

Let this probability density function be  $w(x, y, z, t)$ , which represents the probability of finding the particle (i.e., the identity element  $e$ ) at time  $t$  within an infinitesimal volume  $d\tau$  around point  $(x, y, z)$  of the Lie group. Further, suppose there exists a complex-valued function  $\Phi(x, y, z, t)$  such that

$$w(x, y, z, t)d\tau = C^2 |\Phi(x, y, z, t)|^2 d\tau, \quad (16.3.1)$$

where  $C^2$  is a proportionality constant.  $\Phi$  is called the **wave function**. Since  $\Phi$  is complex, we have  $|\Phi|^2 = \Phi\Phi^*$ ,  $\Phi^*$  is the complex conjugate of  $\Phi$ . The definition of  $\Phi$  is

$$\Phi(x, y, z, t) = \begin{cases} \Phi(x, y, z, t) & \text{inside } G, \\ 0 & \text{outside } G. \end{cases}$$

Since the whole space naturally includes the Lie group  $G$ , the function  $\Phi$  can be regarded as defined over the whole space. However, on the Lie group  $G$ ,  $\Phi$  is non-zero, whereas outside the Lie group  $G$ ,  $\Phi$  equals zero.

Now we determine the constant  $C$ . Because the probability of finding the identity element over the whole space is equal to 1, we integrate equation (16.3.1) over the whole space and require the integral to equal 1:

$$\int w(x, y, z, t) d\tau = \int C^2 |\Phi(x, y, z, t)|^2 d\tau = 1. \quad (16.3.2)$$

We obtain

$$C^2 = \frac{1}{\int |\Phi(x, y, z, t)|^2 d\tau}. \quad (16.3.3)$$

After determining the constant  $C$ , set

$$\psi(x, y, z, t) = C^2 \Phi(x, y, z, t), \quad (16.3.4)$$

here  $\psi(x, y, z, t)$  is also called the **wave function**. Equation (16.3.1) then becomes

$$w(x, y, z, t) d\tau = |\psi(x, y, z, t)|^2 d\tau, \quad (16.3.5)$$

the probability density is

$$w(x, y, z, t) = |\psi(x, y, z, t)|^2. \quad (16.3.6)$$

The meaning of equation (16.3.5) is that **the square of the modulus of the wave function,  $|\psi(x, y, z, t)|^2$ , is proportional to the probability  $w(x, y, z, t)$  of finding the particle in a unit volume at position  $(x, y, z)$  at time  $t$ .**

Rewriting equation (16.3.2) as

$$\int |\psi(x, y, z, t)|^2 d\tau = 1. \quad (16.3.7)$$

A wave function  $\psi$  satisfying (16.3.7) is called a **normalized wave function**. Equation (16.3.7) is called the **normalization condition**. The constant  $C$  determined by equation (16.3.3) is called the **normalization constant**. The process of replacing  $\Phi$  with  $\psi$  is referred to as **normalization**. The variable  $(x, y, z)$  of the wave function  $\psi$  can also be represented by the position vector  $\mathbf{r}$ , so (16.3.7) can be written as

$$\int |\psi(\mathbf{r}, t)|^2 d\tau = 1. \quad (16.3.8)$$

Since the domain of activity of the point-like particle is the local Lie group, which is finite, when  $\mathbf{r} \rightarrow \pm\infty$ , the variable  $\mathbf{r}$  extends beyond the range of the Lie group,  $\psi = 0$ . Consequently, the wave function  $\psi$  is **square-integrable** over the whole space.

Since the probability density function must be single-valued, continuous, and finite, it follows from equation (16.3.6) that the wave function must also be continuous and finite. To ensure the probability density function is single-valued, we require the wave function to be single-valued as well, because if the wave function is single-valued, so is the probability density function. Furthermore, since the wave function is defined on the Lie group, and different points on the Lie group are distinct, we also demand that the wave function be single-valued to reflect the differences among points on the Lie group. In summary, we require the wave function to be **single-valued, continuous, and finite**. All functions defined on the Lie group constitute an infinite-dimensional function space.

## §16.4 Dynamical Variables Are Represented by Linear Hermitian Operators

### 1. Dynamical Variables Represented by Linear Operators

In classical mechanics, when describing the motion of an object, we typically treat it as a point mass. For example, when describing Earth's revolution around the Sun, we regard the large Earth as a point mass and use dynamical variables such as velocity, acceleration, and angular momentum to characterize its orbital motion. According to Example 2.2.1, both velocity and acceleration are first-order tensors; angular momentum is a linear combination of velocity components and is also a first-order tensor. Because we treat the moving object as a point with no size, we use tensors to describe its motion.



For instance, in classical mechanics, if  $f(x, t)$  represents the trajectory of a certain particle, then  $\frac{df(x, t)}{dt}$  is the velocity of that particle. If  $g(x, t)$  is the trajectory of the particle under different circumstances, its velocity is  $\frac{dg(x, t)}{dt}$ . Hence, the particle's velocity can be expressed by the operator  $\frac{d}{dt}$ . Example 2.2.1 has already demonstrated that  $\frac{d}{dt}$  is a first-order contravariant tensor.

However, when we study the internal motion of a particle that includes its field, we can no longer treat the particle as a mere point. Instead, we must regard the particle as a whole—including its field—as a **Lie group**. A Lie group is a continuous group and a smooth manifold composed of infinitely many points. Since a tensor can only describe properties at a single point, we must define a tensor at **every** point of the smooth manifold. These infinitely many tensors together form a **tensor field**. Consequently, we must use a **tensor field** to describe the entire particle, including its field. The advantage of a tensor field is that it provides a **comprehensive description** of the particle as a whole.

Let  $u(x, y, z)$  and  $v(x, y, z)$  be two arbitrary functions, and let  $c_1$  and  $c_2$  be two arbitrary constants. If an operator  $\hat{Q}$  satisfies the following relation:

$$\hat{Q}(c_1 u + c_2 v) = c_1 \hat{Q}u + c_2 \hat{Q}v,$$

then  $\hat{Q}$  is called a **linear operator**.

A tensor field is a **linear mapping** or **linear operator** acting on smooth functions (see Definition 6.1.2 and Theorem 7.6.1). In Lie-group theory, the properties of a smooth manifold are described by tensor fields. Since we regard a particle as a Lie group, we must also use tensor fields to describe the particle's properties. Tensor fields, being linear operators, align perfectly with the quantum-mechanical requirement that dynamical variables be represented by **linear operators**.

For example, when studying the motion of a microscopic particle—including its field—in a central force field, we first treat the particle as a point in the classical-mechanical sense and consider how to describe the point particle's motion in that field. We find that **angular momentum** can be used to describe the point particle. When we then need to describe the motion of the entire microscopic particle including its field, we replace the angular momentum with its corresponding **operator**; the resulting calculations turn out to agree with experiments. This approach is widely adopted in quantum mechanics.

The reason this method is so fruitful is that the microscopic particle, together with its field, is itself a smooth manifold composed of infinitely many points. Replacing a tensor by a **tensor field** successfully shifts the description from a single point particle to a particle consisting of infinitely many points.

Another example: the plane-wave function for a free particle with momentum  $\mathbf{p}$  and energy  $E$  is

$$\psi(\mathbf{r}, t) = A e^{-\frac{i}{\hbar}(Et - \mathbf{r} \cdot \mathbf{p})}. \quad (16.4.1)$$

Taking the partial derivative with respect to time  $t$  on both sides gives

$$i\hbar \frac{\partial}{\partial t} \psi = E \psi. \quad (16.4.2)$$

In classical mechanics,  $E$  describes the total energy of a point particle. However, according to equation (16.4.2), we must use the operator  $i\hbar \frac{\partial}{\partial t}$  to describe the total energy of the particle as a whole, i.e., we must make the substitution

$$E \rightarrow i\hbar \frac{\partial}{\partial t}. \quad (16.4.3)$$

Taking the second partial derivatives of (16.4.1) with respect to  $x, y, z$  yields

$$\frac{\partial^2}{\partial x^2} \psi = -\frac{p_x^2}{\hbar^2} \psi, \quad \frac{\partial^2}{\partial y^2} \psi = -\frac{p_y^2}{\hbar^2} \psi, \quad \frac{\partial^2}{\partial z^2} \psi = -\frac{p_z^2}{\hbar^2} \psi.$$

Adding these three equations gives



$$\left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi = -\frac{p^2}{\hbar^2} \psi \quad \text{or} \quad \nabla^2 \psi = -\frac{p^2}{\hbar^2} \psi ,$$

i.e.,

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = \frac{p^2}{2m} \psi . \quad (16.4.4)$$

In classical mechanics,  $\frac{p^2}{2m}$  describes the kinetic energy of a point particle of mass  $m$ . But according to (16.4.4), we must use the operator  $-\frac{\hbar^2}{2m} \nabla^2$  to describe the kinetic energy of the particle as a whole, i.e., we must make the substitution

$$\frac{p^2}{2m} \rightarrow -\frac{\hbar^2}{2m} \nabla^2 . \quad (16.4.5)$$

In classical mechanics, the total energy  $E$  of a freely moving point particle equals its kinetic energy:

$$E = \frac{p^2}{2m} .$$

In quantum mechanics we must replace both  $E$  and  $\frac{p^2}{2m}$  by operators, giving

$$i\hbar \frac{\partial}{\partial t} \psi = -\frac{\hbar^2}{2m} \nabla^2 \psi . \quad (16.4.6)$$

In classical mechanics, if a point particle is in a field with potential energy  $U(\mathbf{r}, t)$ , its total energy  $E$  is the sum of kinetic and potential energies:

$$E = \frac{p^2}{2m} + U(\mathbf{r}, t) . \quad (16.4.7)$$

In quantum mechanics, using (16.4.3) and (16.4.5) to replace  $E$  and  $\frac{p^2}{2m}$  in (16.4.7), we obtain

$$i\hbar \frac{\partial}{\partial t} \psi = -\frac{\hbar^2}{2m} \nabla^2 \psi + U(\mathbf{r}, t) \psi . \quad (16.4.8)$$

Equation (16.4.8) is the **Schrödinger equation**.

It can be verified that the operator  $i\hbar \frac{\partial}{\partial t}$  satisfies the definition of a tangent vector at every point of a smooth manifold; therefore, it can represent a tangent vector at a point on the smooth manifold. Likewise,  $i\hbar \frac{\partial}{\partial t}$  fulfills the definition of a smooth tangent vector field, so it is a smooth tangent vector field defined on the smooth manifold. Although  $-\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}, t)$  itself does not satisfy the definition of a tangent vector or a smooth tangent vector field, because we equate it to  $i\hbar \frac{\partial}{\partial t}$  via the Schrödinger equation, it also becomes a smooth tangent vector field on the smooth manifold.

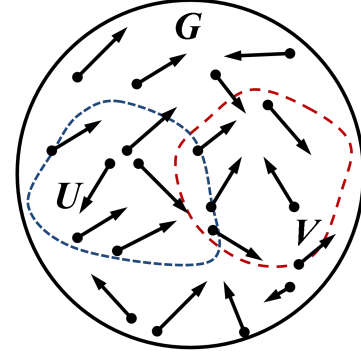
The momentum operators  $-i\hbar \frac{\partial}{\partial x}$ ,  $-i\hbar \frac{\partial}{\partial y}$ ,  $-i\hbar \frac{\partial}{\partial z}$  satisfy the definition of a tangent vector at each point of the smooth manifold; hence, each can represent a tangent vector at a point on the smooth manifold. They also meet the definition of a smooth tangent vector field, so they are smooth tangent vector fields defined on the smooth manifold. Similarly, the angular momentum operators can be regarded as smooth tangent vector fields defined on the smooth manifold.

## 2. Condition for a Dynamical Variable to Have a Definite Value

Since a tensor can be viewed as a multilinear function on the tangent vector space and its dual

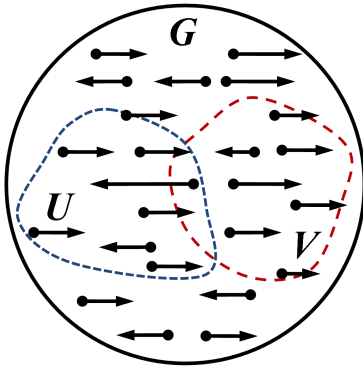
space, and elements of the dual space of the tangent vector space (i.e., the cotangent space) can be regarded as functions on elements of the tangent vector space, and since most operators representing dynamical variables in quantum mechanics can be considered as smooth tangent vector fields defined on a smooth manifold, we will primarily use **smooth tangent vector fields** as an example to discuss the condition under which a dynamical variable has a definite value.

Let  $G$  be the Lie group representing a certain particle.  $G$  is also a smooth manifold. The distribution of a certain tangent vector field  $X$  on this smooth manifold is shown in Figure 16.4.1. Since the point-like particle undergoes random motion, it may appear at the starting point of every arrow in the figure. Suppose this tangent vector field represents the **momentum** of the point-like particle. To measure the momentum, besides observation and instrumentation, we also need computation; therefore, a local coordinate system must be established on the smooth manifold for the measurement. For example, the local coordinate system  $(U; x^i)$  within the left dashed box in Figure 16.4.1. In this local coordinate system, the direction and magnitude of the point-like particle's momentum vary from point to point over the open set  $U$ . Consequently, the measured momentum does **not** have a definite value; measurements taken at different times yield different results.

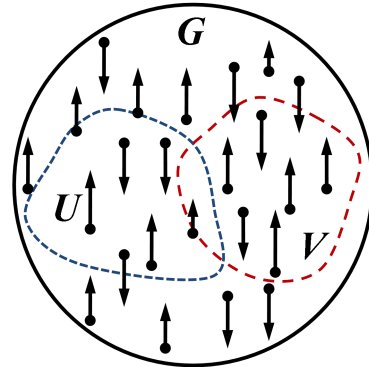


**Figure 16.4.1** Local measurement

If two teams simultaneously measure the momentum of the same point-like particle, they would set up two distinct local coordinate systems, say  $(U; x^i)$  and  $(V; x^i)$ , as illustrated in Figure 16.4.1. In these two local coordinate systems, the direction and magnitude of the point-like particle's momentum differ, so the momentum values they obtain are **not** the same. If this tangent vector field represents other dynamical variables such as energy, the situation is analogous: the measurement results do **not** have a definite value.



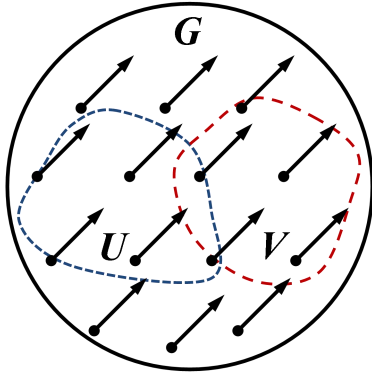
**Figure 16.4.2** The measured value of  $p_x$  is random



**Figure 16.4.3** The measured value of  $p_y$  is random

Each tangent vector can be decomposed into a sum of three components in a local coordinate system; a smooth tangent vector field can naturally be decomposed into the sum of three smooth tangent vector fields. Figure 16.4.2 shows the  $x$ -component  $\hat{p}_x$  of the smooth tangent vector field  $\hat{p}$  representing the momentum of the point-like particle. From the figure, it is evident that the magnitude and direction of  $\hat{p}_x$  vary from point to point on the smooth manifold. Therefore, if we measure the  $x$ -component  $p_x$  of the point-like particle's momentum in the local coordinate system  $(U; x^i)$ , the measured magnitude and direction of  $p_x$  are **uncertain**: at different times, the measured magnitude and direction may sometimes be the same and sometimes different. Similarly, the  $y$ -component  $p_y$  of the point-like particle's momentum measured at different times may also vary in magnitude and direction, as shown in Figure 16.4.3.

When, then, does a measurement yield a definite value? Suppose the distribution of the smooth tangent vector field is as shown in Figure 16.4.4, where the tangent vectors at all points

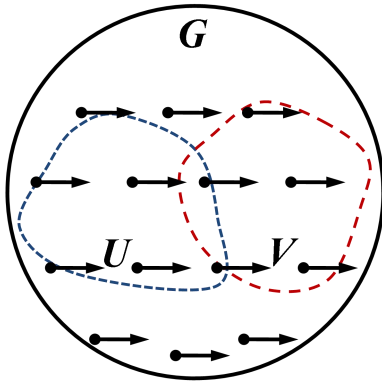


**Figure 16.4.4** The measured value of  $p$  is unique

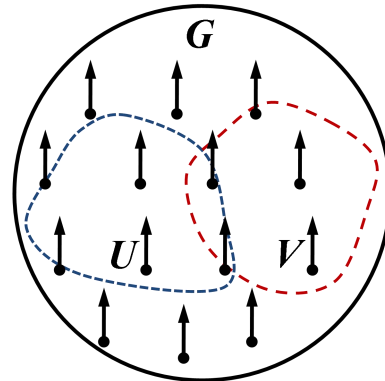
have equal magnitude and the same direction. If this smooth tangent vector field represents the momentum of the point-like particle, then **in any local coordinate system** established on the smooth manifold, the measured momentum of the point-like particle will have a **definite, unique value**; two different measurement teams will obtain the same result.

The  $x$ -component of this smooth tangent vector field is shown in Figure 16.4.5, where the tangent vectors at all points have equal magnitude and the same direction. Therefore, if we measure the  $x$ -component of the point-like particle's momentum, the measured value will be **definite and unique**.

Similarly, the  $y$ -component of this smooth tangent vector field, shown in Figure 16.4.6, also exhibits tangent vectors of equal magnitude and the same direction at all points, so a measurement of the  $y$ -component of the momentum will likewise yield a **definite and unique** value. If this smooth tangent vector field represents the energy of the point-like particle, then the measured energy value will also be **definite and unique**.



**Figure 16.4.5** The measured value of  $p_x$  is unique



**Figure 16.4.6** The measured value of  $p_y$  is unique

Examining Figures 16.4.4 through 16.4.6, we can see that for a measured value to be definite and unique, the smooth tangent vector field representing that dynamical variable must consist of tangent vectors that are **all of equal magnitude and the same direction**; that is, the smooth tangent vector field must be **constant**. The reason is evident: only if the smooth tangent vector field is equivalent to a **constant operator** will all its tangent vectors have equal magnitude and the same direction, yielding a definite and unique value when measured in the same coordinate system. Moreover, we know that a constant is invariant under coordinate transformations—its value does not change when switching from one local coordinate system to another. Therefore, **only when the smooth tangent vector field is equivalent to a constant** will measurements of the same dynamical quantity in different local coordinate systems yield a definite, unique, and identical value.

Most smooth manifolds do **not** admit a tangent vector field whose vectors are everywhere equal in magnitude and parallel in direction. However, for a **particle does** exist, because a particle is a Lie group, and a Lie group or the tangent bundle of a Lie group is **parallelizable** (since the tangent bundle of a Lie group is trivial).

Because the tangent bundle of a Lie group is trivial, when studying the Lie algebra of a Lie group we may focus either on the tangent space at the point-like particle (the identity element) alone, or on the vector space formed by the left- (or right-) invariant vector fields generated by left (or right) translation of elements of the tangent space at the point-like particle over the entire Lie group. In quantum mechanics, particles are regarded as point-like particles, and consequently the Lie algebra of a particle is identified with the tangent space at the point-like particle (the identity element). The resulting conclusions agree remarkably well with experiments, which strongly

encourages the view that a particle is indeed a point. However, the particle itself is both a group and a manifold; therefore, we fail to discover many properties of the particle as a whole and the reasons behind them.

Let the magnitude of this constant vector be  $\lambda$ . Since it is a constant vector, we may focus only on its magnitude, disregarding its direction, and assume the smooth tangent vector field equals a constant  $\lambda$ . Let  $\hat{Q}$  be the smooth tangent vector field shown in Figure 16.4.4, and let  $u(x, y, z)$  be a function defined on the smooth manifold. Then

$$\hat{Q}u(x, y, z) = \lambda u(x, y, z),$$

or

$$\hat{Q}u(\mathbf{r}) = \lambda u(\mathbf{r}). \quad (16.4.9)$$

A smooth tangent vector field is a linear operator acting on functions defined on the smooth manifold, and the constant  $\lambda$  can also be regarded as a linear operator. Equation (16.4.9) is the **eigenvalue equation** for the operator  $\hat{Q}$ , where  $\lambda$  is the eigenvalue and the function  $u$  is the eigenfunction of  $\hat{Q}$ . Therefore, **only when the particle is in an eigenstate  $u$  of the operator  $\hat{Q}$  will a measurement of the dynamical quantity  $Q$  yield a definite, unique value  $\lambda$ .**

The conclusion drawn from the above analysis also applies to cases where dynamical quantities are represented by other types of tensor fields.

### 3. Dynamical Variables Are Represented by Linear Hermitian Operators

The quantity  $\lambda$  obtained from measuring a dynamical variable must be a **real number** to be physically meaningful. Therefore, in equation (16.4.9),  $\lambda$  must be real. This demands that the eigenvalues of the operator  $\hat{Q}$  representing the dynamical variable be real numbers. **Linear Hermitian (self-adjoint) operators** happen to have **real eigenvalues**, which is why we require that dynamical variables of particles be represented by **linear Hermitian operators**.

If, for any two functions  $u(x, y, z)$  and  $v(x, y, z)$ , the operator  $\hat{Q}$  satisfies the equality

$$\int u^*(\mathbf{r}) \hat{Q}v(\mathbf{r}) d\tau = \int v(\mathbf{r}) [\hat{Q}u(\mathbf{r})]^* d\tau, \quad (16.4.10)$$

where the integration is over the whole space and  $*$  denotes complex conjugation, then  $\hat{Q}$  is called a **Hermitian operator**.

**Eigenvalues of a Hermitian operator are real numbers.** For if  $\hat{Q}$  is Hermitian, then from definition (16.4.10) we have

$$\int u^*(\mathbf{r}) \hat{Q}u(\mathbf{r}) d\tau = \int u(\mathbf{r}) [\hat{Q}u(\mathbf{r})]^* d\tau = \left[ \int u^*(\mathbf{r}) \hat{Q}u(\mathbf{r}) d\tau \right]^*. \quad (16.4.11)$$

Substituting the eigenvalue equation (16.4.9) into (16.4.11) gives

$$\lambda \int u^*(\mathbf{r}) u(\mathbf{r}) d\tau = \lambda^* \left[ \int u^*(\mathbf{r}) u(\mathbf{r}) d\tau \right]^*,$$

i.e.,

$$\lambda \int u^*(\mathbf{r}) u(\mathbf{r}) d\tau = \lambda^* \int u(\mathbf{r}) u^*(\mathbf{r}) d\tau,$$

we obtain

$$\lambda = \lambda^*,$$

hence  $\lambda$  is a real number.

Let the operator  $\hat{Q}$  be Hermitian, and let  $u$  be a normalized wave function. Multiplying both sides of (16.4.9) by the complex conjugate  $u^*$  and integrating yields

$$\int u^*(\mathbf{r}) \hat{Q}u(\mathbf{r}) d\tau = \int u^*(\mathbf{r}) \lambda u(\mathbf{r}) d\tau = \lambda. \quad (16.4.12)$$

If operators  $\hat{Q}_1$  and  $\hat{Q}_2$  are both Hermitian, then from (16.4.10) we obtain

$$\begin{aligned}\int u^*(\mathbf{r})(\hat{Q}_1 + \hat{Q}_2)v(\mathbf{r})d\tau &= \int u^*(\mathbf{r})\hat{Q}_1v(\mathbf{r})d\tau + \int u^*(\mathbf{r})\hat{Q}_2v(\mathbf{r})d\tau \\ &= \int v(\mathbf{r})[\hat{Q}_1u(\mathbf{r})]^*d\tau + \int v(\mathbf{r})[\hat{Q}_2u(\mathbf{r})]^*d\tau = \int v(\mathbf{r})[(\hat{Q}_1 + \hat{Q}_2)u(\mathbf{r})]^*d\tau,\end{aligned}$$

Thus we obtain a property of Hermitian operators:

**Property 1:** The sum of two Hermitian operators is still a Hermitian operator.

Let  $g(\mathbf{r})$  be a real-valued function. Then from equation (16.4.10) we have

$$\begin{aligned}\int v(\mathbf{r})[g(\mathbf{r})\hat{Q}u(\mathbf{r})]^*d\tau &= \int v(\mathbf{r})g(\mathbf{r})[\hat{Q}u(\mathbf{r})]^*d\tau = \int g(\mathbf{r})v(\mathbf{r})[\hat{Q}u(\mathbf{r})]^*d\tau \\ &= \int g(\mathbf{r})u^*(\mathbf{r})\hat{Q}v(\mathbf{r})d\tau = \int u^*(\mathbf{r})g(\mathbf{r})\hat{Q}v(\mathbf{r})d\tau,\end{aligned}$$

i.e.,

$$\int u^*(\mathbf{r})g(\mathbf{r})\hat{Q}v(\mathbf{r})d\tau = \int v(\mathbf{r})[g(\mathbf{r})\hat{Q}u(\mathbf{r})]^*d\tau.$$

Comparing this with (16.4.10) shows that  $g(\mathbf{r})\hat{Q}$  is a Hermitian operator. Hence we obtain another property of Hermitian operators:

**Property 2:** The product  $g(\mathbf{r})\hat{Q}$  of a real-valued function  $g(\mathbf{r})$  and a Hermitian operator  $\hat{Q}$  is still a Hermitian operator.

**Example 16.4.1** Regarding the coordinate  $x$  as an operator, prove that the coordinate operator  $x$  is Hermitian.

**Proof:** 
$$\int_{-\infty}^{\infty} u^*(x)xv(x)dx = \int_{-\infty}^{\infty} v(x)[xu(x)]^*dx.$$

Comparing this with equation (16.4.10) shows that  $x$  is a Hermitian operator.

**Example 16.4.2** Prove that  $\frac{\partial}{\partial x}$  is **not** a Hermitian operator.

**Proof:** 
$$\int_{-\infty}^{\infty} u^*(x)\frac{\partial}{\partial x}v(x)dx = \left[u^*(x)v(x)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v(x)\frac{\partial u^*(x)}{\partial x}dx.$$

Since  $u^*(x)$  and  $v(x)$  are square-integrable, the first term on the right-hand side vanishes. Hence,

$$\int_{-\infty}^{\infty} u^*(x)\frac{\partial}{\partial x}v(x)dx = \int_{-\infty}^{\infty} -v(x)\frac{\partial u^*(x)}{\partial x}dx.$$

Comparing this with (16.4.10) shows that the right-hand side carries an extra minus sign; therefore, the operator  $\frac{\partial}{\partial x}$  is **not** Hermitian.

**Example 16.4.3** Prove that  $-i\frac{\partial}{\partial x}$  is a Hermitian operator.

**Proof:** 
$$\begin{aligned}\int_{-\infty}^{\infty} u^*(x)\left(-i\frac{\partial}{\partial x}\right)v(x)dx &= -i\int_{-\infty}^{\infty} u^*(x)\left(\frac{\partial}{\partial x}\right)v(x)dx \\ &= -i\left[u^*(x)v(x)\right]_{-\infty}^{\infty} + i\int_{-\infty}^{\infty} v(x)\left(\frac{\partial u^*(x)}{\partial x}\right)dx.\end{aligned}$$

Because  $u^*(x)$  and  $v(x)$  are square-integrable, the first term on the right-hand side is zero. Thus,

$$\int_{-\infty}^{\infty} u^*(x)\left(-i\frac{\partial}{\partial x}\right)v(x)dx = \int_{-\infty}^{\infty} v(x)\left(-i\frac{\partial}{\partial x}u(x)\right)^*dx.$$

Comparing this with (16.4.10) confirms that the operator  $-i\frac{\partial}{\partial x}$  is Hermitian.

It can be shown that the momentum operator

$$\hat{p} = -i\hbar\nabla = \left( -i\hbar\frac{\partial}{\partial x}, -i\hbar\frac{\partial}{\partial y}, -i\hbar\frac{\partial}{\partial z} \right)$$

is a linear Hermitian operator, and the position vector  $\mathbf{r}$  is also a linear Hermitian operator. By replacing every fundamental dynamical variable  $\mathbf{r}$  and  $\mathbf{p}$  in a classical expression  $Q(\mathbf{r}, \mathbf{p})$  with the corresponding operators  $\mathbf{r}$  and  $-i\hbar\nabla$ , we obtain the operator  $\hat{Q}(\mathbf{r}, -i\hbar\nabla)$  corresponding to the dynamical variable  $Q(\mathbf{r}, \mathbf{p})$ . As long as  $\hat{Q}(\mathbf{r}, -i\hbar\nabla)$  is a linear Hermitian operator, it can serve as the operator representing the dynamical variable  $Q(\mathbf{r}, \mathbf{p})$ .

From a differential-geometry perspective,  $\frac{\partial}{\partial x}$  is a tangent vector field defined on a smooth manifold and is a linear mapping, yet Example 16.4.2 shows that it is **not** a Hermitian operator. Example 16.4.3 demonstrates that  $-i\frac{\partial}{\partial x}$  is a Hermitian operator. Therefore, we regard  $-i\frac{\partial}{\partial x}$  as a tangent vector field on the smooth manifold, because at every point of the smooth manifold it satisfies the definition of a tangent vector on a smooth manifold (see Definition 4.2.1). In fact, **any** tangent vector on a smooth manifold multiplied by  $-i$  fulfills Definition 4.2.1.

According to the eigenvalue equation (16.4.9), a linear Hermitian operator is a linear mapping defined on functions; it maps a function  $u$  to the function  $\lambda u$ . This is analogous to a tensor field on a smooth manifold, so we can regard a linear Hermitian operator as a **tensor field defined on a smooth manifold**. Although we usually define tensors or tensor fields over the real number field, while linear Hermitian operators often contain the imaginary unit  $i$ , as long as we treat the imaginary unit  $i$  as a constant, we can still view linear Hermitian operators as defined over the real number field.

Let the eigenvalue equation of operator  $\frac{\partial}{\partial x}$  be

$$\frac{\partial}{\partial x}u(x, y, z) = \lambda u(x, y, z),$$

where  $\lambda$  is a real number. Since  $\frac{\partial}{\partial x}$  is not Hermitian, we replace  $x$  with  $ix$ , obtaining

$$\begin{aligned} \frac{\partial}{\partial(ix)}u(ix, y, z) &= \lambda u(ix, y, z), \\ -i\frac{\partial}{\partial x}u(ix, y, z) &= \lambda u(ix, y, z). \end{aligned}$$

we see that  $\frac{\partial}{\partial x}$  becomes a **Hermitian operator**  $-i\frac{\partial}{\partial x}$ . Although the function  $u(ix, y, z)$  may become complex-valued,  $\lambda$  remains the original real number; the eigenvalue does not change.

Let us consider the most general case. In a local coordinate system  $(U; x^k)$ , denote  $x^1 = x$ ,  $x^2 = y$ ,  $x^3 = z$ . According to equation (6.1.1), a smooth tangent vector field  $\hat{X}$  can be expressed as

$$\hat{X}|_U = \sum_{k=1}^3 X^k \frac{\partial}{\partial x^k},$$

where the components  $X^k (1 \leq k \leq 3)$  are smooth real-valued functions on  $U$ . Suppose the eigenvalue equation of  $\hat{X}$  is

$$\sum_{k=1}^3 X^k \frac{\partial}{\partial x^k} u(x, y, z) = \lambda u(x, y, z),$$

with  $\lambda$  being a real number. We change the variables  $x, y, z$  in the function  $u(x, y, z)$  to  $ix, iy, iz$ , respectively, while  $X^k$  remain smooth real-valued functions on  $U$ . We then obtain

$$\sum_{k=1}^3 X^k \frac{\partial}{\partial(ix^k)} u(ix, iy, iz) = \lambda u(ix, iy, iz),$$

i.e.,

$$\sum_{k=1}^3 -iX^k \frac{\partial}{\partial x^k} u(ix, iy, iz) = \lambda u(ix, iy, iz).$$

We have already shown that  $-i\frac{\partial}{\partial x}$ ,  $-i\frac{\partial}{\partial y}$ ,  $-i\frac{\partial}{\partial z}$  are Hermitian operators. Therefore, by Property 1 and Property 2 of Hermitian operators, the operator

$$\hat{X} = \sum_{k=1}^3 -iX^k \frac{\partial}{\partial x^k}$$

is a **Hermitian operator**.

By substituting  $x \rightarrow ix$ ,  $y \rightarrow iy$ ,  $z \rightarrow iz$ , we convert the smooth tangent vector field  $\hat{X}$  into a Hermitian operator. The eigenfunction  $u$ , originally a real-valued function, may now become complex-valued, but the eigenvalue  $\lambda$  remains unchanged—it is still the original real number.

Even though the eigenfunction  $u(ix, iy, iz)$  may become complex-valued, we can still regard it as a function defined on the smooth manifold with its domain being the real number field.

Let us now summarize in a general way:

On the Lie group representing a particle, we can define infinitely many complex-valued functions. They can be added, subtracted, and multiplied by arbitrary complex numbers, thereby forming an infinite-dimensional linear space called a **Hilbert space**. The set of smooth functions  $C^\infty(M)$  on the smooth manifold forms a part of this Hilbert space.

A Lie group is also a smooth manifold. On a smooth manifold, a **smooth tangent vector field** is a linear operator from  $C^\infty(M)$  to  $C^\infty(M)$ ; it is a smooth tensor field of type (1,0), also called a **contravariant vector field** (or first-order contravariant tensor field). A **first-order covariant tensor field** (a smooth 1-form) is a smooth tensor field of type (0,1); it acts as a linear functional on smooth tangent vector fields and is thus a linear operator from  $C^\infty(M)$  to  $C^\infty(M)$  on a smooth manifold. A smooth tensor field  $\tau$  of type  $(r,s)$  is a multilinear function of  $r$  covariant tensor fields and  $s$  smooth tangent vector fields; consequently,  $\tau$  is also a linear operator from  $C^\infty(M)$  to  $C^\infty(M)$ . A smooth tensor field of type (0,0) is simply a scalar function, which is likewise a linear operator from  $C^\infty(M)$  to  $C^\infty(M)$ .

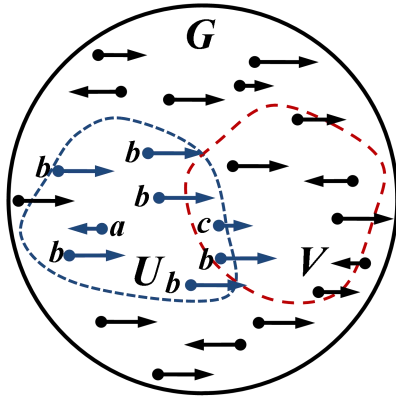
We use a tensor field  $\tau$  to describe the properties of a particle. For the tensor field  $\tau$  to yield a definite, unique measured value, it must be equivalent to a constant—this constant is the **eigenvalue** of  $\tau$ . Since eigenvalues must be measurable, they must be real numbers. Linear operators whose eigenvalues are real numbers are **Hermitian operators**. Therefore, the tensor field  $\tau$  must also be a **linear Hermitian operator**.

## §16.5 Measurement Results of Dynamical Variables

### 1. An Example

As shown in Figure 16.5.1, let the smooth tangent vector field  $\hat{p}_x$  represent the momentum  $p_x$  of the point-like particle. In the local coordinate system  $(U; x^i)$ , a measurement of  $p_x$  does **not** yield a definite value. When the point-like particle appears at point  $a$ , the measured  $p_x$  equals the length of the arrow at point  $a$ , pointing leftward; suppose its magnitude is  $-a$ , i.e.,  $p_x = -a$ . This measurement ( $p_x = -a$ ) occurs only **once**. Let  $u_1(r)$  be the eigenfunction of the point-like particle in this state; then





**Figure 16.5.1** Average measured value of  $p_x$

$$\hat{p}_x u_1(\mathbf{r}) = -a u_1(\mathbf{r}).$$

When the point-like particle appears at point  $b$ , the measured  $p_x$  equals the length of the arrow at point  $b$ , pointing rightward; suppose its magnitude is  $b$ , i.e.,  $p_x = b$ . This measurement ( $p_x = b$ ) occurs **six** times. Let  $u_2(\mathbf{r})$  be the eigenfunction in this state; then

$$\hat{p}_x u_2(\mathbf{r}) = b u_2(\mathbf{r}).$$

When the point-like particle appears at point  $c$ , the measured  $p_x$  equals the length of the arrow at point  $c$ , pointing rightward; its magnitude is  $a$ , i.e.,  $p_x = a$ . This measurement ( $p_x = a$ ) occurs only **once**. Let  $u_3(\mathbf{r})$  be the eigenfunction in this state; then

$$\hat{p}_x u_3(\mathbf{r}) = a u_3(\mathbf{r}).$$

Because the point-like particle appears randomly, the probabilities of obtaining the above  $p_x$  values are as shown in the following table:

Momentum value	Number of occurrences	Probability
$p_x = -a$	1	1/8
$p_x = b$	6	6/8
$p_x = a$	1	1/8

The measured values  $-a, b, a$  are **eigenvalues** of the momentum operator  $\hat{p}_x$ . It is evident that the eigenvalues of  $\hat{p}_x$  are **not unique**. In fact, under other circumstances, the number of eigenvalues of  $\hat{p}_x$  may be finite or infinite; their distribution may be **discrete** (non-continuous) or **continuous**. Therefore, when the eigenvalues are not unique, it is **impossible** to obtain a definite value from a measurement of  $p_x$ .

The point-like particle shown in Figure 16.5.1 may be in one of the three states  $u_1(\mathbf{r}), u_2(\mathbf{r})$ , or  $u_3(\mathbf{r})$  in the local coordinate system  $(U; x^i)$ . Therefore, we form a **linear superposition** of these three states. Let

$$\psi(\mathbf{r}) = c_1 u_1(\mathbf{r}) + c_2 u_2(\mathbf{r}) + c_3 u_3(\mathbf{r}),$$

where  $c_1, c_2, c_3$  are arbitrary constants. The function  $\psi(\mathbf{r})$  is called the **wave function** of the point-like particle, indicating that the point-like particle may be in any one of the states  $u_1(\mathbf{r}), u_2(\mathbf{r})$ , or  $u_3(\mathbf{r})$ . Clearly,  $\psi(\mathbf{r})$  is **not** an eigenstate of the operator  $\hat{p}_x$ , because

$$\hat{p}_x \psi(\mathbf{r}) = \hat{p}_x [c_1 u_1(\mathbf{r}) + c_2 u_2(\mathbf{r}) + c_3 u_3(\mathbf{r})] = -a c_1 u_1(\mathbf{r}) + b c_2 u_2(\mathbf{r}) + a c_3 u_3(\mathbf{r}),$$

and we do **not** obtain a relation of the form  $\hat{p}_x \psi(\mathbf{r}) = \lambda \psi(\mathbf{r})$  with a single eigenvalue  $\lambda$ . Hence, if the point-like particle is in the state  $\psi(\mathbf{r})$ , a measurement of the momentum  $p_x$  will **not** yield a definite value; however, the measured value can be **any one** of  $\{-a, b, a\}$  and cannot be something else. These conclusions can be generalized to the general case.

## 2. Orthonormality

Let the eigenvalue equation of a linear Hermitian operator  $\hat{Q}$  be

$$\begin{aligned} \hat{Q} u_1(\mathbf{r}) &= \lambda_1 u_1(\mathbf{r}), \\ \hat{Q} u_2(\mathbf{r}) &= \lambda_2 u_2(\mathbf{r}), \\ &\vdots \\ \hat{Q} u_n(\mathbf{r}) &= \lambda_n u_n(\mathbf{r}), \\ &\vdots \end{aligned} \tag{16.5.1}$$



According to these eigenvalue equations, the dynamical variable  $Q$  has a definite value  $\lambda_1$  in state  $u_1$ , a definite value  $\lambda_2$  in state  $u_2$ , ..., and a definite value  $\lambda_n$  in state  $u_n$ .

We require the eigenfunctions of  $\hat{Q}$  to be **normalized**, i.e., each eigenfunction  $u_n$  must satisfy

$$\int u_n^* u_n d\tau = 1. \quad (16.5.2)$$

Moreover, for two distinct eigenfunctions  $u_m$  and  $u_n$ , the following holds:

$$\int u_m^* u_n d\tau = 0. \quad (16.5.3)$$

Equation (16.5.2) is called the **normalization** of eigenfunctions, and equation (16.5.3) is called the **orthogonality** of eigenfunctions. Together they are referred to as the **orthonormality** of eigenfunctions.

Let the eigenfunctions of a Hermitian operator  $\hat{Q}$  be

$$u_1, u_2, \dots, u_m, \dots, u_n,$$

all normalized. Then

$$\int u_m^* u_n d\tau = \begin{cases} 0 & \text{when } m \neq n, \\ 1 & \text{when } m = n. \end{cases} \quad (16.5.4)$$

We say that the eigenfunctions of the Hermitian operator  $\hat{Q}$  form an **orthonormal function system**.

### 3.Completeness

Let the eigenfunctions of a Hermitian operator  $\hat{Q}$  be

$$u_1(\mathbf{r}), u_2(\mathbf{r}), \dots, u_n(\mathbf{r}), \dots.$$

Then any function  $\psi(\mathbf{r})$  can be expanded as a linear superposition of the eigenfunctions of  $\hat{Q}$ , i.e.,

$$\psi(\mathbf{r}) = c_1 u_1(\mathbf{r}) + c_2 u_2(\mathbf{r}) + \dots + c_n u_n(\mathbf{r}) + \dots = \sum_n c_n u_n(\mathbf{r}),$$

or

$$\psi(\mathbf{r}) = \sum_n c_n u_n(\mathbf{r}), \quad (16.5.5)$$

where

$$c_n = \int u_n^*(\mathbf{r}) \psi(\mathbf{r}) d\tau. \quad (16.5.6)$$

If in the eigenvalue equation of the operator  $\hat{Q}$ ,

$$\hat{Q} u_\lambda(\mathbf{r}) = \lambda u_\lambda(\mathbf{r}), \quad (16.5.7)$$

the eigenvalue  $\lambda$  takes a continuous range of values, forming a **continuous spectrum**, then equation (16.5.5) should be written in integral form:

$$\psi(\mathbf{r}) = \int c_\lambda u_\lambda(\mathbf{r}) d\lambda, \quad (16.5.8)$$

where

$$c_\lambda = \int u_\lambda^*(\mathbf{r}) \psi(\mathbf{r}) d\tau. \quad (16.5.9)$$

If the operator  $\hat{Q}$  has both discrete and continuous eigenvalues, the expansion of  $\psi(\mathbf{r})$  becomes

$$\psi(\mathbf{r}) = \sum_n c_n u_n(\mathbf{r}) + \int c_\lambda u_\lambda(\mathbf{r}) d\lambda,$$

with the coefficients still given by equations (16.5.6) and (16.5.9).

These two properties of Hermitian-operator eigenfunctions together are stated as: **The eigenfunctions of a Hermitian operator form a complete orthonormal function system.**

#### 4. Meaning of the Expansion Coefficients

In the general case, the wave function  $\psi(\mathbf{r})$  can be expanded using these eigenfunctions:

$$\psi(\mathbf{r}) = \sum_n c_n u_n(\mathbf{r}). \quad (16.5.10)$$

where  $c_n(t)$  are time-dependent coefficients. Taking the complex conjugate of both sides of (16.5.10) gives

$$\psi^*(\mathbf{r}) = \sum_m c_m^* u_m^*(\mathbf{r}), \quad (16.5.11)$$

Substituting equations (16.5.10) and (16.5.11) into the normalization condition (16.3.8) yields

$$1 = \int \psi^*(\mathbf{r}) \psi(\mathbf{r}) d\tau = \sum_{m,n} c_m^* c_n \int u_m^*(\mathbf{r}) u_n(\mathbf{r}) d\tau = \sum_{m,n} c_m^* c_n \delta_{mn} = \sum_n |c_n|^2,$$

i.e.,

$$1 = \int \psi^*(\mathbf{r}) \psi(\mathbf{r}) d\tau = |c_1|^2 + |c_2|^2 + \cdots + |c_n|^2 + \cdots. \quad (16.5.12)$$

The “1” on the far left of equation (16.5.12) represents that the **total probability** of the point-like particle appearing anywhere in the entire integration space is unity. Each term  $|c_n|^2$  on the right-hand side represents the **probability** that the dynamical variable  $Q$  takes the value  $\lambda_n$ . More concretely, the wave function  $\psi(\mathbf{r})$  describes an arbitrary state of the point-like particle. The various possible states are represented by the eigenfunctions  $u_1(\mathbf{r}), u_2(\mathbf{r}), \dots, u_n(\mathbf{r}), \dots$ , respectively. And the wave function  $\psi(\mathbf{r})$  is a linear superposition of these eigenfunctions. The probabilities that the dynamical variable  $Q$  takes the values  $\lambda_1, \lambda_2, \dots, \lambda_n, \dots$  are respectively  $|c_1|^2, |c_2|^2, \dots, |c_n|^2, \dots$ , which are the squared moduli of the coefficients in the expansion of  $\psi(\mathbf{r})$ .

For example, in the scenario discussed above, the wave function  $\psi(\mathbf{r})$  can be written as

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{8}} u_1(\mathbf{r}) + \sqrt{\frac{6}{8}} u_2(\mathbf{r}) + \frac{1}{\sqrt{8}} u_3(\mathbf{r}).$$

When measuring the momentum  $p_x$  of the point-like particle, the probability of obtaining exactly the eigenvalue  $-a$  corresponding to the eigenstate  $u_1(\mathbf{r})$  is  $(1/\sqrt{8})^2 = 1/8$ .

If the expansion of  $\psi(\mathbf{r})$  contains **only one term**, i.e.,

$$\psi(\mathbf{r}) = c_n u_n(\mathbf{r}), \quad (16.5.13)$$

then taking the complex conjugate of (16.5.13) gives

$$\psi^*(\mathbf{r}) = c_n^* u_n^*(\mathbf{r}).$$

Substituting into (16.5.12) yields

$$\int \psi^*(\mathbf{r}) \psi(\mathbf{r}) d\tau = c_n^* c_n \int u_n^*(\mathbf{r}) u_n(\mathbf{r}) d\tau = 1.$$

Since  $u_n(\mathbf{r})$  is normalized, we obtain

$$|c_n|^2 = 1.$$

This means that when the particle is in state  $u_n(\mathbf{r})$ , the probability that the dynamical variable  $Q$  takes the value  $\lambda_n$  is 1. Because  $\lambda_n$  is an observable quantity, the value measured for  $Q$  in state  $u_n(\mathbf{r})$  is precisely  $\lambda_n$ .

When the eigenvalues of  $\hat{Q}$  are continuous, the wave function  $\psi(\mathbf{r})$  must be expanded in integral form:

$$\psi(\mathbf{r}) = \int c_\lambda u_\lambda(\mathbf{r}) d\lambda .$$

The probability that the dynamical variable  $Q$  takes a value in the interval  $[\lambda, \lambda + d\lambda]$  is  $|c_\lambda|^2 d\lambda$ , where

$$c_\lambda = \int u_\lambda^*(\mathbf{r}) \psi(\mathbf{r}) d\tau . \quad (16.5.14)$$

Knowing the probability that the dynamical variable  $Q$  takes the values  $\lambda_1, \lambda_2, \dots, \lambda_n, \dots$ , we can calculate the **average (expectation) value** of  $Q$ :

$$\bar{Q} = |c_1|^2 \lambda_1 + |c_2|^2 \lambda_2 + \dots + |c_n|^2 \lambda_n + \dots .$$

Because

$$\begin{aligned} \int \psi^* \hat{Q} \psi d\tau &= \int \left( \sum_m c_m^* u_m^* \right) \hat{Q} \left( \sum_n c_n u_n \right) d\tau = \sum_{m,n} c_m^* c_n \int u_m^* \hat{Q} u_n d\tau \\ &= \sum_{m,n} c_m^* c_n \lambda_n \int u_m^* u_n d\tau = \sum_{m,n} c_m^* c_n \lambda_n \delta_{mn} = \sum_n |c_n|^2 \lambda_n . \end{aligned} \quad (16.5.15)$$

Therefore, the average value  $\bar{Q}$  can be calculated using the following formula:

$$\bar{Q} = \int \psi^* \hat{Q} \psi d\tau . \quad (16.5.16)$$

Equation (16.5.16) remains valid when the eigenvalue spectrum is continuous.

Suppose we have the normalized wave function  $\psi$ . The average value of the point-like particle's coordinate  $x$  at time  $t$  follows from (16.5.16):

$$\bar{x} = \int \psi^* x \psi d\tau = \int x \psi^* \psi d\tau = \int x |\psi|^2 d\tau .$$

If a dynamical variable  $f$  is a function only of the coordinates  $(x, y, z)$ ,  $f = f(x, y, z)$ , its average value is given by

$$\bar{f} = \int \psi^* f \psi d\tau = \int f \psi^* \psi d\tau = \int f |\psi|^2 d\tau .$$

In summary, once we have the explicit expression of the wave function  $\psi(\mathbf{r})$ , we can not only compute the probabilities for the point-like particle's coordinates to take various possible values, but also calculate the probabilities for **any** dynamical variable  $Q$  to take its possible values. Therefore, determining the explicit form of the wave function  $\psi(\mathbf{r})$  is of fundamental importance. The method to determine  $\psi(\mathbf{r})$  is to represent the dynamical variable  $Q$  by a linear Hermitian operator  $\hat{Q}$ . By solving the eigenvalue equation of  $\hat{Q}$ , we obtain its eigenvalues and eigenfunctions, and consequently we can find the explicit expression of the wave function  $\psi(\mathbf{r})$ .

## 5. Geometric Interpretation of the Expectation-Value Formula

We regard each particle as a Lie group, i.e., as a smooth manifold. A smooth manifold consists of infinitely many points. By defining a tensor at every point of the smooth manifold, these infinitely many tensors form a **tensor field**. Tensor fields provide an excellent description of the particle's properties. In quantum mechanics, most operators representing a dynamical variable  $Q$  can be viewed as **smooth tangent vector fields** defined on the smooth manifold. In the following, we take a smooth tangent vector field as an example to give a geometric interpretation of the expectation-value formula.

As shown in Figure 16.5.2, each arrow (representing the tangent vector at that point) indicates the magnitude and direction of the dynamical variable  $Q$  at that point. We arbitrarily divide the smooth manifold of the particle into  $n$  small regions:

$$\Delta\tau_1, \Delta\tau_2, \dots, \Delta\tau_n,$$

and denote their volumes by  $\Delta\tau_i$  ( $i=1,2,\dots,n$ ). Choose an arbitrary point  $p$  in  $\Delta\tau_i$ , with coordinates  $(x^i, y^i, z^i)$ . Let the magnitude of the tangent vector at point  $p$  be  $\lambda_i$ , and form the product

$$\lambda_i \Delta\tau_i,$$

Summing these products gives the sum

$$\sum_{i=1}^n \lambda_i \Delta\tau_i.$$

Let  $\rho = \max\{d_i\}$  (where  $d_i$  denotes the diameter of  $\Delta\tau_i$ ). Assume that as  $\rho \rightarrow 0$  the limit of the sum exists:

$$\lim_{\rho \rightarrow 0} \sum_{i=1}^n \lambda_i \Delta\tau_i, \quad (16.5.17)$$

and that this limit is independent of the particular partition of  $G$  into small regions  $\Delta\tau_i$  and the choice of points  $p(x^i, y^i, z^i)$ . Then we obtain the integral

$$\lim_{\rho \rightarrow 0} \sum_{i=1}^n \lambda_i \Delta\tau_i = \int_G \lambda(x, y, z) d\tau. \quad (16.5.18)$$

We naturally assume that  $\lambda$  is a continuous function. Hence, by the **mean-value theorem for integrals**, there exists at least one point  $(\xi, \eta, \sigma) \in G$  such that

$$\int_G \lambda(x, y, z) d\tau = \lambda(\xi, \eta, \sigma) \cdot \tau. \quad (16.5.19)$$

From equation (16.5.19) we obtain

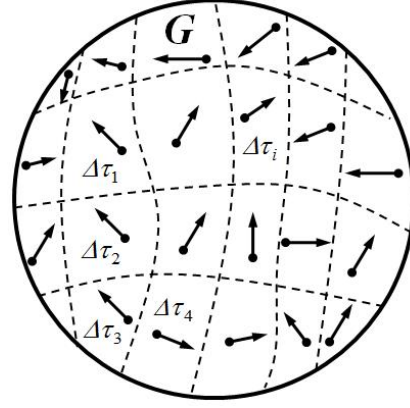
$$\lambda(\xi, \eta, \sigma) = \frac{1}{\tau} \int_G \lambda(x, y, z) d\tau. \quad (16.5.20)$$

$\lambda(\xi, \eta, \sigma)$  is the **average value**  $\bar{Q}$  obtained when measuring the dynamical variable  $Q$ . The reasoning is as follows:

From equations (16.5.18) and (16.5.20) we have

$$\lambda(\xi, \eta, \sigma) = \frac{1}{\tau} \lim_{\rho \rightarrow 0} \sum_{i=1}^n \lambda_i \Delta\tau_i = \lim_{\rho \rightarrow 0} \sum_{i=1}^n \frac{\Delta\tau_i}{\tau} \lambda_i = \sum_{i=1}^n \lim_{\rho \rightarrow 0} \frac{\Delta\tau_i}{\tau} \lambda_i. \quad (16.5.21)$$

From equation (16.5.21),  $\lim_{\rho \rightarrow 0} \frac{\Delta\tau_i}{\tau}$  represents the **proportion** of the smooth manifold  $G$  where the dynamical variable  $Q$  takes the value  $\lambda_i$ . Naturally, this is also the **probability** that a measurement of the dynamical variable  $Q$  yields the value  $\lambda_i$ . For instance, as shown in Figure 16.5.2, the tangent vectors in regions  $\Delta\tau_1, \Delta\tau_2$ , and  $\Delta\tau_3$  are identical, meaning that when measuring the dynamical variable  $Q$  in these three regions, the measured values  $\lambda_1, \lambda_2$ , and  $\lambda_3$  are all equal, i.e.,  $\lambda_1 = \lambda_2 = \lambda_3$ . The tangent vector in region  $\Delta\tau_4$  is unique; no other region has a tangent vector identical to that of region  $\Delta\tau_4$ . Therefore, the probability of measuring the dynamical variable  $Q$  with value  $\lambda_4$  is certainly smaller than the probability of measuring the dynamical variable  $Q$  with value  $\lambda_1$ .



**Figure 16.5.2** Geometric interpretation of the expectation-value formula

Comparing equations (16.5.15) and (16.5.21), we can see that: The term  $\lim_{\rho \rightarrow 0} \frac{\Delta \tau_i}{\tau}$  in equation (16.5.21) corresponds to  $|c_n|^2$  in equation (16.5.15). The term  $\lambda_i$  in equation (16.5.21) corresponds to  $\lambda_n$  in equation (16.5.15).

## §16.6 Conditions for Two Different Dynamical Variables to Have Simultaneous Definite Values

The following two theorems can be proved:

**Theorem 16.6.1** If operators  $\hat{Q}$  and  $\hat{H}$  possess a **common complete set of eigenfunctions**, then  $\hat{Q}$  and  $\hat{H}$  **commute**, i.e.,  $[\hat{Q}, \hat{H}] = 0$ .

**Theorem 16.6.2** If operators  $\hat{Q}$  and  $\hat{H}$  **commute**, i.e.,  $[\hat{Q}, \hat{H}] = 0$ , then  $\hat{Q}$  and  $\hat{H}$  possess a **common complete set of eigenfunctions**.

According to Theorem 16.6.2, if two operators  $\hat{Q}$  and  $\hat{H}$  commute, they share a common complete set of eigenfunctions; consequently, in the same state they can **simultaneously** have definite values. Conversely, if two operators  $\hat{Q}$  and  $\hat{H}$  do **not** commute,  $[\hat{Q}, \hat{H}] \neq 0$ , then in general they **cannot** have simultaneous definite values.

From a differential-geometric perspective, the **Poisson bracket**  $[\hat{Q}, \hat{H}]$  reflects the curvature of the Lie group, i.e., the curvature of the particle as a whole including its field. This can be seen in §6.3 and also from equation (12.8.3). On the other hand, the Poisson bracket  $[\hat{Q}, \hat{H}]$  is essentially the **Lie derivative**

$$L_{\hat{Q}}\hat{H} = [\hat{Q}, \hat{H}].$$

$L_{\hat{Q}}\hat{H}$  is the rate of change of  $\hat{H}$  with respect to  $\hat{Q}$ .  $[\hat{Q}, \hat{H}] = 0$  means that the rate of change of  $\hat{H}$  with respect to  $\hat{Q}$  is zero. Conversely, from  $[\hat{Q}, \hat{H}] = 0$  we also have  $[\hat{H}, \hat{Q}] = 0$ , so the rate of change of  $\hat{Q}$  with respect to  $\hat{H}$  is also zero. Therefore,  $[\hat{Q}, \hat{H}] = 0$  indicates that  $\hat{H}$  and  $\hat{Q}$  are **independent** and do not influence each other; naturally, these two dynamical variables can have simultaneous definite values in the same state.

If  $[\hat{Q}, \hat{H}] \neq 0$ , it means the rate of change of  $\hat{H}$  with respect to  $\hat{Q}$  is **not zero**. Conversely, from  $[\hat{Q}, \hat{H}] \neq 0$  we also have  $[\hat{H}, \hat{Q}] \neq 0$ , so the rate of change of  $\hat{Q}$  with respect to  $\hat{H}$  is also not zero. Therefore,  $[\hat{Q}, \hat{H}] \neq 0$  indicates that  $\hat{H}$  and  $\hat{Q}$  are **not independent** but instead influence each other; consequently, these two dynamical variables **cannot** have simultaneous definite values in the same state.

The relationship between the degrees of uncertainty of two non-commuting operators is precisely the uncertainty principle proposed by Heisenberg.

## §16.7 Uncertainty Principle

Let the commutation relation of operators  $\hat{Q}$  and  $\hat{H}$  be

$$[\hat{Q}, \hat{H}] = i\hat{k},$$

where  $\hat{k}$  is a number or an operator. Then

$$\overline{(\Delta \hat{Q})^2} \cdot \overline{(\Delta \hat{H})^2} \geq \frac{\overline{\hat{k}^2}}{4}, \quad (16.7.1)$$

where  $\overline{(\Delta\hat{Q})^2}$ ,  $\overline{(\Delta\hat{H})^2}$ , and  $\bar{k}^2$  denote the expectation values of  $\hat{Q}$ ,  $\hat{H}$ , and  $\hat{k}$  in the state  $\psi$ . This is the content of the **uncertainty principle**.

Taking the square root of (16.7.1) gives

$$\overline{\Delta\hat{Q}} \cdot \overline{\Delta\hat{H}} \geq \frac{\bar{k}}{2},$$

which is often abbreviated as

$$\Delta Q \cdot \Delta H \geq \frac{\bar{k}}{2}.$$

For example, the commutation relation between position and momentum is

$$x\hat{p}_x - \hat{p}_x x = i\hbar,$$

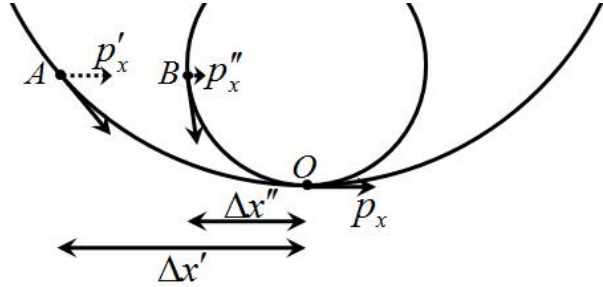
where  $\bar{k} = \hbar$ . Hence,

$$\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}. \quad (16.7.2)$$

Analogously,

$$\Delta y \cdot \Delta p_y \geq \frac{\hbar}{2}, \quad \Delta z \cdot \Delta p_z \geq \frac{\hbar}{2}, \quad \Delta t \cdot \Delta E \geq \frac{\hbar}{2}.$$

The uncertainty principle arises because the particle, as a smooth manifold, is **curved**. We use Figure 16.7.1 to explain the uncertainty principle. As shown, the momentum in the  $x$ -direction at point  $A$  is  $p'_x$ ; the uncertainty in the  $x$ -component of momentum from point  $A$  to point  $O$  is denoted  $\Delta p'_x = p_x - p'_x$ . The momentum in the  $x$ -direction at point  $B$  is  $p''_x$ ; the uncertainty in the  $x$ -component of momentum from point  $B$  to point  $O$  is denoted  $\Delta p''_x = p_x - p''_x$ . Clearly,  $\Delta p'_x < \Delta p''_x$ , but  $\Delta x' > \Delta x''$ . This is precisely the mutual constraint described by the uncertainty principle. The reason for this difference is that the **curvatures** of arc  $AO$  and arc  $BO$  are different: the curvature of arc  $AO$  is **smaller** than that of arc  $BO$ .



**Figure 16.7.1** Why  $x$  and  $p_x$  cannot have simultaneous definite values

denoted  $\Delta p'_x = p_x - p'_x$ . Clearly,  $\Delta p'_x < \Delta p''_x$ , but  $\Delta x' > \Delta x''$ . This is precisely the mutual constraint described by the uncertainty principle. The reason for this difference is that the **curvatures** of arc  $AO$  and arc  $BO$  are different: the curvature of arc  $AO$  is **smaller** than that of arc  $BO$ .

In three dimensions, the curvature of arc  $AO$  can be represented by the **Ricci curvature** along the tangential direction of arc  $AO$ ; similarly, the curvature of arc  $BO$  can be represented by the Ricci curvature along its tangential direction. Because the Ricci curvature  $Ric(v)$  of elementary particles and of particles composed of them has a **strictly positive lower bound** (see Theorem 12.8.4), and these particles are **compact**, it follows from Theorem 11.7.4 that they are also **complete**. Therefore, we can use Theorem 11.7.6 (the **Bonnet–Myers theorem**) to analyze the relationship between the particle's diameter  $d(M)$  and its Ricci curvature  $Ric(v)$ . The following is equation (11.7.1):

$$d(M) \leq \frac{\pi\sqrt{m-1}}{\sqrt{a}}.$$

According to this formula, if the lower bound  $a$  ( $a > 0$ ) of the particle's Ricci curvature is **larger**, then the maximum possible diameter  $d(M)$  is **smaller**. Conversely, if the lower bound  $a$  of the Ricci curvature is **smaller**, the maximum possible diameter  $d(M)$  is **larger**. The uncertainty principle also implies that the Ricci curvature is closely related to the particle's momentum.

Using the uncertainty principle, we can estimate the **extent** of the local Lie group. Taking the one-dimensional case as an example,  $\Delta x$  represents the root-mean-square deviation of the point-like particle's coordinate, i.e., the deviation of the coordinate  $x$  from its average value  $\bar{x}$ :

$$\Delta x = \sqrt{(x - \bar{x})^2} = \sqrt{x^2 - \bar{x}^2},$$

Hence,  $\Delta x$  precisely characterizes the **radius** of the local Lie group. The larger  $\Delta x$  is, the

larger the radius of the local Lie group and, consequently, the larger the domain of activity of the point-like particle; conversely, a smaller  $\Delta x$  implies a smaller domain. In general, since we have  $\Delta x \geq 0$ ,  $\Delta y \geq 0$ , and  $\Delta z \geq 0$ , we may take the **maximum** of these three values.

$$d = \max\{\Delta x, \Delta y, \Delta z\}$$

as the radius of the local Lie group.

$\Delta p_x$  represents the root-mean-square deviation of the point-like particle's momentum, i.e., the deviation of the momentum  $p_x$  from its average value  $\bar{p}_x$ :

$$\Delta p_x = \sqrt{(p_x - \bar{p}_x)^2} = \sqrt{p_x^2 - \bar{p}_x^2}.$$

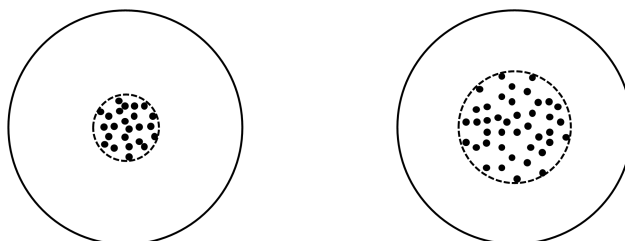
The dependence between  $\Delta x$  and  $\Delta p_x$  is given by equation (16.7.2).

When  $\Delta x$  is large,  $\Delta p_x$  is small; conversely, when  $\Delta p_x$  is small,  $\Delta x$  is large. This indicates that when the local Lie group becomes very large, the interactions among its group elements tend to slow down, so the variation of the tangent vector at each point becomes relatively small, manifesting as a **small**  $\Delta p_x$ . Consequently, the motion of the point-like particle within the local Lie group also tends to become sluggish. This is analogous to placing a fixed number of gas molecules in a very large container: because the container is so large, the collision frequency among the gas molecules becomes very low, resulting in a greatly reduced pressure on the container walls. With a very large  $\Delta x$  and a very small  $\Delta p_x$ , the point-like particle exhibits **very strong wave-like behavior**.

Conversely, when  $\Delta x$  is small,  $\Delta p_x$  is large; or, when  $\Delta p_x$  is large,  $\Delta x$  is small. This shows that when the local Lie group is very small, the interactions among its group elements become **more frequent**, so the variation of the tangent vector at each point becomes very large, manifesting as **large**  $\Delta p_x$ . Consequently, the motion of the point-like particle within the local Lie group also becomes more active. This is analogous to placing a fixed number of gas molecules in a very small container: because the container is so small, collisions among the gas molecules occur very frequently, resulting in a **very high pressure** on the container walls. With a very small  $\Delta x$  and a very large  $\Delta p_x$ , the point-like particle exhibits **very strong particle-like behavior**; in this case, the motion of the point-like particle can be described by **classical mechanics**.

For example, when an electron collides with a proton, the point-like electron may not necessarily hit the point-like proton. However, if a large number of electrons are directed at the proton simultaneously, the chance that a point-like electron strikes the point-like proton increases significantly. Yet, because the point-like proton moves within its local Lie group—i.e., its position changes—incident point-like electrons may hit the point-like proton at different locations. This gives the misleading impression that the point-like proton seems to contain several point-like particles.

When the incident point-like electrons possess very high energy, a large amount of energy is transferred to the point-like proton, causing its  $\Delta p$  to become very large and consequently its  $\Delta x$  to become very small. The point-like proton then moves much more frequently within the tiny  $\Delta x$  range. This further reinforces the misconception that the point-like proton contains countless point-like particles.



**Figure 16.7.2** The domain of activity of the point-like electron forms the volume of the hydrogen atom. A small domain of activity yields a small volume; a large domain of activity yields a large volume

Another example, as shown in Figure 16.7.2, the shape and volume of the hydrogen atom



observed experimentally are actually the shape and volume of the **domain of activity** of the point-like electron within the hydrogen atom.

When experimentalists attempt to measure the radius of a point-like electron by bombarding it with other electrons, they find that the apparent radius of the point-like electron **decreases** as the energy of the incident electrons increases. This occurs because when the incident electrons have very high energy, their impact on the target point-like electron increases the latter's  $\Delta p$ , which in turn reduces its  $\Delta r$ . In short, just as with the hydrogen atom, the observed shape of a point-like particle (if it could be directly observed) is the shape of the **local Lie group** (also called the infinitesimal group), and the measured volume of the point-like particle is also the volume of the local Lie group. Since the local Lie group has no fixed boundary, the volume of the point-like particle is likewise **not fixed**.

## §16.8 Identical Particles

### 1. Homomorphism from the $SU(2)$ Group to the $SO(3)$ Group

First, consider the homomorphism from the  $SU(2)$  group to the  $SO(3)$  group. There exists a homomorphism  $f$  from  $SU(2)$  to  $SO(3)$  whose kernel is  $N(E, -E)$ , where  $E$  is the identity element of  $SU(2)$ . That is, for each element of  $SO(3)$ , there correspond **two** elements of  $SU(2)$ . For example, for the  $SU(2)$  group, one component of a basis in its tangent space at the identity element,  $T_e SU(2)$ , can be chosen as

$$\hat{J}_1 = \frac{1}{2} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

and the corresponding one-parameter subgroup is

$$\begin{aligned} \exp(\alpha \hat{J}_1) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2} \alpha \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} + \frac{1}{2!} \left( \frac{\alpha}{2} \right)^2 \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} + \frac{1}{3!} \left( \frac{\alpha}{2} \right)^3 \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} + \cdots \\ &= \begin{pmatrix} \cos \frac{\alpha}{2} & i \sin \frac{\alpha}{2} \\ i \sin \frac{\alpha}{2} & \cos \frac{\alpha}{2} \end{pmatrix}. \end{aligned}$$

For the  $SO(3)$  group, one component of a basis of its tangent space at the identity,  $T_e SO(3)$ , can be taken as

$$\hat{L}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix},$$

with the corresponding one-parameter subgroup

$$\begin{aligned} \exp(\beta \hat{L}_1) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \beta \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} + \frac{1}{2!} \beta^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} + \cdots \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{pmatrix}. \end{aligned}$$



There exists a homomorphism between  $\exp(\alpha\hat{J}_1)$  and  $\exp(\beta\hat{L}_1)$ , i.e.,

$$f: \begin{pmatrix} \cos \frac{t}{2} & i \sin \frac{t}{2} \\ i \sin \frac{t}{2} & \cos \frac{t}{2} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}.$$

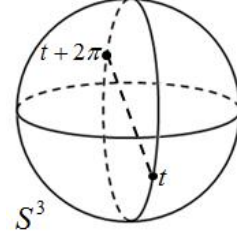


Figure 16.8.1 Antipodal points of  $S^3$

When  $t$  changes to  $t + 2\pi$ , in the  $\exp(\alpha\hat{J}_1)$  group we have

$$\begin{pmatrix} \cos \frac{t}{2} & i \sin \frac{t}{2} \\ i \sin \frac{t}{2} & \cos \frac{t}{2} \end{pmatrix} \rightarrow \begin{pmatrix} \cos \frac{t+2\pi}{2} & i \sin \frac{t+2\pi}{2} \\ i \sin \frac{t+2\pi}{2} & \cos \frac{t+2\pi}{2} \end{pmatrix} = - \begin{pmatrix} \cos \frac{t}{2} & i \sin \frac{t}{2} \\ i \sin \frac{t}{2} & \cos \frac{t}{2} \end{pmatrix}, \quad (16.8.1)$$

so every group element acquires a minus sign, as illustrated in Figure 16.8.1: two antipodal points  $t$  and  $t + 2\pi$  on the diameter of  $S^3$  differ by an angle of  $2\pi$ . In the  $\exp(\beta\hat{L}_1)$  group, however,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(t+2\pi) & -\sin(t+2\pi) \\ 0 & \sin(t+2\pi) & \cos(t+2\pi) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}, \quad (16.8.2)$$

and the group elements remain unchanged.

## 2. Principle of Identity

Particles that are completely identical in all intrinsic properties (such as charge, mass, spin, etc.) are called **identical particles**. For example, all electrons are identical particles because each of them is an  $SU(2)$  group. All protons are identical particles for the same reason—each is an  $SU(2)$  group. All photons are identical particles because each is an  $SO(3)$  group.

**Principle of Identity:** In a system composed of identical particles, exchanging two identical particles does not alter the physical state. This is the principle of identity.

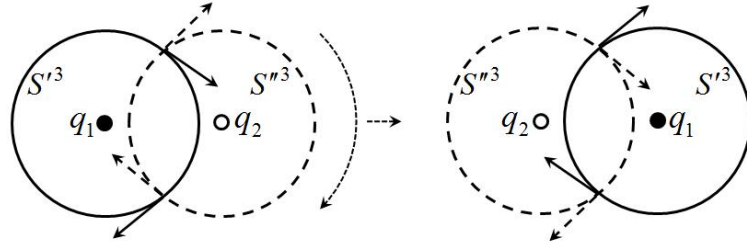
Particles with spin zero or integer spin are called **bosons**. Particles with half-integer spin are called **fermions**.

The wave function of a system of identical fermions must be **antisymmetric**. The wave function of a system of identical bosons must be **symmetric**. Below we illustrate this using electrons and photons as examples.

### (1) Electron System

Consider a system consisting of two electrons that overlap spatially. Let the  $SU(2)$  groups of these two electrons be denoted  $SU(2)'$  and  $SU(2)''$ , and the group  $SU(2)'$  is homeomorphic to the three-sphere  $S^3$ , and the group  $SU(2)''$  is homeomorphic to the three-sphere  $S^3$ . This system can be regarded as the internal direct product group  $SU(2)' \otimes SU(2)''$  of  $SU(2)'$  group and  $SU(2)''$  group. This internal direct product group  $SU(2)' \otimes SU(2)''$  is isomorphic to an  $SU(2)$  group and is also homeomorphic to the unit sphere  $S^3$ . Suppose the state of the two electrons is described by a wave function  $\psi(q_1, q_2)$ , where  $q_1$  represents the coordinate  $x_1, y_1, z_1$  and spin  $S_{z_1}$  of the first electron, and  $q_2$  represents the coordinate  $x_2, y_2, z_2$  and spin  $S_{z_2}$  of the second electron, and  $q_1 \in S^3, q_2 \in S^3$ .

The point  $(q_1, q_2)$  is an element of the internal direct product group  $SU(2)' \otimes SU(2)''$  and also a point on the unit sphere  $S^3$ , with  $(q_1, q_2) \in S^3$ . Therefore, the wave function  $\psi(q_1, q_2)$  is a function defined on the internal direct product group  $SU(2)' \otimes SU(2)''$  and also a function defined on the unit sphere  $S^3$ , as illustrated in Figure 16.8.2.



**Figure 16.8.2** Before exchange

**Figure 16.8.3** After exchange

Exchanging the coordinates  $q_1$  and  $q_2$  in the wave function  $\psi(q_1, q_2)$  corresponds to transforming the point  $(q_1, q_2)$  into the point  $(q_2, q_1)$ . This is equivalent to swapping the positions of the two electrons, which in turn is equivalent to rotating the sphere  $S^3$  by an angle  $2\pi$ , as shown in Figures 16.8.2 and 16.8.3 (compare with Figure 16.8.1). Consequently,

$$(q_1, q_2) \rightarrow (q_1, q_2) + 2\pi = (q_2, q_1).$$

As indicated by equation (16.8.1), when  $t$  changes to  $t + 2\pi$ , the elements of the  $SU(2)$  group change sign. Therefore, when the point  $(q_1, q_2)$  is transformed into  $(q_2, q_1)$ , the wave function  $\psi(q_1, q_2)$  must also change sign:

$$\psi(q_1, q_2) = -\psi(q_2, q_1).$$

Hence, exchanging the coordinates  $q_1$  and  $q_2$  in the wave function  $\psi(q_1, q_2)$  flips its sign, so the wave function  $\psi(q_1, q_2)$  of a system composed of identical electrons is **antisymmetric**. These conclusions hold for all particles with half-integer spin (fermions).

## (2) Photon System

Consider a system consisting of two photons that overlap spatially. Let the  $SO(3)$  groups of these two photons be denoted  $SO(3)'$  and  $SO(3)''$ , and the group  $SO(3)'$  is homeomorphic to the three-sphere  $RP^3$ , and the group  $SO(3)''$  is homeomorphic to the three-sphere  $RP^3$ . This system can be regarded as the internal direct product group  $SO(3)' \otimes SO(3)''$  of the  $SO(3)'$  group and the  $SO(3)''$  group. This internal direct product group  $SO(3)' \otimes SO(3)''$  is isomorphic to an  $SO(3)$  group and is also homeomorphic to the real projective space  $RP^3$ . Suppose the state of the two photons is described by a wave function  $\psi(q_1, q_2)$ , where  $q_1$  represents the coordinate  $x_1, y_1, z_1$  and spin  $S_{z_1}$  of the first photon, and  $q_2$  represents the coordinate  $x_2, y_2, z_2$  and spin  $S_{z_2}$  of the second photon, and  $q_1 \in RP^3, q_2 \in RP^3$ .

The point  $(q_1, q_2)$  is an element of the internal direct product group  $SO(3)' \otimes SO(3)''$  and also a point on the three-dimensional real projective space  $RP^3$ , with  $(q_1, q_2) \in RP^3$ . Therefore, the wave function  $\psi(q_1, q_2)$  is a function defined on the internal direct product group  $SO(3)' \otimes SO(3)''$  and also a function defined on  $RP^3$ .

Identifying antipodal points of  $S^3$  yields  $RP^3$ . For example, as shown in Figure 16.8.1, the point  $t$  and its antipodal point  $t + 2\pi$  are identified. On the diameter of  $S^3$ , the two antipodal points  $q_1$  and  $q_1 + 2\pi = q_2$  are distinct points differing by an angle  $2\pi$ ; but for  $RP^3$ , the antipodal points  $q_1$  and  $q_1 + 2\pi = q_2$  are **the same point**. Exchanging the coordinates  $q_1$  and  $q_2$  in the wave function  $\psi(q_1, q_2)$  corresponds to transforming the point  $(q_1, q_2)$  into  $(q_2, q_1)$ . Transforming  $(q_1, q_2)$  into  $(q_2, q_1)$  is equivalent to rotating the real projective space  $RP^3$  by an angle  $2\pi$ . As indicated by equation (16.8.2), when  $t$  changes to  $t + 2\pi$ , the elements of the  $SO(3)$  group **do not change**. Therefore, exchanging the coordinates  $q_1$  and  $q_2$ , the elements of the  $SO(3)$  group, and consequently all tensor fields defined on the  $SO(3)$  group, remain unchanged. The wave function  $\psi(q_1, q_2)$  defined on the  $SO(3)$  group also remains unchanged. Hence, the

wave function  $\psi(q_1, q_2)$  of a system composed of identical photons is **symmetric**, i.e.,

$$\psi(q_1, q_2) = \psi(q_2, q_1).$$

Although the discussion above focuses on photons, the conclusions also hold for **other bosons besides photons**. Each boson is also an  $SO(3)$  group and is smoothly homeomorphic to  $RP^3$ . An even number of spin-1/2 particles can combine to form a boson; the combination method can be found in §16.13.

## §16.9 The Three Major Equations

In special relativity, the energy  $E$ , momentum  $\mathbf{p}$ , and rest mass  $m$  of a free particle satisfy the relation

$$E^2 = c^2 \mathbf{p}^2 + m^2 c^4.$$

We make the following replacements (quantization prescriptions):

$$E \rightarrow i\hbar \frac{\partial}{\partial t}, \quad \mathbf{p} \rightarrow i\hbar \nabla,$$

and apply them to the wave function  $\psi(\mathbf{r}, t)$ , obtaining the **Klein-Gordon equation** for a free particle:

$$-\hbar^2 \frac{\partial^2}{\partial t^2} \psi = (-\hbar^2 c^2 \nabla^2 + m^2 c^4) \psi. \quad (16.9.1)$$

Taking the **non-relativistic limit** ( $v/c \ll 1$ ) of the Klein-Gordon equation yields the Schrödinger equation. Let

$$\psi(\mathbf{r}, t) = \varphi(\mathbf{r}, t) \exp\left(-\frac{imc^2}{\hbar} t\right),$$

then

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \psi &= \left(i\hbar \frac{\partial \varphi}{\partial t}\right) \exp\left(-\frac{imc^2}{\hbar} t\right) + i\hbar \left(-\frac{imc^2}{\hbar}\right) \varphi \exp\left(-\frac{imc^2}{\hbar} t\right) \\ &= \left(i\hbar \frac{\partial \varphi}{\partial t} + mc^2 \varphi\right) \exp\left(-\frac{imc^2}{\hbar} t\right) \approx mc^2 \varphi \exp\left(-\frac{imc^2}{\hbar} t\right). \quad \left(\because i\hbar \frac{\partial \varphi}{\partial t} \ll mc^2 \varphi\right) \end{aligned}$$

Because

$$\begin{aligned} -\hbar^2 \frac{\partial^2}{\partial t^2} \psi &= \left(i\hbar \frac{\partial}{\partial t}\right)^2 \psi = \left(i\hbar \frac{\partial}{\partial t} + mc^2\right)^2 \psi = -\hbar^2 \frac{\partial^2}{\partial t^2} \psi + 2mc^2 i\hbar \frac{\partial}{\partial t} \psi + m^2 c^4 \psi, \\ -\hbar^2 \frac{\partial^2}{\partial t^2} \psi &= \left(-\hbar^2 \frac{\partial^2 \varphi}{\partial t^2} + 2mc^2 i\hbar \frac{\partial \varphi}{\partial t} + m^2 c^4 \varphi\right) \exp\left(-\frac{imc^2}{\hbar} t\right) \\ &\approx \left(2mc^2 i\hbar \frac{\partial \varphi}{\partial t} + m^2 c^4 \varphi\right) \exp\left(-\frac{imc^2}{\hbar} t\right). \end{aligned}$$

Substituting into equation (16.9.1) gives

$$\left(2mc^2 i\hbar \frac{\partial \varphi}{\partial t} + m^2 c^4 \varphi\right) \exp\left(-\frac{imc^2}{\hbar} t\right) = (-\hbar^2 c^2 \nabla^2 \varphi + m^2 c^4 \varphi) \exp\left(-\frac{imc^2}{\hbar} t\right),$$

i.e.,

$$i\hbar \frac{\partial}{\partial t} \varphi(\mathbf{r}, t) = -\frac{\hbar^2}{2m} \nabla^2 \varphi(\mathbf{r}, t). \quad (16.9.2)$$

Equation (16.9.2) is the **Schrödinger equation** (for a free particle).

The Klein-Gordon equation (16.9.1) can be rewritten as

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 + \frac{m^2 c^2}{\hbar^2}\right) \psi = 0, \quad (16.9.3)$$

or

$$\left[ \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \left( \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + \frac{i^2 m^2 c^2}{\hbar^2} \right) \right] \psi = 0. \quad (16.9.4)$$

This is a differential equation with second-order derivatives in both time and space coordinates. We factorize it into a product of first-order differential operators. Suppose

$$\left[ \frac{1}{c} \frac{\partial}{\partial t} - \left( \alpha_1 \frac{\partial}{\partial x_1} + \alpha_2 \frac{\partial}{\partial x_2} + \alpha_3 \frac{\partial}{\partial x_3} + \frac{imc}{\hbar} \beta \right) \right] \times \left[ \frac{1}{c} \frac{\partial}{\partial t} + \left( \alpha_1 \frac{\partial}{\partial x_1} + \alpha_2 \frac{\partial}{\partial x_2} + \alpha_3 \frac{\partial}{\partial x_3} + \frac{imc}{\hbar} \beta \right) \right] \psi = 0.$$

Expanding and collecting terms gives

$$\left[ \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \sum_{i=1}^3 \alpha_i^2 \frac{\partial^2}{\partial x_i^2} + \frac{m^2 c^2}{\hbar^2} \beta^2 - \sum_{i \neq j} (\alpha_i \alpha_j + \alpha_j \alpha_i) \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} - \frac{imc}{\hbar} \sum_{i=1}^3 (\alpha_i \beta + \beta \alpha_i) \frac{\partial}{\partial x_i} \right] \psi = 0.$$

For this to be equivalent to (16.9.4), the coefficients must satisfy:

$$\begin{aligned} \alpha_i^2 &= \beta^2 = 1 & (i=1,2,3), \\ \alpha_i \alpha_j + \alpha_j \alpha_i &= 0 & (i \neq j; i, j=1,2,3), \\ \alpha_i \beta + \beta \alpha_i &= 0 & (i=1,2,3). \end{aligned}$$

We thus obtain the **Dirac equation**,

$$\left[ \frac{1}{c} \frac{\partial}{\partial t} - \left( \alpha_1 \frac{\partial}{\partial x_1} + \alpha_2 \frac{\partial}{\partial x_2} + \alpha_3 \frac{\partial}{\partial x_3} + \frac{imc}{\hbar} \beta \right) \right] \psi = 0,$$

or

$$i\hbar \frac{\partial \psi}{\partial t} = (-i\hbar c \boldsymbol{\alpha} \cdot \nabla + mc^2 \beta) \psi, \quad (16.9.5)$$

where  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ ,

$$\alpha_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Because the coefficients  $\boldsymbol{\alpha}, \beta$  in equation (16.9.5) are  $4 \times 4$  **matrices**, the wave function  $\psi$  that satisfies (16.9.5) is a **four-component column matrix**, i.e.,

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}.$$

A solution  $\psi$  of the Dirac equation **also** satisfies the Klein-Gordon equation (in the sense that each of its four components satisfies the Klein-Gordon equation). However, a solution of the Klein-Gordon equation does **not** necessarily satisfy the Dirac equation.

## §16.10 Spin Operators of the $SO(3)$ Group

### 1. Matrix Form of the Spin Operators of the $SO(3)$ Group

From equation (12.9.8) we obtain

$$R = g(\theta_1, \theta_2, \theta_3) = \exp[\theta_1 \hat{X}_1 + \theta_2 \hat{X}_2 + \theta_3 \hat{X}_3]$$

$$\begin{aligned}
&= \exp \left[ -i \frac{1}{\hbar} (i\hbar \hat{X}_1 \theta_1 + i\hbar \hat{X}_2 \theta_2 + i\hbar \hat{X}_3 \theta_3) \right] \\
&= \exp \left[ -i \frac{1}{\hbar} (\hat{L}_1 \theta_1 + \hat{L}_2 \theta_2 + \hat{L}_3 \theta_3) \right],
\end{aligned} \tag{16.10.1}$$

where

$$\hat{L}_1 = i\hbar \hat{X}_1 = \hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \tag{16.10.2}$$

$$\hat{L}_2 = i\hbar \hat{X}_2 = \hbar \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \tag{16.10.3}$$

$$\hat{L}_3 = i\hbar \hat{X}_3 = \hbar \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{16.10.4}$$

Their commutation relations are

$$[\hat{L}_1, \hat{L}_2] = i\hbar \hat{L}_3, \quad [\hat{L}_2, \hat{L}_3] = i\hbar \hat{L}_1, \quad [\hat{L}_3, \hat{L}_1] = i\hbar \hat{L}_2. \tag{16.10.5}$$

Let us find the eigenvalues and eigenfunctions of  $\hat{L}_1$ . The eigenvalue equation for  $\hat{L}_1$  is

$$\hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{16.10.6}$$

Simplifying, we obtain

$$\begin{cases} 0x\hbar = \lambda x, \\ -i\hbar z = \lambda y, \\ i\hbar y = \lambda z. \end{cases}$$

From the second and third equations we get  $\lambda = \pm\hbar$ . The solution of the system, i.e., the eigenfunction, is a column matrix

$$\begin{pmatrix} 0 \\ c \\ \pm ic \end{pmatrix},$$

where  $c$  is an arbitrary constant. When  $\lambda = \hbar$ , the third row of the column matrix takes the positive sign; when  $\lambda = -\hbar$ , the third row takes the negative sign.

If  $\lambda = 0$  in eigenvalue equation (16.10.6), then equation (16.10.6) becomes

$$\hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0. \tag{16.10.7}$$

Simplifying, we obtain

$$\begin{cases} 0x = 0, \\ -iz = 0, \\ iy = 0. \end{cases}$$

The solution of the eigenvalue equation, i.e., the eigenfunction, is a column matrix.

$$\begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix},$$

where  $c$  is an arbitrary constant.

Thus, the eigenvalues of  $\hat{L}_1$  are three:  $0, \pm\hbar$ . A similar discussion can be carried out for

$\hat{L}_2$  and  $\hat{L}_3$ ; their eigenvalues are also three:  $0, \pm\hbar$ . Since a photon is assumed to be an  $SO(3)$  group,  $\hat{L}_1, \hat{L}_2$ , and  $\hat{L}_3$  are the **spin operators** of the photon, and the photon's spin is **1**.

The square of the photon's spin angular momentum operator is

$$\hat{S}^2 = \hat{L}_1^2 + \hat{L}_2^2 + \hat{L}_3^2 = \hbar^2 \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The eigenvalue equation is

$$\hbar^2 \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = j \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

i.e.,

$$\begin{cases} 2\hbar^2 x = jx, \\ 2\hbar^2 y = jy, \\ 2\hbar^2 z = jz. \end{cases}$$

Solving gives

$$j = 2\hbar^2.$$

## 2. Differential Form of the Spin Operators of the $SO(3)$ Group

Now let us derive the differential form of the spin operators of the  $SO(3)$  group.

According to equation (12.9.1), the rotation matrix representing a rotation about the x-axis by an angle  $\theta_1$  in terms of the  $SO(3)$  group is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix}. \quad (16.10.8)$$

If  $\theta_1$  is an infinitesimal angle, then using the Taylor expansions

$$\sin \theta = \theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 - \dots, \quad \cos \theta = 1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 - \dots,$$

we obtain

$$\sin \theta_1 \approx \theta_1, \quad \cos \theta_1 \approx 1.$$

Therefore, matrix (16.10.8) becomes

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\theta_1 \\ 0 & \theta_1 & 1 \end{pmatrix}.$$

The coordinate transformation induced by an infinitesimal rotation  $\theta_1$  about the x-axis is

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\theta_1 \\ 0 & \theta_1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (16.10.9)$$

Let  $\psi(x, y, z)$  be a smooth function. Under the coordinate transformation (16.10.9),  $\psi(x, y, z)$  transforms as

$$\psi(x', y', z') = \psi(x, y - \theta_1 z, \theta_1 y + z).$$

According to Theorem 3.1.1, we have

$$\begin{aligned} \psi(x', y', z') &= \psi(x, y - \theta_1 z, \theta_1 y + z) \\ &= \psi(x, y, z) - \theta_1 z \frac{\partial}{\partial y} \psi(x, y, z) + \theta_1 y \frac{\partial}{\partial z} \psi(x, y, z) \end{aligned}$$

$$= \psi(x, y, z) + i\theta_1 \frac{1}{\hbar} \left( i\hbar z \frac{\partial}{\partial y} - i\hbar y \frac{\partial}{\partial z} \right) \psi(x, y, z).$$

Define

$$\hat{L}_x = -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right).$$

Using a similar approach, we obtain

$$\hat{L}_y = -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right), \quad \hat{L}_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right).$$

These  $\hat{L}_x, \hat{L}_y, \hat{L}_z$  are the familiar **angular-momentum operators**. Their commutation relations are

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z, \quad [\hat{L}_y, \hat{L}_z] = i\hbar \hat{L}_x, \quad [\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y.$$

## §16.11 Spin Operators of the $SU(2)$ Group

### 1. Matrix Form of the Spin Operators of the $SU(2)$ Group

From equation (12.10.10) we obtain

$$\begin{aligned} U &= g(\theta_1, \theta_2, \theta_3) = \exp[\theta_1 \hat{X}_1 + \theta_2 \hat{X}_2 + \theta_3 \hat{X}_3] \\ &= \exp \left[ i \frac{1}{\hbar} (-i\hbar \hat{X}_1 \theta_1 - i\hbar \hat{X}_2 \theta_2 - i\hbar \hat{X}_3 \theta_3) \right] \\ &= \exp \left[ i \frac{1}{\hbar} (\hat{S}_1 \theta_1 - \hat{S}_2 \theta_2 + \hat{S}_3 \theta_3) \right], \end{aligned} \quad (16.11.1)$$

where

$$\hat{S}_1 = -i\hbar \hat{X}_1 = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_2 = i\hbar \hat{X}_2 = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{S}_3 = -i\hbar \hat{X}_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (16.11.2)$$

Their commutation relations are

$$[\hat{S}_1, \hat{S}_2] = i\hbar \hat{S}_3, \quad [\hat{S}_2, \hat{S}_3] = i\hbar \hat{S}_1, \quad [\hat{S}_3, \hat{S}_1] = i\hbar \hat{S}_2. \quad (16.11.3)$$

Let us find the eigenvalues and eigenfunctions of  $\hat{S}_1$ . The eigenvalue equation for  $\hat{S}_1$  is

$$\frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}. \quad (16.11.4)$$

Simplifying, we obtain

$$\begin{cases} \frac{\hbar}{2} y = \lambda x, \\ \frac{\hbar}{2} x = \lambda y. \end{cases}$$

Solving gives  $\lambda = \pm \frac{\hbar}{2}$ . The solution of the system, i.e., the eigenfunction, is a column matrix

$$\begin{pmatrix} c \\ \pm c \end{pmatrix},$$

where  $c$  is an arbitrary constant. When  $\lambda = \frac{\hbar}{2}$ , the second row of the column matrix takes the positive sign; when  $\lambda = -\frac{\hbar}{2}$ , the second row takes the negative sign.

Therefore, the eigenvalues of  $\hat{S}_1$  are two:  $\pm \frac{\hbar}{2}$ . A similar discussion can be carried out for

$\hat{S}_2$  and  $\hat{S}_3$ ; their eigenvalues are also two:  $\pm \frac{\hbar}{2}$ . Hence, the infinitesimal generators  $\hat{S}_1, \hat{S}_2$ , and  $\hat{S}_3$  of the  $SU(2)$  group are all **spin- $\frac{1}{2}$  operators**. Since electrons, positrons, protons, antiprotons, neutrinos, and antineutrinos are each assumed to be an  $SU(2)$  group, the spin of these particles is  $\frac{1}{2}$ .

In equation (16.11.2), the matrices inside the three spin operators are precisely the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (16.11.5)$$

It can be verified that the Pauli matrices satisfy the following relations:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = E; \quad (E \text{ is the identity matrix}) \quad (16.11.6)$$

$$\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3, \quad \sigma_2\sigma_3 = -\sigma_3\sigma_2 = i\sigma_1, \quad \sigma_3\sigma_1 = -\sigma_1\sigma_3 = i\sigma_2; \quad (16.11.7)$$

$$[\sigma_1, \sigma_2] = 2i\sigma_3, \quad [\sigma_2, \sigma_3] = 2i\sigma_1, \quad [\sigma_3, \sigma_1] = 2i\sigma_2. \quad (16.11.8)$$

Let  $\mathbf{A}$  and  $\mathbf{B}$  be any two vector operators that commute with  $\boldsymbol{\sigma}$ . Then

$$(\boldsymbol{\sigma} \cdot \mathbf{A})(\boldsymbol{\sigma} \cdot \mathbf{B}) = \mathbf{A} \cdot \mathbf{B} + i\boldsymbol{\sigma} \cdot (\mathbf{A} \times \mathbf{B}). \quad (16.11.9)$$

Proof of equation (16.11.9) is as follows:

$$\begin{aligned} (\boldsymbol{\sigma} \cdot \mathbf{A})(\boldsymbol{\sigma} \cdot \mathbf{B}) &= (\sigma_1 A_1 + \sigma_2 A_2 + \sigma_3 A_3)(\sigma_1 B_1 + \sigma_2 B_2 + \sigma_3 B_3) \\ &= \sigma_1^2 A_1 B_1 + \sigma_1 \sigma_2 A_1 B_2 + \sigma_1 \sigma_3 A_1 B_3 + \sigma_2 \sigma_1 A_2 B_1 + \sigma_2^2 A_2 B_2 + \sigma_2 \sigma_3 A_2 B_3 \\ &\quad + \sigma_3 \sigma_1 A_3 B_1 + \sigma_3 \sigma_2 A_3 B_2 + \sigma_3^2 A_3 B_3 \\ &= A_1 B_1 + A_2 B_2 + A_3 B_3 + i\sigma_3 A_1 B_2 - i\sigma_2 A_1 B_3 - i\sigma_3 A_2 B_1 \\ &\quad + i\sigma_1 A_2 B_3 + i\sigma_2 A_3 B_1 - i\sigma_1 A_3 B_2 \\ &= \mathbf{A} \cdot \mathbf{B} + i\sigma_1 (A_2 B_3 - A_3 B_2) + i\sigma_2 (A_3 B_1 - A_1 B_3) + i\sigma_3 (A_1 B_2 - A_2 B_1) \\ &= \mathbf{A} \cdot \mathbf{B} + i\boldsymbol{\sigma} \cdot (\mathbf{A} \times \mathbf{B}). \end{aligned}$$

## 2. Differential Form of the Spin Operators of the $SU(2)$ Group

Now let us derive the differential form of the spin operators of the  $SU(2)$  group.

### (1) Spin operator $\hat{S}_x$

Consider a vector  $\mathbf{r}$  in two-dimensional complex space, with projections onto the coordinate axes given by  $(u, v)$ . Suppose the vector  $\mathbf{r}$  undergoes a transformation in this two-dimensional complex space analogous to a rotation about the  $x$ -axis in three-dimensional space, with an angle  $\theta_1$ . The corresponding  $SU(2)$  transformation matrix can be obtained from equation (12.10.8) by setting  $\theta_2 = \theta_3 = 0$ , yielding

$$\begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix}. \quad (16.11.10)$$

Assume  $\theta_1$  is an infinitesimal angle; then matrix (16.11.10) becomes

$$\begin{pmatrix} 1 & \frac{i}{2}\theta_1 \\ \frac{i}{2}\theta_1 & 1 \end{pmatrix}.$$

The coordinate transformation induced by an infinitesimal rotation  $\theta_1$  is

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 & \frac{i}{2}\theta_1 \\ \frac{i}{2}\theta_1 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (16.11.11)$$



Let  $\psi(u, v)$  be a smooth function. Under the coordinate transformation (16.11.11),  $\psi(u, v)$  transforms as

$$\begin{aligned}\psi(u', v') &= \psi\left(u + \frac{i}{2}\theta_1 v, \frac{i}{2}\theta_1 u + v\right) \\ &= \psi(u, v) + \frac{i}{2}\theta_1 v \frac{\partial}{\partial u} \psi(u, v) + \frac{i}{2}\theta_1 u \frac{\partial}{\partial v} \psi(u, v) \\ &= \psi(u, v) + i\theta_1 \frac{1}{\hbar} \left( \frac{1}{2}\hbar v \frac{\partial}{\partial u} + \frac{1}{2}\hbar u \frac{\partial}{\partial v} \right) \psi(u, v).\end{aligned}$$

Define

$$\hat{S}_x = \frac{1}{2}\hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \quad (16.11.12)$$

### (2) Spin operator $\hat{S}_y$

Assume the vector  $\mathbf{r}$  undergoes a transformation in this two-dimensional complex space analogous to a rotation about the  $y$ -axis in three-dimensional space, with an angle  $\theta_2$ . The corresponding  $SU(2)$  transformation matrix can be obtained from equation (12.10.8) by setting  $\theta_1 = \theta_3 = 0$ , yielding

$$\begin{pmatrix} \cos \frac{\theta_2}{2} & -\sin \frac{\theta_2}{2} \\ \sin \frac{\theta_2}{2} & \cos \frac{\theta_2}{2} \end{pmatrix}. \quad (16.11.13)$$

Assume  $\theta_2$  is an infinitesimal angle; then matrix (16.11.13) becomes

$$\begin{pmatrix} 1 & -\frac{\theta_2}{2} \\ \frac{\theta_2}{2} & 1 \end{pmatrix}.$$

The coordinate transformation induced by an infinitesimal rotation  $\theta_2$  is

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 & -\frac{\theta_2}{2} \\ \frac{\theta_2}{2} & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (16.11.14)$$

Let  $\psi(u, v)$  be a smooth function. Under the coordinate transformation (16.11.14),  $\psi(u, v)$  transforms as

$$\begin{aligned}\psi(u', v') &= \psi\left(u - \frac{1}{2}\theta_2 v, \frac{1}{2}\theta_2 u + v\right) \\ &= \psi(u, v) - \frac{1}{2}\theta_2 v \frac{\partial}{\partial u} \psi(u, v) + \frac{1}{2}\theta_2 u \frac{\partial}{\partial v} \psi(u, v) \\ &= \psi(u, v) + i\theta_2 \frac{1}{\hbar} \left( \frac{i}{2}\hbar v \frac{\partial}{\partial u} - \frac{i}{2}\hbar u \frac{\partial}{\partial v} \right) \psi(u, v).\end{aligned}$$

Define

$$\hat{S}_y = \frac{i}{2}\hbar \left( v \frac{\partial}{\partial u} - u \frac{\partial}{\partial v} \right). \quad (16.11.15)$$

### (3) Spin operator $\hat{S}_z$

Assume the vector  $\mathbf{r}$  undergoes a transformation in this two-dimensional complex space analogous to a rotation about the  $z$ -axis in three-dimensional space, with an angle  $\theta_3$ . The corresponding  $SU(2)$  transformation matrix can be obtained from equation (12.10.8) by setting  $\theta_1 = \theta_2 = 0$ , yielding

$$\begin{pmatrix} e^{i\frac{\theta_3}{2}} & 0 \\ 0 & e^{-i\frac{\theta_3}{2}} \end{pmatrix}. \quad (16.11.16)$$

Assume  $\theta_3$  is an infinitesimal angle. Since

$$e^\theta = 1 + \theta + \frac{1}{2!}\theta^2 + \dots,$$

we have

$$e^{i\frac{\theta_3}{2}} \approx 1 + i\frac{\theta_3}{2},$$

Therefore, matrix (16.11.16) becomes

$$\begin{pmatrix} 1 + \frac{i}{2}\theta_3 & 0 \\ 0 & 1 - \frac{i}{2}\theta_3 \end{pmatrix}.$$

The coordinate transformation induced by an infinitesimal rotation  $\theta_3$  is

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 + \frac{i}{2}\theta_3 & 0 \\ 0 & 1 - \frac{i}{2}\theta_3 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (16.11.17)$$

Let  $\psi(u, v)$  be a smooth function. Under the coordinate transformation (16.11.17),  $\psi(u, v)$  transforms as

$$\begin{aligned} \psi(u', v') &= \psi\left(u + \frac{i}{2}\theta_3 u, v - \frac{i}{2}\theta_3 v\right) \\ &= \psi(u, v) + \frac{i}{2}\theta_3 u \frac{\partial}{\partial u} \psi(u, v) - \frac{i}{2}\theta_3 v \frac{\partial}{\partial v} \psi(u, v) \\ &= \psi(u, v) + i\theta_3 \frac{1}{\hbar} \left( \frac{1}{2} \hbar u \frac{\partial}{\partial u} - \frac{1}{2} \hbar v \frac{\partial}{\partial v} \right) \psi(u, v). \end{aligned}$$

Define

$$\hat{S}_z = \frac{1}{2} \hbar \left( u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v} \right). \quad (16.11.18)$$

The operators  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  are the familiar **spin angular-momentum operators** for spin- $\frac{1}{2}$  particles. Their commutation relations are

$$[\hat{S}_x, \hat{S}_y] = i\hbar \hat{S}_z, \quad [\hat{S}_y, \hat{S}_z] = i\hbar \hat{S}_x, \quad [\hat{S}_z, \hat{S}_x] = i\hbar \hat{S}_y. \quad (16.11.19)$$

## §16.12 Orbits of Point-Like Particles and Quantum Entanglement

When two or more particles combine, interactions necessarily arise. There are three modes of interaction: the first is the **internal direct product group** approach, which corresponds to the linear superposition of particles—this was already discussed in Chapter 15; the second is the mode where one particle acts as a **transformation group** on the other; the third involves **mutual transformation** of particles (see Chapter 21). This section describes the second mode.

### 1. Orbits of Point-Like Particles

Since elementary particles are Lie groups, the interaction between two or more particles can be regarded as the **direct product** of the Lie groups representing the particles. The interacting particle system thus forms a direct-product Lie group. A Lie group is also a transformation group; therefore, we can treat one particle as a **Lie transformation group** and the other particle as a

**smooth manifold.** The interaction between the two particles is then viewed as the action of the Lie transformation group on the smooth manifold. We illustrate this approach using a system composed of two particles.

Let the Lie group of particle  $A$  be  $G_A$  and that of particle  $B$  be  $G_B$ . These two particles together form a direct-product Lie group  $G_A \times G_B$ . Both  $G_A$  and  $G_B$  are Lie groups and can serve as Lie transformation groups.  $G_A \times G_B$  can be regarded as particle  $A$  **left-acting** on particle  $B$ :  $\varphi: G_A \times G_B \rightarrow G_B$ ; similarly,  $G_A \times G_B$  can be regarded as particle  $B$  **right-acting** on particle  $A$ :  $\omega: G_A \times G_B \rightarrow G_A$ .

The set of points that can be reached by the identity element  $e_B$  of Lie group  $G_B$  under the left (or right) action of all elements of Lie group  $G_A$  is called the **orbit** of  $G_A$  through the point  $e_B$ :

$$O_{e_B} = G_A e_B = \{g_A e_B \mid \forall g_A \in G_A\}, \quad O_{e_B} \subset G_B.$$

If the elements of  $G_A$  act on  $G_B$  **discontinuously** and **randomly**, then the orbit of point-like particle  $B$  (the identity element  $e_B$  of particle  $B$ ) is **chaotic** and **random**. However, if the elements of  $G_A$  act on  $G_B$  **continuously** and in a **deterministic order**, point-like particle  $B$  still undergoes random motion—unless the influence of the transformation group  $G_A$  on the position of point-like particle  $B$  exceeds the influence of the internal elements of particle  $B$  itself on the position of the point-like particle  $B$ . In that case, on a larger scale (i.e., neglecting the effect of particle  $B$ 's internal elements on its point-like particle), the orbit of point-like particle  $B$  follows the sequence determined by the continuous elements of the transformation group  $G_A$ .

Another scenario is as follows: although the point-like particle of an electron performs irregular random motion inside the electron, when a free electron enters a strong, uniform magnetic field, the **Lorentz force** acting on the point-like particle becomes greater than the force exerted by the internal elements of the electron on its point-like particle. Consequently, the point-like particle executes **circular motion** along a **fixed orbit** in the uniform magnetic field.

## 2. Quantum Entanglement

Since the spin operators of particle  $A$  and particle  $B$  are respectively smooth tangent vector fields on the groups  $G_A$  and  $G_B$ , on the direct-product Lie group  $G_A \times G_B$  the spin operators of particle  $A$  and particle  $B$  can be **added**. Let the spin operators of particle  $A$  be  $\hat{S}_{Ax}, \hat{S}_{Ay}, \hat{S}_{Az}$  and those of particle  $B$  be  $\hat{S}_{Bx}, \hat{S}_{By}, \hat{S}_{Bz}$ . Then

$$\hat{S}_{Ax} + \hat{S}_{Bx}, \quad \hat{S}_{Ay} + \hat{S}_{By}, \quad \hat{S}_{Az} + \hat{S}_{Bz} \quad (16.12.1)$$

are smooth tangent vector fields on the direct-product Lie group  $G_A \times G_B$  (see §5.3).

In general, let  $\hat{X}_A$  be a smooth tangent vector field of particle  $A$ , and  $\hat{X}_B$  a smooth tangent vector field of particle  $B$ . Then  $\hat{X} = \hat{X}_A + \hat{X}_B$  is a smooth tangent vector field on the direct-product Lie group  $G_A \times G_B$ . Since  $\hat{X}_A$  and  $\hat{X}_B$  commute,

$$[\hat{X}_A, \hat{X}_B] = \hat{X}_A \hat{X}_B - \hat{X}_B \hat{X}_A = 0,$$

we have

$$\begin{aligned} [\hat{X}, \hat{X}_A] &= [\hat{X}_A + \hat{X}_B, \hat{X}_A] = (\hat{X}_A + \hat{X}_B) \hat{X}_A - \hat{X}_A (\hat{X}_A + \hat{X}_B) \\ &= \hat{X}_A \hat{X}_A + \hat{X}_B \hat{X}_A - \hat{X}_A \hat{X}_A - \hat{X}_A \hat{X}_B = 0. \end{aligned}$$

Similarly,

$$[\hat{X}, \hat{X}_B] = 0.$$

Therefore,  $\hat{X}_A$  and  $\hat{X}_B$  each commute with  $\hat{X}$ ; consequently,  $\hat{X}_A, \hat{X}_B$ , and  $\hat{X}$  can have simultaneous definite values in a certain state and share a common eigenfunction.

Let the eigenfunction of  $\hat{X}$  be  $\psi(\mathbf{r})$  with eigenvalue  $\lambda$ , then

$$\hat{X}\psi(\mathbf{r}) = (\hat{X}_A + \hat{X}_B)\psi(\mathbf{r}) = \lambda\psi(\mathbf{r}).$$

If  $\lambda = 0$ , we obtain

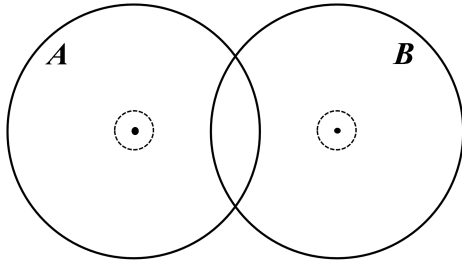
$$\hat{X}_A\psi(\mathbf{r}) = -\hat{X}_B\psi(\mathbf{r}).$$

Let the eigenvalue of  $\hat{X}_A$  be  $\gamma$ ; then

$$\hat{X}_A\psi(\mathbf{r}) = \gamma\psi(\mathbf{r}), \quad (16.12.2)$$

$$\hat{X}_B\psi(\mathbf{r}) = -\gamma\psi(\mathbf{r}). \quad (16.12.3)$$

For example, two **entangled photons**  $A$  and  $B$ —i.e., two interacting photons—form a direct-product Lie group  $G_A \times G_B$ . Assume  $\hat{X}$  is the spin operator of the  $z$ -component on this direct-product Lie group,  $\hat{X}_A$  is the spin operator of the  $z$ -component of photon  $A$ , and  $\hat{X}_B$  is that of photon  $B$ . Suppose the total spin  $z$ -component of the two-photon system is zero, i.e.,  $\lambda = 0$ . According to equations (16.12.2) and (16.12.3), if a measurement on photon  $A$  finds its spin pointing **up**, then a measurement on photon  $B$  will find its spin pointing **down**. Conversely, if a measurement on photon  $A$  finds its spin pointing **down**, then a measurement on photon  $B$  will find its spin pointing **up**.



**Figure 16.12.1** Quantum entanglement

shown in Figure 16.12.1. In this situation, a change in one photon inevitably affects the other. The field elements of photon  $A$  will affect all of photon  $B$ , and conversely, the field elements of photon  $B$  will also affect all of photon  $A$ .

The reason why such correlations appear between two entangled photons  $A$  and  $B$  even when they are separated by more than a thousand kilometers is that, viewed as a whole, a photon is an  $SO(3)$  group consisting of a point-like particle and a **field space**. Although the extent of the field space is finite, it is **very large**, and photons are expanding at the speed of light. Given sufficient time, the fields (the field-space parts) of these two distant photons will **overlap**, forming a direct-product Lie group, as

### §16.13 Forming the Projective Space $RP^3$

An elementary particle with spin-1/2 is a three-dimensional unit sphere. However, **two** spin-1/2 elementary particles can combine to form a projective space  $RP^3$ . Since a photon is precisely an  $RP^3$ , a **pair** of spin-1/2 elementary particles can constitute a photon. This section explores how to assemble  $RP^3$  from spin-1/2 elementary particles. The method can be divided into the following two approaches, which we discuss in four-dimensional Euclidean space  $R^4$ .

#### 1. Antiparallel-Oriented Particles Form $RP^3$

Assume particle  $A$  is a spin-1/2 elementary particle. In  $R^4$ , let the coordinates describing particle  $A$  be  $(x, y, z, w)$ . Set  $a = x + iy$ ,  $b = z + iw$ , where  $i$  is the imaginary unit. Then an element of the  $SU(2)_A$  group describing particle  $A$  can be written as

$$A(x, y, z, w) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} = \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}, \quad (16.13.1)$$

where

$$x^2 + y^2 + z^2 + w^2 = 1.$$

Now suppose particle  $B$  is a spin-1/2 elementary particle whose **orientation** is opposite to that of particle  $A$ . Because particle  $B$  has the opposite orientation to particle  $A$ , we may set its coordinates in  $R^4$  as  $(-x, y, z, w)$ . Consequently, an element of the  $SU(2)_B$  group describing particle  $B$  can be expressed as

$$B(x, y, z, w) = \begin{pmatrix} -x + iy & z + iw \\ -z + iw & -x - iy \end{pmatrix}. \quad (16.13.2)$$

Based on whether the elements of group  $SU(2)_A$  and group  $SU(2)_B$  are inverse elements and whether they are antipodal elements, the elements of these two groups are classified into the following four categories:

$$\begin{aligned} A &= \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}, & -A &= \begin{pmatrix} -x - iy & -z - iw \\ z - iw & -x + iy \end{pmatrix}, \\ A^{-1} &= \begin{pmatrix} x - iy & -z - iw \\ z - iw & x + iy \end{pmatrix}, & -A^{-1} &= \begin{pmatrix} -x + iy & z + iw \\ -z + iw & -x - iy \end{pmatrix}; \\ B &= \begin{pmatrix} -x + iy & z + iw \\ -z + iw & -x - iy \end{pmatrix}, & -B &= \begin{pmatrix} x - iy & -z - iw \\ z - iw & x + iy \end{pmatrix}, \\ B^{-1} &= \begin{pmatrix} -x - iy & -z - iw \\ z - iw & -x + iy \end{pmatrix}, & -B^{-1} &= \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}. \end{aligned}$$

**(1) Condition for Forming the Real Projective Space  $RP^3$**

When particle  $A$  and particle  $B$  combine to form a composite particle, the requirement that the composite particle be **three-dimensional** implies that their two Lie groups should form an **internal direct product group**  $SU(2)_A \otimes SU(2)_B$ . Elements of the internal direct product group are formed by **pairing** elements of group  $SU(2)_A$  with elements of group  $SU(2)_B$  in a one-to-one correspondence. Previously we classified the elements of these two groups into four categories each. Based on this classification, the possible ways for elements of  $SU(2)_A$  and  $SU(2)_B$  to combine into an element of this internal direct product group are the following 16 types:

$$\begin{aligned} (A, B), & \quad (-A, B), & (A^{-1}, B), & \quad (-A^{-1}, B), \\ (A, -B), & \quad (-A, -B), & (A^{-1}, -B), & \quad (-A^{-1}, -B), \\ (A, B^{-1}), & \quad (-A, B^{-1}), & (A^{-1}, B^{-1}), & \quad (-A^{-1}, B^{-1}), \\ (A, -B^{-1}), & \quad (-A, -B^{-1}), & (A^{-1}, -B^{-1}), & \quad (-A^{-1}, -B^{-1}). \end{aligned}$$

However, for an element of this internal direct product group to become an element of group  $RP^3$ , it must satisfy the following two requirements:

**First**, an element of the internal direct product group **cannot** be composed of **identical** elements from group  $SU(2)_A$  and group  $SU(2)_B$ . For example, the combination

$$(A, -B^{-1}) = \left( \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}, \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix} \right)$$

uses the same group element from both groups; this is equivalent to taking the internal direct product of a group with itself. Such a combination yields an internal direct product group **homeomorphic** to  $SU(2)$ , not to  $SO(3)$ , and therefore cannot form the real projective space  $RP^3$ . The other three combinations that also use identical elements from  $SU(2)_A$  and  $SU(2)_B$  are  $(-A, B^{-1})$ ,  $(A^{-1}, -B)$ , and  $(-A^{-1}, B)$ .

**Second**, it must be ensured that the two group elements forming the internal direct product group have the **same orientation**.

Since both  $SU(2)_A$  and  $SU(2)_B$  are **orientable** Lie groups, by Theorem 9.5.1 their internal direct product group  $SU(2)_A \otimes SU(2)_B$  is also orientable. Before forming the internal direct product, the elements of  $SU(2)_A$  and  $SU(2)_B$  belong to distinct Lie groups and their orientations are **opposite**. However, after they combine into an element of the internal direct product group  $SU(2)_A \otimes SU(2)_B$ , they become elements of that single group; therefore, the two group elements must have the **same orientation** within the internal direct product group. If their

orientations are not the same, they cannot combine into an element of an orientable smooth manifold. For example, for the combination

$$(A, -B) = \left( \begin{pmatrix} x+iy & z+iw \\ -z+iw & x-iy \end{pmatrix}, \begin{pmatrix} x-iy & -z-iw \\ z-iw & x+iy \end{pmatrix} \right),$$

we may consider that in the internal direct product group, the coordinates of element  $A$  are  $(x, y, z, w)$ , and the coordinates of element  $-B$  are  $(x, -y, -z, -w)$  as well. For element  $A$ , since its coordinates are  $(x, y, z, w)$ , we have the 4-form

$$dx \wedge (dy \wedge dz \wedge dw). \quad (16.13.3)$$

Because element  $A$  is an element of the three-dimensional  $SU(2)_A$  group, its **orientation** can be determined by the 3-form  $dy \wedge dz \wedge dw$  in equation (16.13.3), while the 1-form  $dx$  in (16.13.3) can be regarded as determining the direction of the **normal vector** at element  $dx$ . For element  $-B$ , its coordinates are also  $(x, -y, -z, -w)$ , giving the 4-form

$$dx \wedge (-dy) \wedge (-dz) \wedge (-dw) = dx \wedge (-dy \wedge dz \wedge dw). \quad (16.13.4)$$

Element  $-B$  is also an element of a three-dimensional group. According to (16.13.4), in order for the normal-vector direction at element  $-B$  to be the **same** as that at element  $A$  (both determined by  $dx$ ), the orientation of element  $-B$  can only be given by the 3-form  $-dy \wedge dz \wedge dw$  in (16.13.4). Consequently, the orientations of element  $A$  and element  $-B$  in the internal direct product group  $SU(2)_A \otimes SU(2)_B$  are **opposite**. Therefore, element  $A$  and element  $-B$  **cannot** be paired together to form an element of an **orientable** smooth manifold  $SU(2)_A \otimes SU(2)_B$ .

Combination modes that **cannot** make the two group elements forming the internal direct product group have the same orientation also include the following seven:

$$\begin{aligned} &(-A, B), \quad (A, B), \quad (-A, -B), \quad (A^{-1}, B^{-1}), \\ &(-A^{-1}, -B^{-1}), \quad (A^{-1}, -B^{-1}), \quad (-A^{-1}, B^{-1}). \end{aligned}$$

After excluding the preceding 12 unsuitable combination modes, only the following four remain:

$$(A, B^{-1}), \quad (-A, -B^{-1}); \quad (A^{-1}, B), \quad (-A^{-1}, -B). \quad (16.13.5)$$

These four combination modes satisfy the two requirements stated above. In fact, the first two are essentially **the same** combination mode because they are antipodal points of each other; similarly, the last two are also the same combination mode because they too are antipodal points. Below we first analyze the first combination mode.

## (2)The First Way to Form $RP^3$

We pair element  $A$  of particle  $A$  with element  $B^{-1}$  of particle  $B$  to form an element of the internal direct product group  $SU(2)_A \otimes SU(2)_B$ :

$$(A, B^{-1}) = \left( \begin{pmatrix} x+iy & z+iw \\ -z+iw & x-iy \end{pmatrix}, \begin{pmatrix} -x-iy & -z-iw \\ z-iw & -x+iy \end{pmatrix} \right), \quad (16.13.6)$$

then

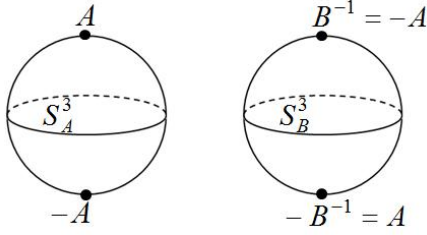
$$(-A, -B^{-1}) = \left( \begin{pmatrix} -x-iy & -z-iw \\ z-iw & -x+iy \end{pmatrix}, \begin{pmatrix} x+iy & z+iw \\ -z+iw & x-iy \end{pmatrix} \right). \quad (16.13.7)$$

Since  $SU(2)_A$  and  $SU(2)_B$  are distinct groups, the product of their elements is **commutative**, i.e.,  $SU(2)_A \otimes SU(2)_B = SU(2)_B \otimes SU(2)_A$ . Hence,

$$(A, B^{-1}) = (B^{-1}, A) = (-A, -B^{-1}),$$

i.e.,

$$(A, B^{-1}) = (-A, -B^{-1}). \quad (16.13.8)$$



**Figure 16.13.1** The first way to form  $RP^3$

As shown in Figure 16.13.1, the group  $SU(2)_A$  is homeomorphic to the unit sphere  $S_A^3$ , and the group  $SU(2)_B$  is homeomorphic to the unit sphere  $S_B^3$ . The point  $A$  and the point  $-A$  on  $S_A^3$  are **antipodal points**; similarly, the point  $B^{-1}$  and the point  $-B^{-1}$  on  $S_B^3$  are also antipodal points (we identify the group  $SU(2)_A$

with the sphere  $S_A^3$ , so  $A$  denotes both an element of  $SU(2)_A$  and a point on  $S_A^3$ . Likewise, we identify  $SU(2)_B$  with  $S_B^3$ , so  $B^{-1}$  denotes both an element of  $SU(2)_B$  and a point on  $S_B^3$ ). The point  $A$  on  $S_A^3$  and the point  $B^{-1}$  on  $S_B^3$  together form an element  $(A, B^{-1})$  of the internal direct product group  $SU(2)_A \otimes SU(2)_B$ . The point  $-A$  on  $S_A^3$  and the point  $-B^{-1}$  on  $S_B^3$  together form another element  $(-A, -B^{-1})$  of the same group  $SU(2)_A \otimes SU(2)_B$ . The points  $(A, B^{-1})$  and  $(-A, -B^{-1})$  are **antipodal points**, but according to equation (16.13.8), these two points are **identified**; i.e., these two antipodal points are **assimilated** into the same point.

Since both  $SU(2)_A$  and  $SU(2)_B$  are **orientable** Lie groups, their internal direct product group  $SU(2)_A \otimes SU(2)_B$  is also orientable. Before being paired, element  $A$  and element  $B^{-1}$  belong to different Lie groups and have **opposite** orientations. However, after they combine into an element of the internal direct product group  $SU(2)_A \otimes SU(2)_B$ , they become elements of that single group; therefore, the two elements  $A$  and  $B^{-1}$  should have the **same orientation** within the internal direct product group. According to equation (16.13.6), we may regard the coordinates of element  $A$  in the internal direct product group as  $(x, y, z, w)$  and the coordinates of element  $B^{-1}$  as  $(-x, -y, -z, -w)$ . For element  $A$ , with coordinates  $(x, y, z, w)$ , we have the 4-form

$$dx \wedge (dy \wedge dz \wedge dw). \quad (16.13.9)$$

Since element  $A$  is an element of the three-dimensional  $SU(2)_A$  group, its orientation can be determined by the 3-form  $dy \wedge dz \wedge dw$  in equation (16.13.9), while the 1-form  $dx$  in (16.13.9) can be regarded as determining the direction of the **normal vector** at element  $A$ . Similarly, for element  $B^{-1}$ , with coordinates  $(-x, -y, -z, -w)$ , we have the 4-form

$$-dx \wedge (-dy) \wedge (-dz) \wedge (-dw) = dx \wedge (dy \wedge dz \wedge dw). \quad (16.13.10)$$

Since element  $B^{-1}$  is an element of the three-dimensional  $SU(2)_B$  group, its orientation can be determined by the 3-form  $dy \wedge dz \wedge dw$  in equation (16.13.10), while the 1-form  $dx$  on the right-hand side of (16.13.10) can be regarded as determining the direction of the **normal vector** at element  $B^{-1}$ . Therefore, the two elements  $A$  and  $B^{-1}$  have the **same orientation** in the internal direct product group  $SU(2)_A \otimes SU(2)_B$ —both determined by  $dy \wedge dz \wedge dw$ —and their normal-vector directions are also the same. Consequently, the internal direct product group  $SU(2)_A \otimes SU(2)_B$  formed by pairing these two elements  $A$  and  $B^{-1}$  is an **orientable smooth manifold**.

### (3) The Second Way to Form $RP^3$

We now pair element  $A^{-1}$  of particle  $A$  with element  $B$  of particle  $B$  to form an element of the internal direct product group  $SU(2)_A \otimes SU(2)_B$ :

$$(A^{-1}, B) = \left( \begin{pmatrix} x - iy & -z - iw \\ z - iw & x + iy \end{pmatrix}, \begin{pmatrix} -x + iy & z + iw \\ -z + iw & -x - iy \end{pmatrix} \right), \quad (16.13.11)$$

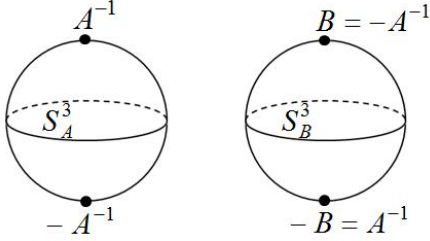
then



$$(-A^{-1}, -B) = \left( \begin{pmatrix} -x+iy & z+iw \\ -z+iw & -x-iy \end{pmatrix}, \begin{pmatrix} x-iy & -z-iw \\ z-iw & x+iy \end{pmatrix} \right). \quad (16.13.12)$$

From equations (16.13.11) and (16.13.12) we obtain

$$(A^{-1}, B) = (-A^{-1}, -B). \quad (16.13.13)$$



**Figure 16.13.2** The second way to form  $RP^3$

As shown in Figure 16.13.2, point  $A^{-1}$  and point  $-A^{-1}$  on  $S^3_A$  are antipodal points, while point  $B$  and point  $-B$  on  $S^3_B$  are also antipodal points. Point  $A^{-1}$  on  $S^3_A$  and point  $B$  on  $S^3_B$  together form an element  $(A^{-1}, B)$  of the inner direct product group  $SU(2)_A \otimes SU(2)_B$ , and point  $-A^{-1}$  on  $S^3_A$

together with point  $-B$  on  $S^3_B$  also form an element  $(-A^{-1}, -B)$  of the inner direct product group  $SU(2)_A \otimes SU(2)_B$ . Points  $(A^{-1}, B)$  and  $(-A^{-1}, -B)$  are antipodal points, but according to equation (16.13.13), these two points are identified, meaning these two antipodal points are identified as the same point.

Although the elements  $A^{-1}$  and  $B$  originally belong to different Lie groups with opposite orientations before pairing, once they are combined to form an element of the inner direct product group  $SU(2)_A \otimes SU(2)_B$ , they become elements of the inner direct product group  $SU(2)_A \otimes SU(2)_B$ . Consequently, the orientations of these two elements  $A^{-1}$  and  $B$  within the inner direct product group  $SU(2)_A \otimes SU(2)_B$  should be the same. According to equation (16.13.11), it can be considered that within the inner direct product group, the coordinate of element  $A^{-1}$  is  $(x, -y, -z, -w)$ , and the coordinate of element  $B$  is  $(-x, y, z, w)$ . For element  $A^{-1}$ , since its coordinate is  $(x, -y, -z, -w)$ , there exists a 4-form differential expression

$$dx \wedge (-dy) \wedge (-dz) \wedge (-dw) = -dx \wedge (dy \wedge dz \wedge dw). \quad (16.13.14)$$

Since element  $A^{-1}$  is an element of the three-dimensional group  $SU(2)_A$ , its orientation can be determined by the 3-form differential expression  $dy \wedge dz \wedge dw$  in equation (16.13.14). The  $-dx$  on the right-hand side of equation (16.13.14) can be regarded as a 1-form differential expression that determines the direction of the normal vector at element  $A^{-1}$ . Similarly, for element  $B$ , since its coordinate is  $(-x, y, z, w)$ , there exists a 4-form differential expression

$$-dx \wedge (dy \wedge dz \wedge dw). \quad (16.13.15)$$

Since element  $B$  belongs to the three-dimensional group  $SU(2)_B$ , its orientation can be determined by the 3-form differential expression  $dy \wedge dz \wedge dw$  in equation (16.13.15). The  $-dx$  in equation (16.13.15) can be regarded as a 1-form differential expression that determines the direction of the normal vector at element  $B$ . Therefore, the orientations of these two elements  $A^{-1}$  and  $B$  within the inner direct product group  $SU(2)_A \otimes SU(2)_B$  are the same—both are determined by  $dy \wedge dz \wedge dw$  to define their orientations, and their normal vector directions are also identical. Consequently, the inner direct product group  $SU(2)_A \otimes SU(2)_B$  formed by pairing these two elements  $A^{-1}$  and  $B$  is an orientable smooth manifold.

The inner direct product group  $SU(2)_A \otimes SU(2)_B$  is isomorphic to the group  $SU(2)_A$  and the group  $SU(2)_B$ , i.e.,

$$SU(2)_A \otimes SU(2)_B \cong SU(2)_A \cong SU(2)_B,$$

and thus it is also homeomorphic to the three-dimensional unit sphere. However, based on the analysis above, the three-dimensional sphere  $S^3_{\otimes}$  that is homeomorphic to the inner direct



product group  $SU(2)_A \otimes SU(2)_B$ —formed by two oppositely oriented groups  $SU(2)_A$  and  $SU(2)_B$  according to equation (16.13.6) or (16.13.11)—has all its antipodal points identified. Therefore,  $S^3_\otimes$  can be regarded as a projective space  $RP^3$ . Furthermore, according to the preceding analysis, the formation of this projective space  $RP^3$  is definite and possesses inevitability.

## 2. Projective Space $RP^3$ Formed by Particles with the Same Orientation

When two spin-1/2 elementary particles with the same orientation combine to form a projective space  $RP^3$ , the process is relatively straightforward.

Let particles  $A$  and  $B$  both be spin-1/2 elementary particles with the same orientation. In  $R^4$ , the coordinates describing particles  $A$  and  $B$  can be identical, both assumed to be  $(x, y, z, w)$ . Suppose the Lie groups describing particles  $A$  and  $B$  are the  $SU(2)_A$  group and the  $SU(2)_B$  group, respectively. According to equation (16.13.1), the group elements of the  $SU(2)_A$  group and the  $SU(2)_B$  group are

$$A = \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}, \quad B = \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}, \quad (16.13.16)$$

respectively, where

$$x^2 + y^2 + z^2 + w^2 = 1.$$

The corresponding  $-A$  and  $-B$  are

$$-A = \begin{pmatrix} -x - iy & -z - iw \\ z - iw & -x + iy \end{pmatrix}, \quad -B = \begin{pmatrix} -x - iy & -z - iw \\ z - iw & -x + iy \end{pmatrix}. \quad (16.13.17)$$

respectively.

The two Lie groups of particles  $A$  and  $B$  also form an inner direct product group  $SU(2)_A \otimes SU(2)_B$ . Suppose the pairing of elements is given by

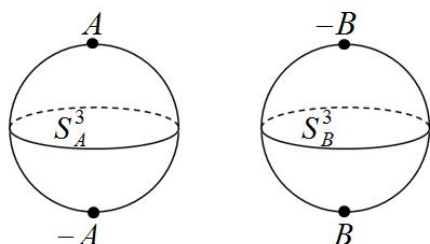
$$(A, -B) = \left( \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix}, \begin{pmatrix} -x - iy & -z - iw \\ z - iw & -x + iy \end{pmatrix} \right), \quad (16.13.18)$$

then

$$(-A, B) = \left( \begin{pmatrix} -x - iy & -z - iw \\ z - iw & -x + iy \end{pmatrix}, \begin{pmatrix} x + iy & z + iw \\ -z + iw & x - iy \end{pmatrix} \right). \quad (16.13.19)$$

From equations (16.13.18) and (16.13.19), we obtain:

$$(A, -B) = (-A, B). \quad (16.13.20)$$



**Figure 16.13.3**  $RP^3$  composed of two particles with the same orientation

As shown in Figure 16.13.3, the  $SU(2)_A$  group is homeomorphic to the unit sphere  $S_A^3$ , and the  $SU(2)_B$  group is also homeomorphic to the unit sphere  $S_B^3$ . Point  $A$  and point  $-A$  on  $S_A^3$  are antipodal points, while point  $B$  and point  $-B$  on  $S_B^3$  are also antipodal points. Point  $A$  on  $S_A^3$  and point  $-B$  on  $S_B^3$  together form an

element  $(A, -B)$  of the inner direct product group  $SU(2)_A \otimes SU(2)_B$ , while point  $-A$  on  $S_A^3$  and point  $B$  on  $S_B^3$  also form an element  $(-A, B)$  of the inner direct product group  $SU(2)_A \otimes SU(2)_B$ . Points  $(A, -B)$  and  $(-A, B)$  are antipodal points, but according to equation (16.13.20), these two points are identified, meaning these two antipodal points are identified as the same point.

Therefore, based on the analysis above, although the inner direct product group  $SU(2)_A \otimes SU(2)_B$  formed by spin-1/2 elementary particles with the same orientation is homeomorphic to the three-dimensional sphere  $S^3_\otimes$ , as long as their points (or elements) are paired in the manner described above, the antipodal points of  $S^3_\otimes$  become identified. Thus,  $S^3_\otimes$  can also be regarded as a projective space  $RP^3$ . However, according to the previous analysis, in the absence of external factors, the probability of forming such a projective space  $RP^3$  in this way is very low and highly coincidental.

The above conclusion has a good verification: two electrons with opposite spin  $z$ -components and opposite momenta meet the above requirements and can combine to form a particle pair exhibiting  $RP^3$ -like properties (i.e., bosonic characteristics). This particle pair is the well-known Cooper pair.

## Chapter 17 Particles as Principal Bundles

We have assumed that electrons, positrons, protons, antiprotons, neutrinos, and antineutrinos are all  $SU(2)$  groups, implying that we assume the properties of these particles are identical. However, in reality, these particles differ significantly. That is to say, these particles have both common properties and distinct properties. Similarly, we assume that photons and antiphotons are all  $SO(3)$  groups, also implying that we assume all photons and antiphotons share identical properties, but in fact, they also have distinct properties. We also assume that elementary dark particles  $\chi_0, \chi_1, \chi_2$ , and their antiparticles are all  $L_+^\uparrow$  groups, implying that we assume the properties of these particles are identical, but in fact, they also have distinct properties. Therefore, it is necessary to further subdivide these particles. Experiments have measured that the rest mass of protons and antiprotons is greater than that of electrons and positrons, while the rest mass of photons and neutrinos/antineutrinos equals zero. Consequently, this chapter further classifies these particles using principal bundle theory based on their rest mass. The further classification of dark particles  $\chi_0, \chi_1, \chi_2$ , and their antiparticles will be addressed in the final chapter.

### §17.1 The Relationship Between Inertial Force and Rest Mass

The mass  $m$  referred to in this chapter denotes the rest mass of a particle. Classical mechanics asserts that rest mass  $m$  is a measure of a particle's inertia: the greater the  $m$ , the harder it is to alter its state of motion. A particle possesses rest mass because it experiences an inertial force—an inherent force intrinsic to the particle, acting as a resistive force that tends to keep all particles either at rest or in a state of uniform rectilinear motion. When an external force is applied to a particle in an attempt to change its state of motion, the inertial force endeavors to maintain the particle's original state. This inevitably raises the question: Why do particles possess inertia or inertial force? Or, what is the mechanism that generates inertial force? This chapter aims to explore this issue. This chapter's discussion of rest mass is speculative in nature; a precise definition of rest mass will be provided in Chapter 20.

#### 1. Expressing Inertial Force Using Divergence and Curl

We have already considered particles as smooth manifolds, on which smooth tangent vector fields exist. A smooth tangent vector field is analogous to a velocity field. Taking the divergence and curl of a smooth tangent vector field is similar to computing the divergence and curl of a velocity field, with the results resembling an acceleration field. According to Newton's second law, acceleration is proportional to force, so we can relate the divergence and curl of a smooth tangent vector field to force. Since the smooth tangent vector field resides within the particle, this force is intrinsic to the particle. We can associate this force with inertial force and refer to it as such.

For example, consider a velocity field  $\mathbf{V} = V_x \mathbf{i} + V_y \mathbf{j} + V_z \mathbf{k}$  in three-dimensional Euclidean space  $R^3$ . Its divergence is

$$\operatorname{div} \mathbf{V} = \nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}.$$

$\frac{\partial V_x}{\partial x}$  represents the rate of change of the  $x$ -component  $V_x$  of the velocity field  $\mathbf{V}$  along the  $x$ -direction. It can be assumed to be proportional to the  $x$ -component  $F_x$  of a certain force  $\mathbf{F}$ ,

$$\frac{\partial V_x}{\partial x} \propto F_x.$$

Similarly,  $\frac{\partial V_y}{\partial y}$  represents the rate of change of the  $y$ -component  $V_y$  of the velocity field  $\mathbf{V}$  along the  $y$ -direction and can be assumed proportional to the  $y$ -component  $F_y$  of the force  $\mathbf{F}$ ,

$$\frac{\partial V_y}{\partial y} \propto F_y .$$

Also,  $\frac{\partial V_z}{\partial z}$  represents the rate of change of the  $z$ -component  $V_z$  of the velocity field  $\mathbf{V}$  along the  $z$ -direction. It can be assumed to be proportional to the  $z$ -component  $F_z$  of a certain force  $\mathbf{F}$ ,

$$\frac{\partial V_z}{\partial z} \propto F_z .$$

The divergence  $\text{div}\mathbf{V}$  is proportional to the sum of  $F_x, F_y$ , and  $F_z$  :

$$\text{div}\mathbf{V} \propto F_x + F_y + F_z .$$

According to Newton's second law, the force  $\mathbf{F}$  is proportional to the rest mass  $m$ ,  $\mathbf{F} \propto m$ . Thus, we obtain

$$\text{div}\mathbf{V} \propto m .$$

Therefore, it can be concluded that when  $\text{div}\mathbf{V} = 0, m = 0$ ; and when  $\text{div}\mathbf{V} \neq 0, m \neq 0$ .

In addition to calculating the rate of change of the  $x$ -component  $V_x$  of the velocity field  $\mathbf{V}$  along the  $x$ -direction, we can also compute its rate of change along other directions. Specifically, we can compute the curl of the velocity field  $\mathbf{V}$ :

$$\text{curl}\mathbf{V} = \left( \frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right) \mathbf{i} + \left( \frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \right) \mathbf{j} + \left( \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) \mathbf{k} .$$

The  $x$ -component of the curl,  $\left( \frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right)$ , can be assumed proportional to the  $x$ -component  $G_x$  of a certain force  $\mathbf{G}$ :

$$\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \propto G_x .$$

Similarly, it can be assumed that  $\left( \frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \right)$  is proportional to the  $y$ -component  $G_y$  of the force  $\mathbf{G}$ :

$$\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \propto G_y ,$$

and  $\left( \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right)$  is proportional to the  $z$ -component  $G_z$  of the force  $\mathbf{G}$ :

$$\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \propto G_z .$$

That is, we assume  $\text{curl}\mathbf{V}$  is proportional to the force  $\mathbf{G}$ :

$$\text{curl}\mathbf{V} \propto \mathbf{G} .$$

According to Newton's second law, the force  $\mathbf{G}$  is proportional to the mass  $m$ :  $\mathbf{G} \propto m$ . Thus, we obtain

$$\text{curl}\mathbf{V} \propto m .$$

Therefore, it can be concluded that when  $\text{curl}\mathbf{V} = 0, m = 0$ ; and when  $\text{curl}\mathbf{V} \neq 0, m \neq 0$ .

This example suggests that when both the divergence and curl of a smooth tangent vector field are zero, the particle can be considered to have no inertial force and thus no rest mass. Conversely, if either the divergence or the curl of the smooth tangent vector field is non-zero, the particle can be considered to possess inertial force and, consequently, rest mass.

## 2.Relationship Between the Inertial Force of a Free Particle and Its Rest Mass

Quantum field theory asserts that the Klein-Gordon equation (16.9.1) describes spin-0 particles, while the Dirac equation describes spin-1/2 particles. Since the Dirac equation can be derived by factorizing the Klein-Gordon equation, conversely, the Klein-Gordon equation can also be obtained from the Dirac equation. Therefore, a solution  $\psi$  that satisfies the Dirac equation

simultaneously satisfies the Klein-Gordon equation (in this case, all four components of  $\psi$  satisfy the Klein-Gordon equation).

The Klein-Gordon equation for a free particle is

$$\left( \frac{\hbar^2}{c^2} \nabla^2 - \frac{\hbar^2}{c^4} \frac{\partial^2}{\partial t^2} \right) \psi = m^2 \psi. \quad (17.1.1)$$

Equation (17.1.1) can be rewritten as

$$\left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial t} \right) \cdot \left( \frac{\hbar^2}{c^2} \frac{\partial}{\partial x}, \frac{\hbar^2}{c^2} \frac{\partial}{\partial y}, \frac{\hbar^2}{c^2} \frac{\partial}{\partial z}, -\frac{\hbar^2}{c^4} \frac{\partial}{\partial t} \right) \psi = m^2 \psi. \quad (17.1.2)$$

Let

$$\diamond = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial t} \right), \quad \mathbf{X} = \left( \frac{\hbar^2}{c^2} \frac{\partial}{\partial x}, \frac{\hbar^2}{c^2} \frac{\partial}{\partial y}, \frac{\hbar^2}{c^2} \frac{\partial}{\partial z}, -\frac{\hbar^2}{c^4} \frac{\partial}{\partial t} \right) \psi,$$

then equation (17.1.2) can be written as

$$\text{div} \mathbf{X} = \diamond \cdot \mathbf{X} = m^2 \psi. \quad (17.1.3)$$

We often say that when in equation (17.1.1)  $m = 0$ ,

$$\left( \frac{\hbar^2}{c^2} \nabla^2 - \frac{\hbar^2}{c^4} \frac{\partial^2}{\partial t^2} \right) \psi = 0,$$

but we can also say that when

$$\left( \frac{\hbar^2}{c^2} \nabla^2 - \frac{\hbar^2}{c^4} \frac{\partial^2}{\partial t^2} \right) \psi = 0,$$

$m = 0$ . Similarly, we can say that when the divergence of  $\mathbf{X}$  is  $\text{div} \mathbf{X} = \diamond \cdot \mathbf{X} = 0$ , the rest mass of the particle is  $m = 0$ ; when  $\text{div} \mathbf{X} = \diamond \cdot \mathbf{X} \neq 0$ , the rest mass of the particle is  $m \neq 0$ .

In equation (17.1.1), setting  $m = 0$ ,  $\psi = \mathbf{E}$  or  $\psi = \mathbf{B}$ , yields the d'Alembert equation satisfied by the electric field intensity  $\mathbf{E}$  and magnetic field intensity  $\mathbf{B}$  of a photon:

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \quad \nabla^2 \mathbf{B} - \frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} = 0.$$

Let

$$\diamond = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial t} \right), \quad \mathbf{X} = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, -\frac{1}{c^2} \frac{\partial}{\partial t} \right) \mathbf{E}, \quad \tilde{\mathbf{X}} = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, -\frac{1}{c^2} \frac{\partial}{\partial t} \right) \mathbf{B},$$

then the d'Alembert equation can be written in the following form:

$$\text{div} \mathbf{E} = \diamond \cdot \mathbf{X} = 0, \quad \text{div} \mathbf{B} = \diamond \cdot \tilde{\mathbf{X}} = 0.$$

Therefore, we assert that precisely because  $\text{div} \mathbf{E} = \diamond \cdot \mathbf{X} = 0$  and  $\text{div} \mathbf{B} = \diamond \cdot \tilde{\mathbf{X}} = 0$ , the rest mass of the photon equals zero.

Next, let us examine the Dirac equation. Writing the Dirac equation (16.9.5) as

$$\left( i\hbar c \boldsymbol{\beta} \boldsymbol{\alpha} \cdot \nabla + i\hbar \beta \frac{\partial}{\partial t} \right) \psi = mc^2 \psi. \quad (17.1.4)$$

the operator  $\hat{H} = i\hbar c \boldsymbol{\beta} \boldsymbol{\alpha} \cdot \nabla + i\hbar \beta \frac{\partial}{\partial t}$  possesses properties akin to both divergence and curl. Here,  $\psi$  consists of four components and is a four-row, one-column matrix. When  $\hat{H}\psi = 0$ , the rest mass of the particle is zero,  $m = 0$ . When  $\hat{H}\psi \neq 0$ , the rest mass of the particle is non-zero,  $m \neq 0$ .

## §17.2 Transformations with Only a One-Parameter Transformation Group

Since both the  $SO(3)$  group and the  $SU(2)$  group are compact smooth manifolds, and their spin operators are smooth tangent vector fields, according to Theorem 6.6.1, these spin operators each define a one-parameter differentiable transformation group on the  $SO(3)$  and  $SU(2)$  groups, respectively.

The spin operators of the  $SU(2)$  group are  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$ . They each define a one-parameter differentiable transformation group on the  $SU(2)$  group. Let the one-parameter differentiable transformation groups determined by  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  be denoted as  $G_1$ ,  $G_2$ , and  $G_3$ , respectively. Then

$$\begin{aligned}\kappa_1 : G_1 \times SU(2) &\rightarrow SU(2), \\ \kappa_2 : G_2 \times SU(2) &\rightarrow SU(2), \\ \kappa_3 : G_3 \times SU(2) &\rightarrow SU(2).\end{aligned}$$

$G_1$  is a differentiable homeomorphism transformation group acting on the  $SU(2)$  group. Precisely because  $G_1$  is a group, every element in  $G_1$  has an inverse, enabling the  $SU(2)$  group to potentially return to its original state after being transformed by each element of  $G_1$ . The same applies to  $G_2$  and  $G_3$ .

Here, the transformations of the  $SU(2)$  group we analyze are those that do not map the  $SU(2)$  group into other groups—i.e., transformations that do not involve changing particles into other particles.  $G_1$  is such a transformation, as are  $G_2$  and  $G_3$ . We might then ask: If the  $SU(2)$  group is simultaneously transformed by the  $G_1$  group and the  $G_2$  group, does it remain an  $SU(2)$  group? The following answers this question.

The  $SU(2)$  group, after being simultaneously transformed by the  $G_1$  group and the  $G_2$  group, remains an  $SU(2)$  group. This process can be represented as

$$\kappa_{12} : (G_1 \times G_2) \times SU(2) \rightarrow SU(2), \quad (17.2.1)$$

where the set  $G_1 \times G_2$  must form a group.

Since both  $G_1$  and  $G_2$  are subgroups of the same  $SU(2)$  group, the multiplication of elements between  $G_1$  and  $G_2$  is permissible, and the multiplication of elements between  $G_1$  and  $G_2$  is the same as the multiplication of elements in the  $SU(2)$  group. Let  $x_1, y_1 \in G_1$  and  $x_2, y_2 \in G_2$ . According to our problem, the multiplication of elements in the set  $G_1 \times G_2$  should be defined as

$$G_1 \times G_2 = \{x_1 x_2 : x_1 \in G_1, x_2 \in G_2\}. \quad (17.2.2)$$

For the set  $G_1 \times G_2$  to form a group, it must satisfy the four group axioms. If any one of these axioms is not satisfied, then  $G_1 \times G_2$  is not a group.

Since  $x_1, y_1 \in G_1$  and  $x_2, y_2 \in G_2$ , we have

$$x = x_1 x_2 \in G_1 \times G_2, \quad y = y_1 y_2 \in G_1 \times G_2.$$

For  $G_1 \times G_2$  to be a group, the group multiplication must be closed, i.e., we must have

$$xy = x_1 x_2 y_1 y_2 \in G_1 \times G_2.$$

But according to equation (17.2.2), we must have  $x_1 x_2 \in G_1$  and  $y_1 y_2 \in G_2$ , which is obviously not necessarily true. Because  $x_1$  and  $x_2$  do not belong to the same group, and  $y_1$  and  $y_2$  also do not belong to the same group, therefore  $xy$  is not necessarily an element of the set  $G_1 \times G_2$ , i.e., the group multiplication is not necessarily closed.

If  $x_2 y_1 = y_1 x_2$ , then

$$xy = x_1 x_2 y_1 y_2 = x_1 y_1 x_2 y_2 \in G_1 \times G_2.$$

In this case, the group multiplication is closed. Requiring  $x_2 y_1 = y_1 x_2$  to hold is equivalent to requiring that the multiplication of elements in the  $SU(2)$  group is commutative, but this is impossible because the  $SU(2)$  group is not an Abelian group.

If  $G_1 \times G_2$  is to be a group, then every element must have an inverse. Since  $x_1 x_2 \in G_1 \times G_2$ , if  $x_1 x_2$  has an inverse, its inverse would be  $(x_1 x_2)^{-1}$ . But because  $(x_1 x_2)^{-1} = x_2^{-1} x_1^{-1}$ , according to the definition of  $G_1 \times G_2$ , we would necessarily have  $x_2^{-1} \in G_1$  and  $x_1^{-1} \in G_2$ . This contradicts our assumption  $x_2 \in G_2$ ,  $x_1 \in G_1$ . Therefore,  $x_1 x_2$  does not have an inverse. However, if

$x_2^{-1}x_1^{-1} = x_1^{-1}x_2^{-2}$ , then according to equation (17.2.2), we would necessarily have  $x_1^{-1} \in G_1, x_2^{-1} \in G_2$ , and thus  $x_1x_2$  would have an inverse  $(x_1x_2)^{-1}$ . But requiring  $x_2^{-1}x_1^{-1} = x_1^{-1}x_2^{-2}$  to hold is equivalent to requiring that the multiplication of elements in the  $SU(2)$  group is commutative, which is impossible. Hence, in the set  $G_1 \times G_2$ , not every element has an inverse. It is evident that the set  $G_1 \times G_2$  cannot form a group.

Since it cannot be guaranteed that every element in  $G_1 \times G_2$  has an inverse, if the set  $G_1 \times G_2$  acts on the  $SU(2)$  group, it cannot be guaranteed that this transformation is invertible.

Because  $G_1 \times G_2$  cannot form a group,  $(G_1 \times G_2) \times SU(2)$  is not an  $SU(2)$  group. Let  $g$  be any element of the  $SU(2)$  group, i.e.,  $g \in SU(2)$ , and let  $z \in G_1 \times G_2$ . If  $(G_1 \times G_2) \times SU(2)$  were an  $SU(2)$  group, then  $zg$  would be an element of the  $SU(2)$  group, and it would necessarily have an inverse  $(zg)^{-1}$  such that  $(zg)^{-1}zg = e$ , where  $e$  is the identity element of the  $SU(2)$  group. From  $(zg)^{-1}zg = e$ , we obtain  $(zg)^{-1}zg = g^{-1}z^{-1}zg = e$ . This requires that  $z^{-1}$  is the inverse of  $z$ , but we have already seen that  $G_1 \times G_2$  is not a group, so  $z$  may not have an inverse  $z^{-1}$ . Consequently,  $(G_1 \times G_2) \times SU(2)$  cannot form a group, and certainly is not an  $SU(2)$  group.

Since the set  $G_1 \times G_2$  cannot form a group, it cannot be a Lie transformation group. Therefore, the mapping

$$\kappa_{12} : (G_1 \times G_2) \times SU(2) \rightarrow SU(2) \quad (17.2.3)$$

is not valid. The parentheses in (17.2.3) indicate that  $G_1$  and  $G_2$  act simultaneously as a whole on the  $SU(2)$  group. If they do not act simultaneously as a whole on the  $SU(2)$  group, but instead act sequentially on the  $SU(2)$  group, then the following two mappings hold:

$$\begin{aligned} \kappa_1 : G_1 \times SU(2) &\rightarrow SU(2), \\ \kappa_2 : G_2 \times (G_1 \times SU(2)) &\rightarrow G_2 \times SU(2) \rightarrow SU(2). \end{aligned}$$

Similarly, if the  $SU(2)$  group, after being simultaneously transformed by the  $G_1, G_2$ , and  $G_3$  groups, remains an  $SU(2)$  group, then this process can be represented as

$$\kappa_{123} : (G_1 \times G_2 \times G_3) \times SU(2) \rightarrow SU(2), \quad (17.2.4)$$

where the set  $G_1 \times G_2 \times G_3$  must form a group. Let  $x_3, y_3 \in G_3$ . According to this requirement, the multiplication of elements in the set  $G_1 \times G_2 \times G_3$  should be defined as

$$G_1 \times G_2 \times G_3 = \{x_1x_2x_3 : x_1 \in G_1, x_2 \in G_2, x_3 \in G_3\}. \quad (17.2.5)$$

For the set  $G_1 \times G_2 \times G_3$  to be a group, the group multiplication must be closed. Let

$$a = x_1x_2x_3 \in G_1 \wedge G_2 \wedge G_3, \quad b = y_1y_2y_3 \in G_1 \wedge G_2 \wedge G_3,$$

then the closure property requires that the following holds:

$$ab = x_1x_2x_3y_1y_2y_3 \in G_1 \wedge G_2 \wedge G_3.$$

But according to equation (17.2.5), we must have  $x_1x_2 \in G_1$ ,  $x_3y_1 \in G_2$ ,  $y_2y_3 \in G_2$ , which is obviously not true. Because  $x_1$  and  $x_2$  do not belong to the same group  $G_1$ ,  $x_3$  and  $y_1$  do not belong to the same group  $G_2$ , and  $y_2$  and  $y_3$  do not belong to the same group  $G_3$ . Therefore,  $ab$  is not necessarily an element of the set  $G_1 \times G_2 \times G_3$ , i.e., the group multiplication is not necessarily closed. If the multiplication of elements in the  $SU(2)$  group were commutative, then the following would hold:

$$ab = x_1x_2x_3y_1y_2y_3 = x_1y_1x_2y_2x_3y_3 \in G_1 \times G_2 \times G_3,$$

and the group multiplication would be closed. However, since the multiplication of elements in the  $SU(2)$  group is not commutative, the set  $G_1 \times G_2 \times G_3$  also cannot form a group.

Therefore, because the set  $G_1 \times G_2 \times G_3$  cannot form a group, the mapping

$$\kappa_{123} : (G_1 \times G_2 \times G_3) \times SU(2) \rightarrow SU(2)$$

is also not valid.

However, we know that the mapping  $\kappa_1 : G_1 \times SU(2) \rightarrow SU(2)$  is valid. Therefore, the group of differentiable homeomorphic transformations on the  $SU(2)$  group can only be one of the



three one-parameter differentiable transformation groups determined by  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$ . The product of any two or all three of these one-parameter differentiable transformation groups cannot serve as a differentiable transformation group for the  $SU(2)$  group.

The transformations of the  $SU(2)$  group analyzed here are those that do not alter the  $SU(2)$  group—meaning they do not involve transforming particles into other particles. The requirement that all such transformations on the  $SU(2)$  group form a group arises because every element in a group has an inverse, enabling the  $SU(2)$  group to potentially return to its original state after undergoing various transformations. Additionally, demanding that this group be a continuous group is because a continuous group possesses infinitely many elements, corresponding to infinitely many transformations, thereby encompassing all possible transformations on the  $SU(2)$  group. The one-parameter differentiable transformation group generated by a single spin operator is precisely such a continuous group, meeting our requirements.

Experimental measurements have revealed that among the three spin operators, only one of them yields definite measurement values, while the other two do not. This indicates that among these three operators, only one plays a decisive role.

Using a similar method as above, it can also be concluded that although the  $SO(3)$  group has three one-parameter differentiable transformation groups determined by  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$ , only one of these one-parameter differentiable transformation groups can act on the  $SO(3)$  group while keeping the  $SO(3)$  group as  $SO(3)$ .

Each photon is an  $SO(3)$  group. Let the three one-parameter differentiable transformation groups determined by  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$  be denoted as  $L_x$ ,  $L_y$ , and  $L_z$ , respectively. Then, we have the mapping:

$$\begin{aligned}\tau_x &: L_x \times SO(3) \rightarrow SO(3), \\ \tau_y &: L_y \times SO(3) \rightarrow SO(3), \\ \tau_z &: L_z \times SO(3) \rightarrow SO(3).\end{aligned}$$

Since the expressions for  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$  are similar (or symmetric), the properties of photons under the action of the one-parameter differentiable transformation groups  $L_x$ ,  $L_y$ , and  $L_z$  are also similar (or symmetric).

Similarly, each spin-1/2 elementary particle is an  $SU(2)$  group. Let the three one-parameter differentiable transformation groups determined by  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  be denoted as  $S_x$ ,  $S_y$ , and  $S_z$ , respectively. Then, we have the mapping:

$$\begin{aligned}\kappa_x &: S_x \times SU(2) \rightarrow SU(2), \\ \kappa_y &: S_y \times SU(2) \rightarrow SU(2), \\ \kappa_z &: S_z \times SU(2) \rightarrow SU(2).\end{aligned}$$

Since the expressions for  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  differ significantly, the properties of particles under the action of the one-parameter differentiable transformation groups  $S_x$ ,  $S_y$ , and  $S_z$  also vary considerably. Below, we will specify:  $S_x \times SU(2)$  corresponds to protons and antiprotons;  $S_y \times SU(2)$  corresponds to neutrinos and antineutrinos; and  $S_z \times SU(2)$  corresponds to electrons and antielectrons.

Although protons and antiprotons are represented in the form of a product manifold (or direct product of Lie groups)  $S_x \times SU(2)$ , protons and antiprotons are still an  $SU(2)$  group—but rather an  $SU(2)$  group that has been transformed by  $S_x$ . The same understanding applies to other particles.



### §17.3 Physical Significance of Complex Function Integration

This section is adapted from the article by Song Xiangnuan<sup>2</sup>. Consider a complex function  $f(z) = u(x, y) + iv(x, y)$  defined in a region  $D$  of the complex plane, and let vector  $A = f(z)^* = u(x, y) - iv(x, y)$ , with  $C$  being any smooth oriented curve within region  $D$ . Define

$$dz = dx + idy = ds\mathbf{l}^0, \quad -idz = dy - idx = ds\mathbf{n}^0,$$

Where  $ds$  is the arc differential,  $\mathbf{l}^0$  is the unit vector along the direction of curve  $C$  at point  $z$ , and  $\mathbf{n}^0$  is the unit normal vector at that point, as illustrated in Figure 17.3.1.

In the Euclidean plane Cartesian coordinate system, vector  $A$  is expressed as

$$A = u(x, y)e_x - v(x, y)e_y,$$

$ds\mathbf{l}^0$  is expressed as

$$ds\mathbf{l}^0 = dx\mathbf{e}_x + dy\mathbf{e}_y,$$

and  $ds\mathbf{n}^0$  is expressed as

$$ds\mathbf{n}^0 = dy\mathbf{e}_x - dx\mathbf{e}_y,$$

where  $\mathbf{e}_x, \mathbf{e}_y$  are the unit vectors along the  $x$ -axis and

$y$ -axis in the Euclidean plane, respectively. Computing the inner product of  $ds\mathbf{l}^0$  and  $ds\mathbf{n}^0$ , we obtain

$$ds\mathbf{l}^0 \cdot ds\mathbf{n}^0 = (dx\mathbf{e}_x + dy\mathbf{e}_y) \cdot (dy\mathbf{e}_x - dx\mathbf{e}_y) = dx dy - dy dx = 0,$$

so  $ds\mathbf{l}^0$  and  $ds\mathbf{n}^0$  are mutually perpendicular. Translating this to the complex plane,  $ds\mathbf{l}^0$  and  $ds\mathbf{n}^0$  are naturally mutually perpendicular. In the Euclidean plane:

$$A \cdot \mathbf{l}^0 ds = [u(x, y)\mathbf{e}_x - v(x, y)\mathbf{e}_y] \cdot (dx\mathbf{e}_x + dy\mathbf{e}_y) = u(x, y)dx - v(x, y)dy,$$

$$A \cdot \mathbf{n}^0 ds = [u(x, y)\mathbf{e}_x - v(x, y)\mathbf{e}_y] \cdot (dy\mathbf{e}_x - dx\mathbf{e}_y) = v(x, y)dx + u(x, y)dy.$$

The integral of the complex function  $f(z)$  along curve  $C$  in the complex plane is

$$\begin{aligned} \int_C f(z)dz &= \int_C [u(x, y) + iv(x, y)]d(x + iy) \\ &= \int_C u(x, y)dx - v(x, y)dy + i \int_C v(x, y)dx + u(x, y)dy \\ &= \int_C A \cdot \mathbf{l}^0 ds + i \int_C A \cdot \mathbf{n}^0 ds. \end{aligned}$$

The above formula endows the integral of a complex function with physical significance. From the perspective of field theory, the real part  $\int_C A \cdot \mathbf{l}^0 ds$  of  $\int_C f(z)dz$  represents the circulation of the

vector field  $A$  along curve  $C$ , while the imaginary part  $\int_C A \cdot \mathbf{n}^0 ds$  represents the flux of the vector

field  $A$  through curve  $C$ . If the integral value is a real number, it indicates that the circulation of the vector field  $A$  along curve  $C$  is nonzero, while the flux of the vector field  $A$  through curve  $C$  is zero. Conversely, if the integral value is a purely imaginary number, it indicates that the circulation of the vector field  $A$  along curve  $C$  is zero, while the flux of the vector field  $A$  through curve  $C$  is nonzero.

Although vector  $A$  is not the vector represented by  $f(z)$ , its magnitude is equal to that of

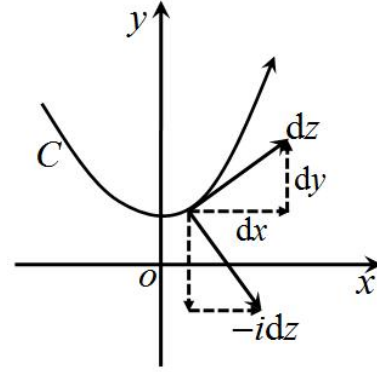


Figure 17.3.1 Relationship between  $dz$  and  $-idz$

<sup>2</sup>Song Xiangnuan. The Physical Significance of the Integral of Complex Functions [J]. Tianjin: Journal of Tianjin University of Commerce, 1996, (3): 41–42.

the vector represented by  $f(z)$ , and it is symmetric about the  $x$ -axis. Therefore, vector  $A$  is the mirror image of the vector represented by  $f(z)$ . The circulation and flux of vector  $A$  can thus represent the cumulative variation of the vector represented by  $f(z)$  along the integration curve  $C$ .

Since four of the six spin operators of the  $SO(3)$  and  $SU(2)$  groups involve complex numbers, we will now utilize the physical significance of complex function integration. Specifically, we will compute the integrals of the spin operators of these two groups along the unit circle  $C$  in the complex plane. This will allow us to obtain the circulation and flux of the spin operators along the unit circle  $C$ , thereby gaining insight into and comparing the characteristics of these spin operators and exploring what discoveries can be made from this analysis.

## §17.4 Photon as a Principal Bundle

### 1. Circulation and Flux of the Spin Operators of the $SO(3)$ Group

We begin by calculating with the  $\hat{L}_z$  operator of the  $SO(3)$  group,

$$\hat{L}_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right).$$

We assume  $\hat{L}_z$  is an operator in the complex plane  $xy$ . Since the integration curve  $C$  is the unit circle,

$$x = \cos \theta, \quad y = \sin \theta, \quad \text{or} \quad z = e^{i\theta}, \quad dz = ie^{i\theta} d\theta.$$

$\hat{L}_z$  is also a tangent vector field, which can be represented in the complex plane by a complex function  $f$ :

$$\begin{aligned} f &= i\hbar y + (-i\hbar x)i = \hbar x + i\hbar y = \hbar \cos \theta + i\hbar \sin \theta = \hbar e^{i\theta}. \\ \int_0^{2\pi} f dz &= \int_0^{2\pi} \hbar e^{i\theta} ie^{i\theta} d\theta = i\hbar \int_0^{2\pi} e^{i2\theta} d\theta = i\hbar \left[ \frac{1}{2i} e^{i2\theta} \right]_0^{2\pi} = \frac{\hbar}{2} [e^{i4\pi} - 1] = \frac{\hbar}{2} [\cos 4\pi + i \sin 4\pi - 1] = 0. \end{aligned}$$

We find that both the circulation and flux of  $\hat{L}_z$  around the unit circle in the complex plane  $xy$  are zero.

Since  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$  are symmetric, using the same calculation method as above, we similarly find that both the circulation and flux of  $\hat{L}_x$  and  $\hat{L}_y$  around the unit circle in the complex plane are zero. The fact that both circulation and flux are zero indicates that their curl and divergence are zero, meaning these tangent vector fields are source-free. Therefore, we conclude that precisely because the tangent vector fields represented by the spin operators of the  $SO(3)$  group are source-free, there is no source for generating rest mass, leading to the rest mass of photons being zero.

### 2. Treating Photons as Principal Bundles

We can regard photons as principal bundles. Based on the three different spin operators of photons, there are three distinct types of principal bundles, leading to three different types of photons.

#### (1) Principal Bundle Photon ( $SO(3), BC, \pi, G_{\hat{L}_1}$ )

A photon is an  $SO(3)$  group. The local coordinate representation of the  $SO(3)$  group is given by equation (12.9.4), namely:

$$\begin{aligned} SO(3) &= g(\theta_1, \theta_2, \theta_3) \\ &= ABC = \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix}. \end{aligned} \quad (17.4.1)$$

Matrices  $A, B$ , and  $C$  are each homeomorphic to the unit circle, and they are all one-dimensional smooth manifolds.

Matrix  $A$  is a one-parameter subgroup generated by the spin operator

$$\hat{L}_1 = i\hbar\hat{X}_1 = \hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad (17.4.2)$$

from equation (16.10.2). that is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix} = \exp \left[ -i \frac{1}{\hbar} \theta_1 \hat{L}_1 \right]. \quad (17.4.3)$$

Matrix  $A$  is also a closed subgroup of  $SO(3)$ . According to Theorem 12.5.6, by introducing a unique smooth structure on the closed subgroup  $A$ , it becomes a Lie subgroup of  $SO(3)$ .

According to Theorem 12.13.4,  $\pi : SO(3) \rightarrow SO(3)/A$  is a principal  $A$ -bundle, where  $SO(3)/A$  is a quotient space. The quotient space  $SO(3)/A$  is obtained by regarding all elements of the closed subgroup  $A$  as equivalent elements, or equivalently, by collapsing the unit circle homeomorphic to the closed subgroup  $A$  into a single point. Since, according to equation (17.4.1), the  $SO(3)$  group can be expressed locally as the product of three matrix groups  $A, B$ , and  $C$ , the quotient space  $SO(3)/A$  locally corresponds to the product space  $BC$ . Therefore, we can regard a photon as a principal bundle  $\pi : SO(3) \rightarrow BC$ , denoted as  $(SO(3), BC, \pi, G_{\hat{L}_1})$ . Here,  $SO(3)$  is referred to as the bundle space,  $BC$  is the product space of matrices  $B$  and  $C$  from equation (17.4.1), called the base space, and  $G_{\hat{L}_1}$  denotes the Lie subgroup (matrix group (17.4.3)) generated by the spin operator  $\hat{L}_1$ , called the fiber type or structure group. This is the first type of photon.

### (2)Principal Bundle Photon $(SO(3), AC, \pi, G_{\hat{L}_2})$

Matrix  $B$  is a one-parameter subgroup generated by the spin operator

$$\hat{L}_2 = i\hbar\hat{X}_2 = \hbar \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \quad (17.4.4)$$

from equation (16.10.3). that is

$$B = \begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{pmatrix} = \exp \left[ -i \frac{1}{\hbar} \theta_2 \hat{L}_2 \right]. \quad (17.4.5)$$

Therefore, there also exists a photon that can be represented by another type of principal bundle  $\pi : SO(3) \rightarrow AC$ . Denote this principal bundle as  $(SO(3), AC, \pi, G_{\hat{L}_2})$ . Here,  $AC$  is the product of matrices  $A$  and  $C$  from equation (17.4.1), called the base space, and  $G_{\hat{L}_2}$  denotes the Lie subgroup (17.4.5) generated by the spin operator  $\hat{L}_2$ , referred to as the fiber type or structure group. This is the second type of photon.

### (3)Principal Bundle Photon $(SO(3), AB, \pi, G_{\hat{L}_3})$

Matrix  $C$  is a one-parameter subgroup generated by the spin operator

$$\hat{L}_3 = i\hbar\hat{X}_3 = \hbar \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (17.4.6)$$

from equation (16.10.4). that is

$$C = \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \exp \left[ -i \frac{1}{\hbar} \theta_3 \hat{L}_3 \right]. \quad (17.4.7)$$

Therefore, there also exists a photon that can be represented by another type of principal bundle  $\pi : SO(3) \rightarrow AB$ . Denote this principal bundle as  $(SO(3), AB, \pi, G_{\hat{L}_3})$ , where  $AB$  is the product of matrices  $A$  and  $B$  from equation (17.4.1), referred to as the base space, and  $G_{\hat{L}_3}$  denotes the Lie subgroup (17.4.7) generated by the spin operator  $\hat{L}_3$ , called the fiber type or structure group. This is the third type of photon.

Although we classify photons into three types, their properties are similar, and no significant differences among them are observable at the macroscopic level.

## §17.5 Spin-1/2 Elementary Particles as Principal Bundles

### 1. Circulation and Flux of the Spin Operator $\hat{S}_x$

We calculate the circulation and flux of the spin operator  $\hat{S}_x$  of the  $SU(2)$  group around the unit circle in the complex plane,

$$\hat{S}_x = \frac{1}{2}\hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \quad (17.5.1)$$

We assume  $\hat{S}_x$  is an operator in the complex plane  $uv$ , where  $u$  is the real axis and  $v$  is the imaginary axis. Since the integration curve  $C$  is the unit circle,

$$u = \cos \theta, \quad v = \sin \theta, \quad \text{or} \quad z = e^{i\theta}, \quad dz = ie^{i\theta} d\theta.$$

$\hat{S}_x$  is also a tangent vector field, which can be represented in the complex plane by a complex function  $f$ :

$$\begin{aligned} f &= \frac{1}{2}\hbar v + i \left( \frac{1}{2}\hbar u \right) = \frac{1}{2}\hbar \sin \theta + i \left( \frac{1}{2}\hbar \cos \theta \right). \\ \int_0^{2\pi} f dz &= \frac{1}{2}\hbar \int_0^{2\pi} (\sin \theta + i \cos \theta) ie^{i\theta} d\theta = \frac{1}{2}\hbar \int_0^{2\pi} (-\cos \theta + i \sin \theta) e^{i\theta} d\theta \\ &= \frac{1}{2}\hbar \int_0^{2\pi} [-\cos(-\theta) - i \sin(-\theta)] e^{i\theta} d\theta = -\frac{1}{2}\hbar \int_0^{2\pi} e^{-i\theta} e^{i\theta} d\theta = -\frac{1}{2}\hbar [\theta]_0^{2\pi} = -\pi\hbar. \end{aligned}$$

The calculation shows that the circulation of  $\hat{S}_x$  around the unit circle in the complex plane  $uv$  is  $-\pi\hbar$ , while the flux is zero.

### 2. Circulation and Flux of the Spin Operator $\hat{S}_y$

Next, we compute the circulation and flux of the spin operator  $\hat{S}_y$  of the  $SU(2)$  group around the unit circle in the complex plane,

$$\hat{S}_y = \frac{i}{2}\hbar \left( v \frac{\partial}{\partial u} - u \frac{\partial}{\partial v} \right). \quad (17.5.2)$$

We also assume  $\hat{S}_y$  is an operator in the complex plane  $uv$ . Since the integration curve  $C$  is the unit circle,

$$u = \cos \theta, \quad v = \sin \theta, \quad \text{or} \quad z = e^{i\theta}, \quad dz = ie^{i\theta} d\theta.$$

$\hat{S}_y$  is also a tangent vector field, which can be represented in the complex plane by a complex function  $f$ :

$$f = \frac{i}{2}\hbar v + i \left( -\frac{i}{2}\hbar u \right) = \frac{1}{2}\hbar u + i \left( \frac{1}{2}\hbar v \right) = \frac{1}{2}\hbar \cos \theta + i \left( \frac{1}{2}\hbar \sin \theta \right).$$

$$\begin{aligned}\int_0^{2\pi} f dz &= \frac{1}{2} \hbar \int_0^{2\pi} (\cos \theta + i \sin \theta) i e^{i\theta} d\theta = \frac{i}{2} \hbar \int_0^{2\pi} e^{i\theta} e^{i\theta} d\theta \\ &= \frac{i}{2} \hbar \int_0^{2\pi} e^{2i\theta} d\theta = \frac{i}{2} \hbar \left[ \frac{1}{2i} e^{2i\theta} \right]_0^{2\pi} = \frac{\hbar}{4} [e^{i4\pi} - 1] = 0.\end{aligned}$$

The calculation shows that both the circulation and flux of  $\hat{S}_y$  around the unit circle in the complex plane  $uv$  are zero.

### 3. Circulation and Flux of the Spin Operator $\hat{S}_z$

Finally, we compute the circulation and flux of the spin operator  $\hat{S}_z$  of the  $SU(2)$  group around the unit circle in the complex plane,

$$\hat{S}_z = \frac{1}{2} \hbar \left( u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v} \right). \quad (17.5.3)$$

We also assume  $\hat{S}_z$  is an operator in the complex plane  $uv$ . Since the integration curve  $C$  is the unit circle,

$$u = \cos \theta, \quad v = \sin \theta, \quad \text{or} \quad z = e^{i\theta}, \quad dz = i e^{i\theta} d\theta.$$

$\hat{S}_z$  is also a tangent vector field, which can be represented in the complex plane by a complex function  $f$ :

$$\begin{aligned}f &= \frac{1}{2} \hbar u + i \left( -\frac{1}{2} \hbar v \right) = \frac{1}{2} \hbar \cos \theta - i \left( \frac{1}{2} \hbar \sin \theta \right). \\ \int_0^{2\pi} f dz &= \frac{1}{2} \hbar \int_0^{2\pi} (\cos \theta - i \sin \theta) i e^{i\theta} d\theta = \frac{i}{2} \hbar \int_0^{2\pi} [(\cos(-\theta) + i \sin(-\theta))] e^{i\theta} d\theta \\ &= \frac{i}{2} \hbar \int_0^{2\pi} e^{-i\theta} e^{i\theta} d\theta = \frac{i}{2} \hbar [\theta]_0^{2\pi} = i\pi \hbar.\end{aligned}$$

The calculation shows that the circulation of  $\hat{S}_z$  around the unit circle in the complex plane  $uv$  is zero, while the flux is  $\pi \hbar$ .

### 4. Reasons for Unequal Rest Masses

Based on the above calculation results, we can now understand why photons and neutrinos/antineutrinos have zero rest mass, while electrons/positrons and protons/antiprotons have unequal masses.

The one-parameter differentiable transformation groups determined by  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$  act on the  $SO(3)$  group. However, since both their circulation and flux around the unit circle are zero, this leads to photons having zero rest mass.

The one-parameter differentiable transformation groups determined by  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  act on the  $SU(2)$  group, but their circulation and flux around the unit circle differ. This results in different rest masses for the three types of spin-1/2 elementary particles.

First, spin-1/2 elementary particles (neutrinos/antineutrinos, electrons/positrons, and protons/antiprotons) all possess three spin operators:  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$ . Do these three spin operators act simultaneously? In classical mechanics, we can precisely determine the magnitudes of all three components of angular momentum. However, in the microscopic realm, experiments have shown that we can only measure the magnitude of one of these three spin operators, typically chosen as the operator along the  $z$ -direction. The other two spin operators remain unknown (i.e., their magnitudes cannot be measured). This indicates that these three spin operators play different roles in determining the properties of the particle.

The circulation and flux of  $\hat{S}_y$  around the unit circle are both zero. Zero circulation and flux

indicate that its curl and divergence are zero, meaning the  $\hat{S}_y$  tangent vector field is source-free.

Therefore, we believe that precisely because the tangent vector field represented by  $\hat{S}_y$  is source-free and lacks a source for generating rest mass, this implies that the rest mass of the particle is zero. This particle is the neutrino/antineutrino. Although neutrinos/antineutrinos, as an  $SU(2)$  group, possess three spin operators  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$ , only the spin operator  $\hat{S}_y$  plays a dominant role in determining the properties of neutrinos/antineutrinos.

Let the one-parameter differentiable transformation groups determined by  $\hat{S}_x$  and  $\hat{S}_z$  be denoted as  $S_x$  and  $S_z$ , respectively. Then, the group  $S_x$  plays the dominant role for protons/antiprotons, while the group  $S_z$  plays the dominant role for electrons/positrons. This is because the circulation of  $\hat{S}_x$  around the unit circle is  $-\pi\hbar$ , while its flux is zero; whereas the circulation of  $\hat{S}_z$  around the unit circle is zero, but its flux is  $\pi\hbar$ . Non-zero circulation or flux indicates that their curl or divergence is non-zero, meaning these tangent vector fields are sourced. Consequently, the rest masses of particles influenced by these two spin operators are non-zero.

The assertion that protons/antiprotons are influenced by the one-parameter differentiable transformation group  $S_x$ , while electrons/positrons are influenced by  $S_z$ , is based on the fact that the circulation of  $\hat{S}_x$  around the unit circle is  $-\pi\hbar$ , whereas that of  $\hat{S}_z$  is zero. Although the flux of  $\hat{S}_z$  around the unit circle is  $\pi\hbar$ , which is equal in magnitude to the circulation of  $\hat{S}_x$ , non-zero circulation implies non-zero curl, and non-zero flux implies non-zero divergence. Divergence is a scalar, so changing divergence only requires altering its magnitude. In contrast, curl is a vector, and altering curl necessitates changing both its magnitude and direction. Therefore, modifying curl is more challenging than modifying divergence. For instance, a rapidly spinning bullet does not tumble during flight, whereas a non-spinning bullet is easily overturned by air resistance. This is because altering the state of motion of a bullet with curl is more difficult than that of a bullet without curl.

Thus, the inertial force generated by the tangent vector field  $\hat{S}_x$  (with non-zero curl) is greater than that generated by the tangent vector field  $\hat{S}_z$  (with non-zero divergence). Consequently, the mass produced by  $\hat{S}_x$  is greater than that produced by  $\hat{S}_z$ . Given that the mass of protons/antiprotons is greater than that of electrons/positrons, we conclude that protons/antiprotons are influenced by the one-parameter differentiable transformation group  $S_x$ , while electrons/positrons are influenced by  $S_z$ .

So, can the rest masses of electrons and protons be calculated? Currently, they cannot. However, we can calculate the spin magnetic moments of electrons and protons. Knowing their spin magnetic moments allows us to compute their rest masses. This will be covered in Chapter 20.

## 5. Distinguishing Spin-1/2 Elementary Particles via Principal Bundles

Electrons, protons, neutrinos, and their antiparticles are all  $SU(2)$  groups. The representation of the  $SU(2)$  group is given by equation (12.10.8), namely:

$$U = g(\theta_1, \theta_2, \theta_3) = ABC$$

$$= \begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix} \begin{pmatrix} \cos \frac{\theta_2}{2} & -\sin \frac{\theta_2}{2} \\ \sin \frac{\theta_2}{2} & \cos \frac{\theta_2}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i}{2}\theta_3} & 0 \\ 0 & e^{-\frac{i}{2}\theta_3} \end{pmatrix}. \quad (17.5.4)$$

Matrices  $A$ ,  $B$ , and  $C$  are each homeomorphic to the unit circle and are one-dimensional smooth manifolds.

### (1) Protons and antiprotons as $(SU(2), BC, \pi, G_{\hat{S}_1})$ principal bundles

Matrix  $A$  is a one-parameter subgroup generated by the spin operator

$$\hat{S}_1 = -i\hbar\hat{X}_1 = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (17.5.5)$$

from equation (16.11.2), that is

$$A = \begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix} = \exp \left[ i \frac{1}{\hbar} \theta_1 \hat{S}_1 \right]. \quad (17.5.6)$$

The neighborhood of the identity element in the  $SU(2)$  group is an infinitesimal group or a local Lie group, which is naturally a smooth manifold. Equation (17.5.4) shows that this smooth manifold is the product of three matrices, i.e., three unit circles. Therefore, according to Theorem 12.13.3, protons and antiprotons can be regarded as a principal bundle  $\pi: SU(2) \rightarrow BC$ , denoted as  $(SU(2), BC, \pi, G_{\hat{S}_1})$ . Here,  $BC$  is the product of matrices  $B$  and  $C$  from equation (17.5.4), referred to as the base space, and  $G_{\hat{S}_1}$  denotes the Lie subgroup (17.5.6) generated by the spin operator  $\hat{S}_1$ , called the fiber type or structure group.

### (2) Electrons and positrons as $(SU(2), AB, \pi, G_{\hat{S}_3})$ principal bundles

Matrix  $C$  is also a one-parameter subgroup, generated by the spin operator

$$\hat{S}_3 = -i\hbar\hat{X}_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (17.5.7)$$

from equation (16.11.2), that is

$$C = \begin{pmatrix} e^{\frac{i}{2}\theta_3} & 0 \\ 0 & e^{-\frac{i}{2}\theta_3} \end{pmatrix} = \exp \left[ i \frac{1}{\hbar} \theta_3 \hat{S}_3 \right]. \quad (17.5.8)$$

Based on the analysis above, we also regard electrons and positrons as a principal bundle  $\pi: SU(2) \rightarrow AB$ , denoted as  $(SU(2), AB, \pi, G_{\hat{S}_3})$ . Here,  $AB$  is the product of matrices  $A$  and  $B$  from equation (17.5.4), referred to as the base space, and  $G_{\hat{S}_3}$  denotes the Lie subgroup (17.5.8) generated by the spin operator  $\hat{S}_3$ , called the fiber type or structure group.

### (3) Neutrinos and antineutrinos as $(SU(2), A'C', \pi, G_{\hat{S}_2})$ principal bundles

The  $SU(2)$  group can also be expressed in the form of equation (12.11.15), namely:

$$\begin{aligned} U(\alpha, \beta, \gamma) &= \begin{pmatrix} e^{\frac{i}{2}(\alpha+\gamma)} \cos \frac{\beta}{2} & -e^{\frac{i}{2}(\alpha-\gamma)} \sin \frac{\beta}{2} \\ e^{-\frac{i}{2}(\alpha-\gamma)} \sin \frac{\beta}{2} & e^{\frac{i}{2}(\alpha+\gamma)} \cos \frac{\beta}{2} \end{pmatrix} = A'B'C' \\ &= \begin{pmatrix} e^{\frac{i}{2}\alpha} & 0 \\ 0 & e^{-\frac{i}{2}\alpha} \end{pmatrix} \begin{pmatrix} \cos \frac{\beta}{2} & -\sin \frac{\beta}{2} \\ \sin \frac{\beta}{2} & \cos \frac{\beta}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i}{2}\gamma} & 0 \\ 0 & e^{-\frac{i}{2}\gamma} \end{pmatrix}. \end{aligned} \quad (17.5.9)$$

Matrix  $B'$  is also a one-parameter subgroup, generated by the spin operator

$$\hat{S}_2 = i\hbar\hat{X}_2 = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (17.5.10)$$

from equation (16.11.2), that is

$$B' = \begin{pmatrix} \cos \frac{\beta}{2} & -\sin \frac{\beta}{2} \\ \sin \frac{\beta}{2} & \cos \frac{\beta}{2} \end{pmatrix} = \exp \left[ -i \frac{1}{\hbar} \beta \hat{S}_2 \right]. \quad (17.5.11)$$

Based on the analysis above, we also regard neutrinos and antineutrinos as a principal bundle  $\pi: SU(2) \rightarrow A'C'$ , denoted as  $(SU(2), A'C', \pi, G_{\hat{S}_2})$ . Here,  $A'C'$  is the product of matrices  $A'$

and  $C'$  from equation (17.5.9), referred to as the base space, and  $G_{\hat{S}_2}$  denotes the Lie subgroup (17.5.11) generated by the spin operator  $\hat{S}_2$ , called the fiber type or structure group.

In summary, all elementary particles are principal bundles, and their structure groups are Lie subgroups generated by their respective spin operators. According to Theorem 12.13.2, the structure groups act freely and on the right as Lie transformation groups on the principal bundle. The term "free action," as defined in Definition 12.6.4, means that for any non-identity element  $g$  in the structure group acting on any point  $x$  in the principal bundle, we have  $x \cdot g \neq x$ .

How is  $x \cdot g$  calculated? According to principal bundle theory, any point  $x$  in the principal bundle  $B$  can be expressed as  $x = \psi(p, g)$ , where  $p$  is a point in the base manifold  $M$  and  $g$  is an element of the structure group  $G$ . Let  $h \in G$ , then according to Theorem 12.13.2:

$$\psi(p, g) \cdot h = \psi(p, g \cdot h).$$

This means that the action of  $h$  on  $x$  is equivalent to the action of  $h$  on  $g$ . In other words, the action of an element of the structure group on a point  $x$  in the principal bundle equals the interaction between elements of the structure group. This demonstrates that the structure group generated by the spin operator has a significant influence on the properties of the particle.



## Chapter 18 The Origin of Electromagnetic Fields

### §18.1 The Reason Why Electrons/Positrons and Protons/Antiprotons Carry Charge

#### 1. Gaussian Curvature of $S^2$

The metric of  $R^3$  is flat. In Cartesian coordinates  $(x, y, z)$ , it is given by

$$ds^2 = dx^2 + dy^2 + dz^2, \quad (18.1.1)$$

and in spherical coordinates  $(r, \theta, \phi)$ , it is

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2). \quad (18.1.2)$$

$S^2$  is the set of all points in  $R^3$  at a fixed distance  $r$  from a given point (typically taken as the origin), where  $r$  is a positive real constant. Thus,  $dr^2 = 0$ . Substituting into equation (18.1.2) gives the metric of  $S^2$ :

$$ds^2 = r^2(d\theta^2 + \sin^2 \theta d\phi^2). \quad (18.1.3)$$

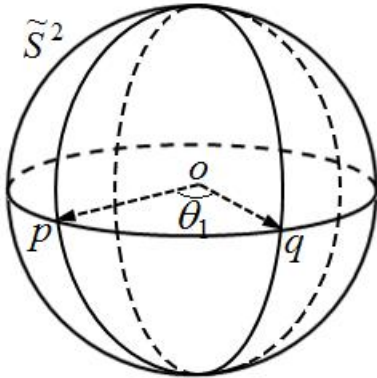
The metric (18.1.3) is called the induced metric on the sphere  $S^2$  from the metric  $R^3$ . From metric (18.1.3), the Gaussian curvature of a sphere with radius  $r$  is found to be

$$K = -\frac{R_{1212}}{g_{11}g_{22} - g_{12}^2} = \frac{1}{r^2}. \quad (18.1.4)$$

#### 2. Concentric Spherical Structure of Electrons/Positrons and Protons/Antiprotons

According to equation (12.10.7), the  $SU(2)$  group is homeomorphic to the unit sphere  $S^3$ . The local representation of the  $SU(2)$  group (see Theorem 12.3.1 and Theorem 12.3.2) is given by equation (17.5.4):

$$U = g(\theta_1, \theta_2, \theta_3) = ABC = \begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix} \begin{pmatrix} \cos \frac{\theta_2}{2} & -\sin \frac{\theta_2}{2} \\ \sin \frac{\theta_2}{2} & \cos \frac{\theta_2}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i}{2}\theta_3} & 0 \\ 0 & e^{-\frac{i}{2}\theta_3} \end{pmatrix}. \quad (18.1.5)$$



**Figure 18.1.1** The meridian circle  $S_p^1$  at point  $p$  rotates eastward by an angle  $\theta_1$

Chapter 17 regards protons/antiprotons as a principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ , with the base space  $BC$  being the product of matrices  $B$  and  $C$ . Let us analyze this base space. Matrices  $A$ ,  $B$ , and  $C$  are each homeomorphic to the unit circle. However, matrix  $B$  can be further interpreted as a transformation group that acts on circle  $C$  by rotating it around its diameter. After one full rotation, circle  $C$  forms a two-dimensional sphere  $S^2$ . Thus, the base space  $BC$  is homeomorphic to the sphere  $S^2$ , representing a spherical surface. Here,  $G_{\hat{S}_1}$  is matrix  $A$ , generated by the spin operator  $\hat{S}_1$  (equation (17.5.5)), and is referred to as the fiber type or structure group—that is, a transformation group.

As shown in Figure 18.1.1, a meridian circle (denoted as  $S_p^1$ ) passes through the equatorial

point  $p$  of  $\tilde{S}^2$ . When moving from point  $p$  to point  $q$ , another meridian circle  $S_q^1$  passes through point  $q$ . This can also be understood as follows: the equator is the orbit of the group  $e^{i\theta}$ . The meridian circle  $S_p^1$  at point  $p$  is represented as  $e^{i\varphi}$ . When the group element  $e^{i\theta_q} = e^{i\theta_1}$  at point  $q$  acts on the meridian circle  $S_p^1$  at point  $p$ ,  $e^{i\varphi}$  transforms into  $e^{i(\varphi+\theta_1)}$ , causing all points on the meridian circle  $S_p^1$  to rotate eastward by an angle  $\theta_1$ . This changes the meridian circle  $S_p^1$  at point  $p$  into the meridian circle  $S_q^1$  at point  $q$ .

Similarly, the action of  $A$  on the sphere  $S^2$  can be represented as  $S^2 \times S$ . This means that, locally,  $S^3$  is homeomorphic to  $S^2 \times S$ .  $S^2 \times S$  can be interpreted as having a sphere  $S^2$  at each point of circle  $A$  (i.e., circle  $S$ ). When moving from one point  $p$  on circle  $A$  to another point  $q$ , the sphere  $S^2$  at point  $p$  transforms into the sphere  $S^2$  at point  $q$ . Alternatively, it can be understood that when the group element at point  $q$  on circle  $A$  acts on the sphere  $S^2$  at point  $p$ , the sphere  $S^2$  at point  $p$  transforms into the sphere  $S^2$  at point  $q$ .

Chapter 17 also regards electrons and positrons as a principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , where the base space  $AB$  is the product of matrices  $A$  and  $B$ . The base space  $AB$  is homeomorphic to the sphere  $S^2$ , meaning it represents a spherical surface  $S^2$ . Here,  $G_{\hat{S}_3}$  is matrix  $C$ , generated by the spin operator  $\hat{S}_3$  (equation (17.5.7)), and is referred to as the structure group, i.e., a transformation group.

Protons/antiprotons are principal bundles:  $(SU(2), BC, \pi, G_{\hat{S}_1})$ , and electrons/positrons are principal bundles:  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , respectively. The base manifolds of these two principal bundles are both homeomorphic to the two-dimensional sphere  $S^2$ , meaning that  $S^3$  is regarded as the principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$  or  $(SU(2), AB, \pi, G_{\hat{S}_3})$  which locally homeomorphics to  $S^2 \times S$ . The curvature (Gaussian curvature) of  $S^2$  is non-zero. Let  $\Omega_\alpha$  be the curvature 2-form of  $S^2$ , then  $\Omega_\alpha \neq 0$ . Substituting  $\Omega_\alpha$  into equation (12.13.10) yields the curvature forms  $\tilde{\Omega}$  of these two principal bundles. Clearly, the curvatures of both principal bundles are non-zero and equal. Consequently, the overall curvature of protons/antiprotons and electrons/positrons is also non-zero and equal.

The properties discussed above are all intrinsic to  $S^3$ . Now let us embed  $S^3$  into Euclidean space to examine its additional properties.

Since  $S^3$  is three-dimensional, it cannot be embedded into three-dimensional Euclidean space  $R^3$ . Instead, it can only be embedded into  $R^4$ , becoming a submanifold of  $R^4$ . Establish a Cartesian coordinate system  $(x, y, z, w)$  in  $R^4$ , where  $x, y, z$  are the coordinates in  $R^3$ . Then, the equation of  $S^3$  in  $R^4$  is

$$x^2 + y^2 + z^2 + w^2 = 1. \quad (18.1.6)$$

Rearranging gives

$$x^2 + y^2 + z^2 = 1 - w^2. \quad (18.1.7)$$

Treat equation (18.1.7) as the equation of a two-dimensional sphere  $S^2$  with its center at the origin of  $R^3$  and radius  $\sqrt{1-w^2}$ . This interpretation aligns with our intuition because  $R^3$  is a realistic, everyday three-dimensional space. Since the range of  $w$  is  $0 \leq w \leq 1$ , we have:

When  $w=0$ , we obtain a two-dimensional sphere  $S^2$  with radius 1.

When  $w=\frac{1}{4}$ , we obtain a two-dimensional sphere  $S^2$  with radius  $\frac{\sqrt{15}}{4}$ .

When  $w = \frac{1}{2}$ , we obtain a two-dimensional sphere  $S^2$  with radius  $\frac{\sqrt{3}}{2}$ .

When  $w = 1$ , we obtain a two-dimensional sphere  $S^2$  with radius 0, i.e., a single point; and so on.

Although it is difficult for us to visualize the specific position of the fourth dimension of  $R^4$ , i.e., the fourth coordinate axis  $w$ , based on the analysis above, we can imagine the fourth dimension  $w$  as the radius of this two-dimensional sphere  $S^2$ .

Since  $w$  can take infinitely many real values within  $[0, 1]$ , there exist infinitely many two-dimensional spheres  $S^2$  with the origin of  $R^3$  as their center and radius  $\sqrt{1-w^2}$ .

In the four-dimensional Minkowski spacetime  $M^4$ , let  $w = ict$ , where  $i$  is the imaginary unit,  $c$  is the speed of light, and  $t$  is time. Then equation (18.1.7) becomes

$$x^2 + y^2 + z^2 = 1 + c^2 t^2. \quad (18.1.8)$$

Let  $r^2 = 1 + c^2 t^2$ . Then we obtain

$$x^2 + y^2 + z^2 = r^2. \quad (18.1.9)$$

Equation (18.1.9) represents the equation of a two-dimensional sphere  $S^2$  with radius  $r$  in  $R^3$ , where the radius of  $S^2$  increases with time  $t$ . As time  $t$  takes different values, spheres  $S^2$  with different radii are obtained. Since time  $t$  is continuous, there are infinitely many spheres  $S^2$  with varying radii. All these spheres are centered at the origin of  $R^3$ , and their radii continuously expand with the passage of time. However, when time  $t = 0$ ,  $r = 1$ . Therefore, we denote the collection of these infinitely many spheres  $S^2$ , centered at the origin of  $R^3$  with radii ranging from 1 to  $r = \sqrt{1 + c^2 t^2}$  ( $t > 0$ ), as  $B^3$ ,

$$B^3 = \{(x, y, z) \in R^3 \mid 1 \leq x^2 + y^2 + z^2 \leq \sqrt{1 + c^2 t^2} (t > 0)\}.$$

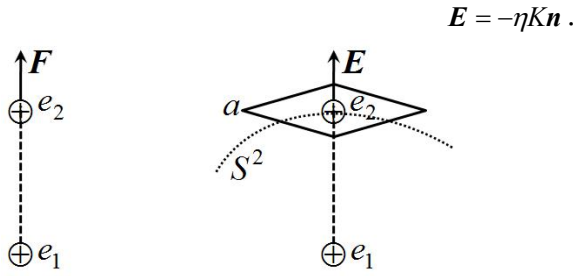
Since we have assumed that electrons, protons, and their antiparticles are all  $SU(2)$  groups, electrons, protons, and their antiparticles can be regarded as  $B^3$  in  $M^4$ , consisting of infinitely many concentric spheres  $S^2$  centered at the origin of  $R^3$ , with radii starting from 1 and gradually increasing.

### 3. Electric Field Strength

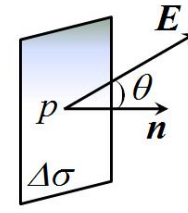
We introduce the concept of electric field strength  $E$ . Since electric field strength  $E$  is a vector, we must define its direction. As shown in Figure 18.1.2, two positively charged protons  $e_1$  and  $e_2$  have centers separated by a distance  $r$ . These two positively charged protons repel each other. For example, proton  $e_2$  is repelled by proton  $e_1$ , and the direction of this repulsive force  $F$  is defined as the direction of the electric field strength  $E$  at proton  $e_2$ . Since electrons/positrons and protons/antiprotons can be regarded as a collection  $B^3$  consisting of infinitely many concentric spherical surfaces, at the center  $e_2$  of proton  $e_2$ , there is a concentric spherical surface belonging to proton  $e_1$ . The Gaussian curvature  $K$  (also the sectional curvature of the tangent plane  $a$ ) of this concentric spherical surface at point  $e_2$  is defined as the magnitude of the electric field strength  $E$  produced by proton  $e_1$  at point  $e_2$ . Let the direction from the center of proton  $e_1$  to the center of proton  $e_2$  be the direction of the unit vector  $n$ . Then, the electric field strength produced by proton  $e_1$  at point  $e_2$  is

$$E = \eta K n,$$

where  $\eta$  is a proportionality constant. If the particle at point  $e_1$  is an electron instead of a proton, then since a proton is attracted by electron  $e_1$ , the electric field strength produced by electron  $e_1$  at point  $e_2$  is



**Figure 18.1.2** Repulsive force  $F$  on proton  $e_2$  due to proton  $e_1$



**Figure 18.1.3** Electric flux through surface element  $\Delta\sigma$

#### 4. Electric Flux

As shown in Figure 18.1.3, let the electric field intensity at a point  $p$  in the electric field be  $E$ . Consider a surface element (area element)  $\Delta\sigma$  containing point  $p$ , where  $n$  is the unit normal vector (normal vector) of the surface element, and  $\theta$  is the angle between  $E$  and  $n$ . Then the electric flux through the surface element  $\Delta\sigma$  is

$$\Delta\Phi = E\Delta\sigma \cos\theta = E \cdot \Delta\sigma n = E \cdot \Delta\sigma.$$

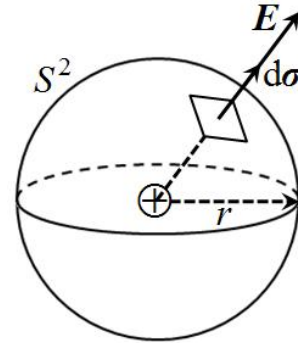
That is, the electric flux through the surface element  $\Delta\sigma$  equals the product of the electric field intensity  $E$  and the projection of the surface element vector  $\Delta\sigma$  in the direction of the field intensity. If the surface element is expressed in differential form  $d\sigma$ , then the electric flux through  $d\sigma$  is

$$d\Phi = E d\sigma \cos\theta = E \cdot d\sigma n = E \cdot d\sigma.$$

$d\Phi$  is a second-order exterior differential form.

#### 5. Gauss's Theorem for a Single Charge

As previously established, electrons, protons, and their antiparticles can be regarded in  $M^4$  as a collection  $B^3$  consisting of infinitely many concentric spherical surfaces  $S^2$  centered at the origin of  $R^3$ , with radii starting from 1 and gradually increasing. Since each spherical surface  $S^2$  is a compact, orientable two-dimensional Riemannian manifold, we may select any one of these infinitely many two-dimensional concentric spheres, denoted as  $S^2$ , as shown in Figure 18.1.4. Let the radius of this sphere be  $r$ . According to the Gauss-Bonnet theorem (Theorem 11.3.1), we have



**Figure 18.1.4**  $E d\sigma$  on the sphere

$$\int_{S^2} K d\sigma = 2\pi\chi, \quad (18.1.10)$$

where  $K$  is the Gaussian curvature on  $S^2$ ,  $d\sigma$  is the oriented area element on  $S^2$ , and  $\chi$  is the Euler characteristic of  $S^2$ . For  $S^2$ ,  $\chi = 2$ . Taking the outward normal vector  $n$  as the positive orientation of  $S^2$ , we rewrite equation (18.1.10) in vector form:

$$\int_{S^2} \eta K n \cdot n d\sigma = 2\pi\eta\chi. \quad (18.1.11)$$

Equation (18.1.11) is similar to Gauss's theorem in electromagnetic field theory:

$$\oint_{S^2} E \cdot d\sigma = \frac{Q}{\epsilon_0}. \quad (18.1.12)$$

Therefore, we compare equation (18.1.11) with Gauss's theorem. By comparing, we treat the Gaussian curvature  $K$  as the magnitude of the electric field intensity of electrons, protons, and their antiparticles, i.e., let

$$\mathbf{E} = \eta K \mathbf{n}, \quad (18.1.13)$$

and

$$d\sigma = n d\sigma, \quad 2\pi\eta\chi = \frac{Q}{\epsilon_0}.$$

Furthermore,

$$Q = 2\pi\eta\chi\epsilon_0 = 4\pi\eta\epsilon_0. \quad (18.1.14)$$

Equation (18.1.11) becomes

$$\int_{S^2} \mathbf{E} \cdot d\sigma = \frac{Q}{\epsilon_0}. \quad (18.1.15)$$

Substituting equation (18.1.14) into equation (18.1.13) yields

$$\mathbf{E} = \eta K \mathbf{n} = \frac{Q}{4\pi\epsilon_0} \frac{1}{r^2} \mathbf{n} = \frac{Q\mathbf{r}}{4\pi\epsilon_0 r^3}, \quad (18.1.16)$$

where  $K = \frac{1}{r^2}$  is the Gaussian curvature of the sphere  $S^2$ .

## 6. The Origin of Positive and Negative Charges

In  $M^4$ , electrons, protons, and their antiparticles can all be regarded as a collection  $B^3$  consisting of infinitely many concentric spherical surfaces  $S^2$  centered at the origin of  $R^3$ , with radii starting from 1 and gradually increasing. Whether an electron (proton, positron, or antiproton) is treated as a principal bundle or as  $B^3$  in  $M^4$ , their orientations should be the same.

From equation (18.1.15), we obtain

$$-\int_{S^2} \mathbf{E} \cdot d\sigma = -\frac{Q}{\epsilon_0},$$

in classical electromagnetic theory, from this equation we obtain

$$\int_{S^2} (-\mathbf{E}) \cdot d\sigma = -\frac{Q}{\epsilon_0},$$

this is a phenomenological description. However, to explain why electrons and protons have different charges, from this equation we should instead obtain

$$\int_{S^2} \mathbf{E} \cdot (-d\sigma) = -\frac{Q}{\epsilon_0}.$$

Based on experimental observations, for protons and positrons, we set  $Q > 0$ , i.e., we define them as carrying positive charges. Setting  $Q > 0$  means setting  $\chi > 0$ . For antiprotons and electrons, we set  $Q < 0$ , i.e., we define them as carrying negative charges. Setting  $Q < 0$  means setting  $\chi < 0$ .

Why do positive and negative charges arise? This is due to their different orientations. The orientation of electrons and protons aligns with the left-hand rule, which is referred to as the positive orientation. In a local Cartesian coordinate system  $(U; x^i)$ , their orientation is represented by the 3-form differential  $-dx \wedge dy \wedge dz$ . Although electrons and protons share the same orientation, there are still differences, as compared below:

Electron	Proton
$(-dx \wedge dy) \wedge dz = -d\sigma \wedge dz$	$(dx \wedge dy) \wedge (-dz) = d\sigma \wedge (-dz)$

In the above equations,  $d\sigma = dx \wedge dy$ . Electrons correspond to the principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , while protons correspond to the principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ .

However, locally, both are product spaces  $S^2 \times S$ , where the sphere  $S^2$  serves as the base space. For electrons, the orientation of  $S^2$  is determined by  $-d\sigma = -dx \wedge dy$ . In equation (18.1.15), the

direction of  $\mathbf{E}$  is opposite to that of  $d\sigma$ , i.e.,  $\mathbf{E} \cdot d\sigma < 0$ . Therefore, according to equation (18.1.15),  $Q < 0$ , indicating that electrons carry a negative charge. For protons, the orientation of  $S^2$  is determined by  $d\sigma = dx \wedge dy$ . In equation (18.1.15), the direction of  $\mathbf{E}$  aligns with that of  $d\sigma$ , i.e.,  $\mathbf{E} \cdot d\sigma > 0$ . Thus, according to equation (18.1.15),  $Q > 0$ , indicating that protons carry a positive charge. The orientation of the structure group  $G_{\hat{S}_3}$  for electrons is determined by  $dz$ , while the orientation of the structure group  $G_{\hat{S}_1}$  for protons is determined by  $-dz$ . Their orientations are opposite because their directions of rotation are opposite.

Positrons and antiprotons have an orientation consistent with the right-hand rule, referred to as the reverse orientation. In a local Cartesian coordinate system  $(U; x^i)$ , their orientation is represented by the 3-form differential  $dx \wedge dy \wedge dz$ . Although positrons and antiprotons share the same orientation, there are still differences, as compared below:

Positron	Antiproton
$(dx \wedge dy) \wedge dz = d\sigma \wedge dz$	$(-dx \wedge dy) \wedge (-dz) = -d\sigma \wedge (-dz)$

Positrons, like electrons, correspond to the principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , while antiprotons, like protons, correspond to the principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ . However, their base spaces are both the sphere  $S^2$ . For positrons, the orientation of  $S^2$  is determined by  $d\sigma = dx \wedge dy$ . In equation (18.1.15), the direction of  $\mathbf{E}$  aligns with that of  $d\sigma$ , i.e.,  $\mathbf{E} \cdot d\sigma > 0$ . Therefore, according to equation (18.1.15),  $Q > 0$ , indicating that positrons carry a positive charge. For antiprotons, the orientation of  $S^2$  is determined by  $-d\sigma = -dx \wedge dy$ . In equation (18.1.15), the direction of  $\mathbf{E}$  is opposite to that of  $d\sigma$ , i.e.,  $\mathbf{E} \cdot d\sigma < 0$ . Thus, according to equation (18.1.15),  $Q < 0$ , indicating that antiprotons carry a negative charge. The orientation of the structure group  $G_{\hat{S}_3}$  for positrons is determined by  $dz$ , while the orientation of the structure group  $G_{\hat{S}_1}$  for antiprotons is determined by  $-dz$ . Their orientations are opposite because their directions of rotation are opposite.

Setting  $Q > 0$  means setting  $\chi > 0$ . Similarly, setting  $Q < 0$  means setting  $\chi < 0$ . When  $Q < 0$ , equation (18.1.11) should be

$$\int_{S^2} -\eta K \mathbf{n} \cdot \mathbf{n} d\sigma = 2\pi\eta\chi, \quad (18.1.17)$$

and equation (18.1.15) should be

$$\oint_{S^2} \mathbf{E} \cdot d\sigma = -\frac{Q}{\epsilon_0}. \quad (18.1.18)$$

Whether for electrons, protons, positrons, or antiprotons, in calculations, we conventionally take the outward normal vector  $\mathbf{n}$  of the integration surface  $S^2$  as the positive direction (positive orientation), i.e.,  $d\sigma$  is positive. Therefore, for positively charged positrons and protons, the electric field intensity  $\mathbf{E}$  is also positive; for negatively charged electrons and antiprotons, the electric field intensity is negative,  $-\mathbf{E}$ .

The orientations of the structure groups for electrons and positrons are the same, both determined by  $dz$ ; similarly, the orientations of the structure groups for protons and antiprotons are also the same, both determined by  $-dz$ .

The orientations of the structure groups of protons (antiprotons) and electrons (positrons) are opposite, which can be seen from the matrix forms of their structure groups. The structure group of protons (antiprotons) is the  $A$  matrix in equation (18.1.5),

$$A = \begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix}.$$

We determine the direction of rotation for matrix  $A$ . Suppose there is a vector  $(1,0)$  in the complex plane. Under the action of matrix  $A$ , it rotates by an angle  $\pi$ , i.e.,  $\theta_1 = \pi$ . Then we have

$$\begin{pmatrix} \cos \frac{\theta_1}{2} & i \sin \frac{\theta_1}{2} \\ i \sin \frac{\theta_1}{2} & \cos \frac{\theta_1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \frac{\pi}{2} & i \sin \frac{\pi}{2} \\ i \sin \frac{\pi}{2} & \cos \frac{\pi}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ i \end{pmatrix}.$$

The vector  $(1,0)$  on the horizontal axis  $u$  transforms into the vector  $(0,i)$  on the vertical axis  $v$ . From this, it can be seen that the vector  $(1,0)$  rotates counterclockwise to become the vector  $(0,i)$ .

The structure group of electrons (positrons) is the  $C$  matrix in equation (18.1.5),

$$C = \begin{pmatrix} e^{\frac{i}{2}\theta_3} & 0 \\ 0 & e^{-\frac{i}{2}\theta_3} \end{pmatrix}.$$

We determine the direction of rotation for matrix  $C$ . Suppose there is a vector  $(0,i)$  in the complex plane. Under the action of matrix  $C$ , it rotates by an angle  $\pi$ , i.e.,  $\theta_3 = \pi$ . Then we have

$$\begin{pmatrix} e^{\frac{i}{2}\theta_3} & 0 \\ 0 & e^{-\frac{i}{2}\theta_3} \end{pmatrix} \begin{pmatrix} 0 \\ i \end{pmatrix} = \begin{pmatrix} e^{\frac{i\pi}{2}} & 0 \\ 0 & e^{-\frac{i\pi}{2}} \end{pmatrix} \begin{pmatrix} 0 \\ i \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 0 \\ i \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The vector  $(0,i)$  on the vertical axis  $v$  transforms into the vector  $(1,0)$  on the horizontal axis  $u$ . From this, it is evident that the vector  $(0,i)$  rotates clockwise to become the vector  $(1,0)$ .

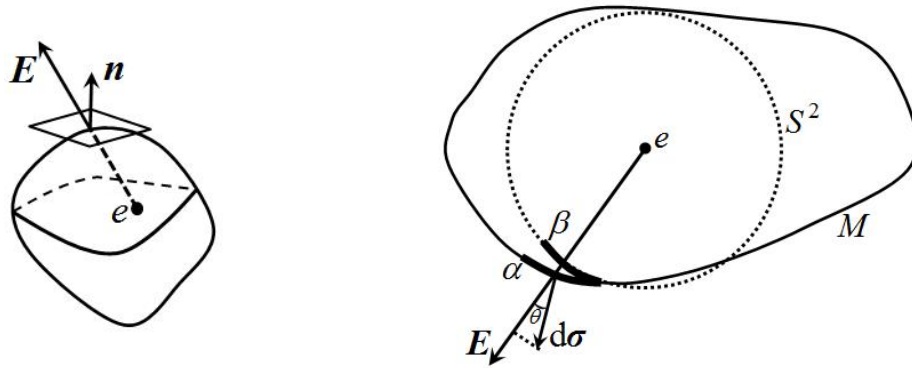
From the above analysis, it can be seen that the directions of rotation described by matrix  $A$  and matrix  $C$  are exactly opposite. This indicates that their orientations are opposite, and thus it also shows that the orientations of the structure groups of protons (antiprotons) and electrons (positrons) are opposite.

## 7.The Reason for Charge Conservation

Equation (18.1.14) tells us that the charge  $Q$  is a constant, and its magnitude is closely related to the topological structure of the spherical surface—specifically, to the Euler characteristic  $\chi$ . If  $\chi = 0$ , the charge  $Q$  would be zero. If  $\chi = 1$ , the charge  $Q$  would be halved. However, since the Euler characteristic  $\chi$  is a topological invariant, as long as electrons, protons, and their antiparticles remain the same particles after any transformation, the magnitude of their charge will not change. For example, when these four types of particles undergo free motion without transforming into other particles, each spherical surface  $S^2$  remains  $S^2$  under translation or rotation, and its Euler characteristic is fixed at 2. Therefore, the charge remains unchanged during free motion of these particles. This is the reason for charge conservation.

A homeomorphic transformation can be applied to the spherical surface  $S^2$ , transforming it into other forms of closed surfaces, such as the closed surface shown in the left diagram of Figure 18.1.5. Since this is a homeomorphic transformation, the Euler characteristic of the closed surface remains 2. This also explains charge conservation. For any closed surface, the Gauss-Bonnet theorem still holds, allowing for analysis similar to the above. The conclusions drawn from such analysis are consistent with those obtained earlier.





**Figure 18.1.5** The Gauss-Bonnet theorem holds for any closed surface

As shown in the right diagram of Figure 18.1.5, an arbitrary closed surface  $M$  encloses a proton  $e$ . The sphere  $S^2$  is a two-dimensional spherical surface with the proton  $e$  at its center. The bold line segment on the sphere  $S^2$  represents a surface element  $d\beta$  on the sphere. Since the electric field intensity  $E$  is parallel to the surface element  $d\beta$ , equation (18.1.15) gives

$$\int_{S^2} E \cdot d\beta = \int_{S^2} E d\beta = \frac{e}{\epsilon_0}.$$

The thick line segment on  $M$  represents a surface element  $d\alpha$  on the closed surface  $M$ , and the angle between  $d\alpha$  and the electric field intensity  $E$  is  $\theta$ . The projection of the surface element  $d\alpha$  onto the sphere  $S^2$  is the surface element  $d\beta$ , with  $d\beta = \cos\theta d\alpha$ . Therefore,

$$\int_{S^2} E \cdot d\beta = \int_{S^2} E d\beta = \int_M E \cos\theta d\alpha = \frac{e}{\epsilon_0},$$

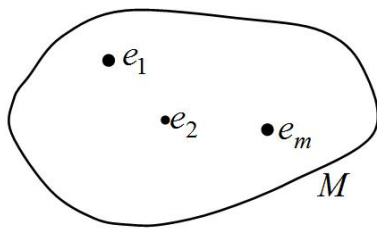
which, on the closed surface  $M$ , gives

$$\int_M E \cos\theta d\alpha = \int_M E \cdot d\alpha = \frac{e}{\epsilon_0}. \quad (18.1.19)$$

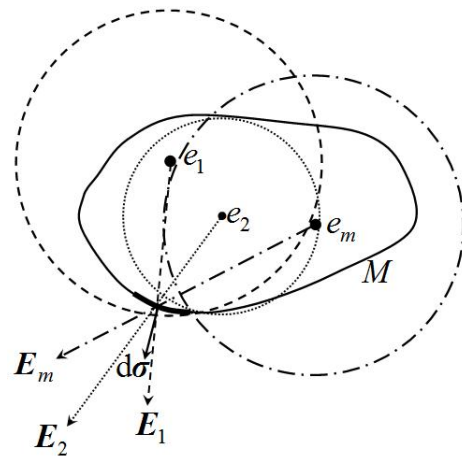
Here,  $E \cdot d\alpha$  represents the electric flux through the surface element  $d\alpha$ , i.e.,

$$d\Phi = E \cdot d\alpha = E \cos\theta d\alpha.$$

### 8. Gauss's Theorem for Multiple Charges



**Figure 18.1.6** Closed Surface Enclosing Multiple Charges



**Figure 18.1.7** Gauss's Theorem for Multiple Charges

If an arbitrary closed surface  $M$  encloses several charged particles, as shown in Figure 18.1.6, we can adopt a method of analysis similar to that illustrated in Figure 18.1.5. As shown in Figure 18.1.7, we obtain



$$\int_M \mathbf{E}_1 \cdot d\boldsymbol{\sigma} = \frac{e_1}{\epsilon_0}, \quad \int_M \mathbf{E}_2 \cdot d\boldsymbol{\sigma} = \frac{e_2}{\epsilon_0}, \quad \dots, \quad \int_M \mathbf{E}_m \cdot d\boldsymbol{\sigma} = \frac{e_m}{\epsilon_0}.$$

Summing these equations gives

$$\int_M \mathbf{E}_1 \cdot d\boldsymbol{\sigma} + \int_M \mathbf{E}_2 \cdot d\boldsymbol{\sigma} + \dots + \int_M \mathbf{E}_m \cdot d\boldsymbol{\sigma} = \frac{e_1}{\epsilon_0} + \frac{e_2}{\epsilon_0} + \dots + \frac{e_m}{\epsilon_0},$$

that is

$$\int_M (\mathbf{E}_1 + \mathbf{E}_2 + \dots + \mathbf{E}_m) \cdot d\boldsymbol{\sigma} = \sum_{i=1}^m \frac{e_i}{\epsilon_0},$$

or

$$\int_M \mathbf{E} \cdot d\boldsymbol{\sigma} = \sum_{i=1}^m \int_M \mathbf{E}_i \cdot d\boldsymbol{\sigma} = \sum_{i=1}^m \frac{Q_i}{\epsilon_0}, \quad (18.1.20)$$

where

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \dots + \mathbf{E}_m. \quad (18.1.21)$$

Here,  $\mathbf{E}$  is the total electric field intensity produced by all charged particles inside the closed surface  $M$ . Equation (18.1.21) is the mathematical representation of the superposition principle for electric field intensity.

Equation (18.1.20) represents the case where an arbitrary closed surface  $M$  contains multiple charges, which may include both positive and negative charges. If the charges are continuously distributed inside the closed surface  $M$ , the right-hand side of equation (18.1.20) can be expressed in integral form as

$$\int_M \mathbf{E} \cdot d\boldsymbol{\sigma} = \frac{1}{\epsilon_0} \int_V \rho dV, \quad (18.1.22)$$

where  $V$  is the volume enclosed by  $M$ , and  $\rho$  is the charge density.

When two or more particles combine, they are linearly superposed, i.e., they exist in the form of an inner direct product group. The inner direct product group is the disjoint union of the elements of the groups involved in the inner direct product operation.

For example, consider two groups  $G_1 = \{a, b, c, d\}$  and  $G_2 = \{\alpha, \beta, \gamma, \lambda\}$ . Their inner direct product group is

$$G_1 \otimes G_2 = \{(a, \alpha), (b, \beta), (c, \gamma), (d, \lambda)\}.$$

This inner direct product group can be viewed as first taking the disjoint union of the elements of the two groups, forming a union set

$$\{a, b, c, d, \alpha, \beta, \gamma, \lambda\},$$

and then classifying the elements in this union set

$$\{(a, \alpha), (b, \beta), (c, \gamma), (d, \lambda)\},$$

Therefore, the inner direct product group is the disjoint union (with no increase or decrease in elements) of the elements of the groups involved in the inner direct product operation.

Precisely because the inner direct product group is the disjoint union of the elements of the groups involved, according to Theorem 1.10.1, when two or more particles combine, their Euler characteristics can be added. Since we assign positive or negative signs to the Euler characteristic, the Euler characteristic of a composite particle becomes the algebraic sum of the Euler characteristics of its constituent particles.

If the elements of the groups representing the constituent particles are combined in a non-disjoint manner, fractional charges may arise, as suggested by Theorem 1.10.1.

In summary, electrons, protons, and their antiparticles carry electric charges because they are all  $SU(2)$  groups, homeomorphic to  $S^3$ . Specifically, protons and antiprotons correspond to the principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ , while electrons and positrons correspond to the principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ . Consequently, the curvature inside  $S^3$  is non-zero. Since  $S^3$  is equivalent to a collection of infinitely many two-dimensional spheres  $S^2$  in  $R^3$ , centered at the origin of  $R^3$  with radius  $(1 + c^2 t^2)^{1/2}$ , and each  $S^2$  satisfies the Gauss-Bonnet theorem, the

electric field intensity is proportional to the Gaussian curvature of  $S^2$ . Furthermore, since the Euler characteristic of each  $S^2$  is  $\chi = 2$ , the magnitudes of the charges carried by these four types of particles are equal. The opposite directions of their charges are related to their orientations. Because the Euler characteristic of each  $S^2$ , as well as every closed surface homeomorphic to  $S^2$ , equals 2, the magnitudes of the charges of electrons, protons, and their antiparticles are equal. Moreover, since the Euler characteristic is a topological invariant, charge is conserved.

There exists a close relationship between the spin operator and electric charge. Protons (antiprotons) are a principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ , with the structure group  $G_{\hat{S}_1}$ . Electrons (positrons) are a principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , with the structure group  $G_{\hat{S}_3}$ . Precisely because their structure groups are  $G_{\hat{S}_1}$  and  $G_{\hat{S}_3}$ , respectively, their base manifolds are both a two-dimensional sphere  $S^2$ , leading to the presence of an electric field within these two types of particles. Additionally, the opposite orientations of these two structure groups result in opposite charges for the two types of particles. The structure groups  $G_{\hat{S}_1}$  and  $G_{\hat{S}_3}$  are determined by the spin operators  $\hat{S}_x$  and  $\hat{S}_z$ , respectively. Therefore, the charge properties of these two types of particles are determined by the spin operators  $\hat{S}_x$  and  $\hat{S}_z$ .

Since the left or right translations generated by the group elements of  $SU(2)$  are diffeomorphisms or isomorphisms of  $SU(2)$  itself, and the spin operators of electrons/positrons and protons/antiprotons are left-invariant (or right-invariant) tangent vector fields that remain invariant under left or right translations, the charge of electrons/positrons and protons/antiprotons remains invariant in a free state.

## §18.2 The Reason Why Neutrinos and Antineutrinos Carry No Charge

Neutrinos and antineutrinos are both  $SU(2)$  groups. Although the  $SU(2)$  group is homeomorphic to the unit sphere  $S^3$ , locally it can also be viewed as consisting of infinitely many two-dimensional tori  $T^2$ . This is precisely the reason why neutrinos and antineutrinos carry no charge.

### 1. Gaussian Curvature of the Torus

The torus  $T^2$  has two Riemannian structures, and their Gaussian curvatures differ.

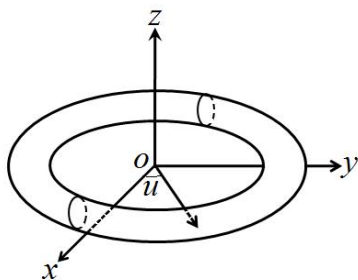


Figure 18.2.1 Torus

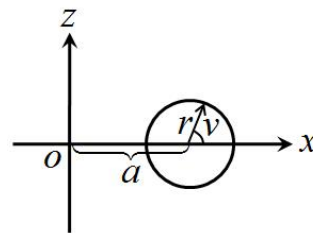


Figure 18.2.2 Parameters of the torus

The first Riemannian structure is that of the torus as a submanifold of  $R^3$ , endowed with the Riemannian metric induced from  $R^3$ . As shown in Figures 18.2.1 and 18.2.2, the parametric equations of the torus are

$$x = (a + r \cos v) \cos u, \quad y = (a + r \cos v) \sin u, \quad z = r \sin v, \\ 0 < r < a, \quad 0 \leq u, v < 2\pi.$$

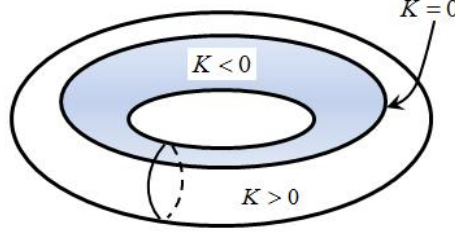
The metric of the torus is

$$ds^2 = (a + r \cos v)^2 du^2 + r^2 dv^2.$$

The Gaussian curvature of the torus is

$$K = \frac{\cos v}{r(a + r \cos v)},$$

This expression shows that the Gaussian curvature on the torus is not uniformly constant. The approximate distribution of Gaussian curvature across the torus is illustrated in Figure 18.2.3.

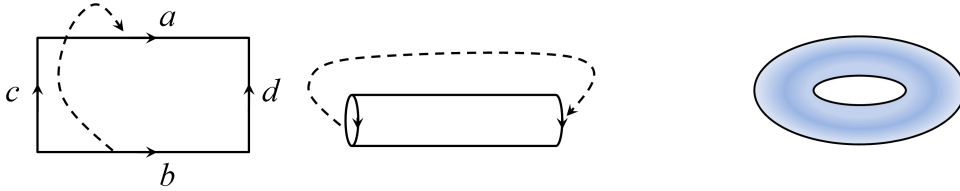


**Figure 18.2.3** Distribution of Gaussian curvature on a torus in  $R^3$

The second Riemannian structure defines the torus as the direct product of two circles. This definition can be found in Example 3.4.2. Since the torus  $T^2$  is the direct product of two circles,  $T^2 = S^1 \times S^1$ , its Gaussian curvature is zero, i.e.,  $K = 0$ . This is because  $T^2$  can be viewed as a rectangle (like a sheet of paper) on a plane  $R^2$  formed by folding opposite edges together. As shown in Figure 18.2.4, first, the edges  $a$  and  $b$  are glued together to form a cylindrical surface, and then the two circular edges of the cylinder are aligned and glued together in the same direction to form a torus  $T^2$ . Since the Gaussian curvature  $K$  of the rectangle on the plane  $R^2$  is zero, the Gaussian curvature  $K$  of the resulting cylindrical surface is also zero. Further, the Gaussian curvature  $K$  of the torus formed by rolling the cylinder remains zero. That is, because the mapping

$$R^2 \rightarrow S^1 \times S^1$$

is a local isometry, and the rectangle in  $R^2$  has  $K = 0$ , the Gaussian curvature  $K$  of the torus is locally zero as well.



**Figure 18.2.4** Edges  $a$  and  $b$  glued into a cylinder, then edges  $c$  and  $d$  glued into a torus

The Gaussian curvature  $K$  of the torus being locally zero does not imply that the torus is globally isometric to the plane, because the torus has a hole in the middle and is not homeomorphic to  $R^2$ . Since the Gaussian curvature of the torus as the direct product of two circles ( $T^2 = S^1 \times S^1$ ) is locally zero, it cannot be embedded into  $R^3$  as a submanifold (because the Gaussian curvature of a torus in  $R^3$  is locally positive in some regions and negative in others, as shown in Figure 18.2.3). Instead, it can only be embedded into  $R^4$  as a submanifold.

## 2.The Reason Why Neutrinos and Antineutrinos Carry No Charge

The expression for the  $SU(2)$  group can also take the form of equation (17.5.9), namely:

$$U(\alpha, \beta, \gamma) = A'B'C' = \begin{pmatrix} e^{\frac{i}{2}\alpha} & 0 \\ 0 & e^{-\frac{i}{2}\alpha} \end{pmatrix} \begin{pmatrix} \cos \frac{\beta}{2} & -\sin \frac{\beta}{2} \\ \sin \frac{\beta}{2} & \cos \frac{\beta}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i}{2}\gamma} & 0 \\ 0 & e^{-\frac{i}{2}\gamma} \end{pmatrix}. \quad (18.2.1)$$

In Chapter 17, neutrinos and antineutrinos are regarded as a principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$ , where  $\pi: SU(2) \rightarrow A'C'$  is a smooth mapping. The matrices  $A'$ ,  $B'$ , and

$C'$  in equation (18.2.1) are each homeomorphic to the unit circle. The base space  $A'C'$  is the product of matrices  $A'$  and  $C'$ . Since  $A'$  and  $C'$  are both diagonal matrices, their product is commutative, i.e.,  $A'C' = C'A'$ . Thus, the base space  $A'C'$  represents a torus  $T^2$ .

Let  $\{U_\alpha; \alpha \in I\}$  be an open cover of  $T^2$ , where  $I$  is the set of natural numbers,  $I = \{1, 2, \dots\}$ . On the principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$ , there exists a diffeomorphism  $\psi_\alpha: U_\alpha \times G_{\hat{S}_2} \rightarrow \pi^{-1}(U_\alpha)$  for each  $\alpha \in I$  satisfying the condition:

$$\pi \circ \psi_\alpha(p, g) = p, \quad \forall (p, g) \in U_\alpha \times G_{\hat{S}_2}.$$

The meaning of this mapping is that the principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$  is locally (over  $\pi^{-1}(U_\alpha)$ ) a product space  $U_\alpha \times G_{\hat{S}_2}$ . Since  $U_\alpha$  is an open subset of  $T^2$ , its Gaussian curvature  $K$  is zero, and the curvature 2-form  $\Omega_\alpha = 0$ . Therefore,  $U_\alpha$  is flat. Substituting  $\Omega_\alpha = 0$  into equation (12.13.10), we find that the curvature form  $\tilde{\Omega}$  of the principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$  is also zero. Hence, the principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$  is locally flat.

We have interpreted the electric field intensity as being proportional to the Gaussian curvature  $K$ . Since the Gaussian curvature  $K$  of the principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$  is locally zero, the electric field intensity of neutrinos and antineutrinos is zero, meaning they carry no electric charge. This is the reason why neutrinos and antineutrinos are electrically neutral.

The properties discussed above are intrinsic to neutrinos and antineutrinos and are independent of Euclidean space. Even if neutrinos and antineutrinos are embedded into Euclidean space  $R^4$ , they remain electrically neutral.

Neutrinos and antineutrinos are both homeomorphic to the  $SU(2)$  group. From the properties of the  $SU(2)$  group, we can derive the following equations:

$$x^2 + y^2 + z^2 + w^2 = 1. \quad (18.2.2)$$

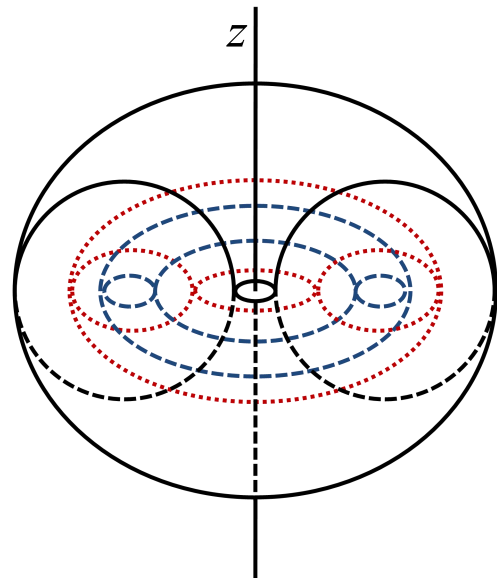
This equation can be interpreted as the equation of the unit sphere  $S^3$  in  $R^4$ , or alternatively as the equation describing a collection of tori  $T^2$  of varying sizes in  $R^4$ , as follows:

$R^4$  can be regarded as the direct product of two complex planes. Therefore, we analyze this equation using complex variables. Let  $u = x + iy$  and  $v = z + iw$ , then equation (18.2.2) transforms into

$$|u|^2 + |v|^2 = 1.$$

When the modulus of  $u$  is fixed at a certain real value, such as  $|u| = d$  ( $0 \leq d \leq 1$ ), the modulus of  $v$  is also fixed. Here,  $|u| = d$  implies that the variables  $x$  and  $y$  form a circle of radius  $d$ , while  $z$  and  $w$  form a circle of radius  $\sqrt{1-d^2}$ . The direct product of these two circles yields a torus. When  $d$  takes another positive value, a corresponding torus is obtained. Since  $d$  can take any value in the interval  $[0, 1]$ , there are infinitely many tori of different sizes corresponding to these values, collectively constituting the three-sphere  $S^3$ .

The three-sphere  $S^3$  can be viewed as the union of three-dimensional Euclidean space  $R^3$  with a point at infinity, i.e.,  $R^3 \cup \{\infty\} = S^3$ . Therefore, we can represent the toroidal structure of



**Figure 18.2.5** The three-sphere can be seen as nested two-dimensional tori

$S^3$  in  $R^3$ , as illustrated in Figure 18.2.5. Figure 18.2.5 visualizes the structure of  $S^3$  through a transformation into  $R^3$ . The method involves using polar coordinates to express  $|u|^2 + |v|^2 = 1$  as

$$r_0^2 + r_1^2 = 1,$$

where  $u = r_0 e^{i\theta_0}$ ,  $v = r_1 e^{i\theta_1}$ , and defining the mapping

$$p(u, v) = \frac{r_0}{r_1} e^{i(\theta_0 - \theta_1)}, \quad \lambda = \frac{r_0}{r_1}.$$

Here,  $r_0$  is assumed to be the radius of the vertical circle, and  $r_1$  the radius of the horizontal circle. When  $r_0$  and  $r_1$  are fixed,  $\lambda$  is also fixed, yielding a torus  $T_\lambda$ . When  $r_0 = 0$ , then  $\lambda = 0$ , and the torus degenerates into a circle in the plane  $(x, y)$ . When  $r_1 = 0$ , then  $\lambda \rightarrow \infty$ , and the torus degenerates into the  $z$ -axis, as shown in Figure 18.2.5. The two ends of the  $z$ -axis can be considered to meet at the point at infinity, forming a circle of infinite radius. The union of this point at infinity with  $R^3$  precisely forms  $S^3$ .

We further interpret equation (18.2.2) as describing a collection of tori  $T^2$  of varying sizes in Minkowski spacetime  $M^4$ . Transforming the coordinate system from  $(x, y, z, w)$  to  $(x, y, z, ict)$ , where  $w = ict$ , then  $R^4$  becomes  $M^4$ , and equation (18.2.2) becomes

$$x^2 + y^2 + z^2 - c^2 t^2 = 1. \quad (18.2.3)$$

Let:

$$r'^2 = x^2 + y^2 + z^2,$$

then equation (18.2.3) becomes

$$r'^2 - c^2 t^2 = 1. \quad (18.2.4)$$

Factoring, we obtain

$$(r' + ct)(r' - ct) = 1. \quad (18.2.5)$$

Let

$$R = r' + ct, \quad r = r' - ct, \quad (18.2.6)$$

then equation (18.2.5) becomes

$$Rr = 1. \quad (18.2.7)$$

Equation (18.2.8) indicates that the area of a rectangle with length  $R$  and width  $r$  equals 1.

Equation (18.2.7) can be written as

$$|R e^{i2\pi\theta}| |r e^{i2\pi\omega}| = 1, \quad (18.2.8)$$

where  $0 \leq \theta \leq 1, 0 \leq \omega \leq 1$ . Rewriting equation (18.2.8) as

$$|R \cos(2\pi\theta) + iR \sin(2\pi\theta)| |r \cos(2\pi\omega) + ir \sin(2\pi\omega)| = 1. \quad (18.2.9)$$

Let:

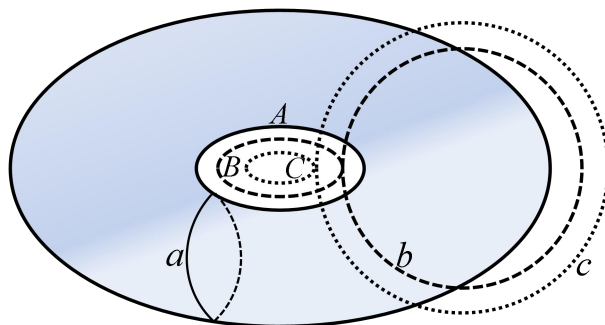
$$\tilde{x} = R \cos(2\pi\theta), \quad \tilde{y} = R \sin(2\pi\theta), \quad \tilde{z} = r \cos(2\pi\omega), \quad \tilde{w} = r \sin(2\pi\omega), \quad (18.2.10)$$

then  $R e^{i2\pi\theta}$  represents a circle  $C_R$  of radius  $R$  in the complex plane  $\tilde{x}\tilde{y}$  ( $\tilde{C}^1$ ), and  $r e^{i2\pi\omega}$  represents a circle  $C_r$  of radius  $r$  in the complex plane  $\tilde{z}\tilde{w}$  ( $\tilde{C}^1$ ). Equation (18.2.8) indicates the multiplication of the circles  $C_R$  and  $C_r$ , so equation (18.2.8) describes a two-dimensional torus  $T^2$ . When  $t$  takes different values,  $R$  and  $r$  vary accordingly, leading to tori  $T^2$  of different sizes.

According to equation (18.2.4), when time  $t$  increases,  $r'$  also increases. Therefore, as time  $t$  increases,  $R$  increases while  $r$  decreases, but  $r$  never becomes less than zero. This is because, in the plane  $(r', ct)$ , equation (18.2.4) represents a hyperbolic equation, with asymptotes  $ct = r'$ . As  $ct \rightarrow \infty$ , the point on the hyperbola in the first quadrant approaches the asymptote. Thus, as time  $t$  increases,  $r$  decreases, and when  $ct \rightarrow \infty, (r' - ct) \rightarrow 0$  while  $r \rightarrow 0$ .

When  $t = 0$ , equation (18.2.5) gives  $r'r' = 1$ . This can be interpreted as the product of two circles of radius 1, forming a torus  $T_0^2$ . As shown in Figure 18.2.6, the direct product of the

horizontal circle  $A$  and the vertical circle  $a$  yields a torus  $T_0^2$ . As previously noted, in  $M^4$ , the point-like particles of electrons, protons, and their antiparticles reside on the two-dimensional sphere  $S_0^2$ . Similarly, in  $M^4$ , the point-like particles of neutrinos and antineutrinos reside on the torus  $T_0^2$ .



**Figure 18.2.6** The generation of new tori leads to a progressively smaller hole in the center

Let  $R e^{i2\pi\theta}$  in equation (18.2.8) represent the vertical circle  $a$ , and  $r e^{i2\pi\omega}$  represent the horizontal circle  $A$ . As time  $t$  increases, the radius of the vertical circle  $a$  continuously expands, for instance, successively increasing to circles  $b$  and  $c$ , while the radius of the horizontal circle  $A$  continuously shrinks, for example, successively decreasing to circles  $B$  and  $C$ . Consequently, over time, new tori emerge around the periphery of  $T_0^2$ , with the holes at the centers of the newly formed tori becoming progressively smaller, while the thickness of the tori increases. When  $ct \rightarrow \infty$ ,  $r = (r' - ct) \rightarrow 0$ , meaning the horizontal circle  $A$  shrinks to a point, while the radius of the vertical circle  $a$  tends to infinity, becoming a straight line, as shown by the central line in Figure 18.2.5. Therefore, when  $t \rightarrow \infty$ , the hole at the center of the torus disappears. However, within a finite time after the neutrino or antineutrino is produced, the hole at the center of the torus exists.

A torus is a compact, orientable two-dimensional Riemannian manifold. Thus, among these infinitely many tori, selecting any one torus  $T^2$ , according to the Gauss-Bonnet theorem (Theorem 11.3.1), we have

$$\int_{T^2} K d\sigma = 2\pi\chi, \quad (18.2.11)$$

where  $K$  is the Gaussian curvature (sectional curvature) on  $T^2$ ,  $d\sigma$  is the oriented area element on  $T^2$ , and  $\chi$  is the Euler characteristic of  $T^2$ . Since the Euler characteristic of a torus is  $\chi = 0$ , we obtain

$$\int_{T^2} K d\sigma = 0. \quad (18.2.12)$$

Additionally, because the Gaussian curvature on the torus  $T^2$  is locally  $K = 0$ , the left-hand side of equation (18.2.12) naturally equals zero. Since the electric field intensity is proportional to the Gaussian curvature  $K$ , and the Gaussian curvature on the torus  $T^2$  is  $K = 0$ , the electric field intensity of neutrinos and antineutrinos is zero. Consequently, neutrinos and antineutrinos carry no electric charge.

We previously interpreted Equation (18.2.3) as describing a set

$$B^3 = \{(x, y, z) \in R^3 \mid 1 \leq x^2 + y^2 + z^2 \leq \sqrt{1 + c^2 t^2} (t > 0)\},$$

composed of infinitely many spheres  $S^2$  centered at the origin of  $R^3$  with radii ranging from 1 to  $r = \sqrt{1 + c^2 t^2} (t > 0)$ . In this set  $B^3$ , different Riemannian metrics can be chosen—either a metric with sectional curvature  $K > 0$  or a metric with sectional curvature  $K = 0$ , see Example 11.9.1. For electrons and protons, we take the metric with  $K > 0$ . For neutrinos, we take the



metric with  $K = 0$ , and thus by taking a sphere  $S^2$  surrounding the origin of  $R^3$ , we obtain

$$\int_{S^2} K d\sigma = 0.$$

### 3.Free Neutrinos and Antineutrinos Move at the Speed of Light

We write equation (18.2.8) in the form of a mapping:

$$f(\theta, \omega) = (R e^{i2\pi\theta}, r e^{i2\pi\omega}), \quad (18.2.13)$$

where  $Rr=1$ . When  $0 \leq \theta \leq 1$  and  $0 \leq \omega \leq 1$ , this mapping  $f(\theta, \omega)$  transforms a unit square in the plane  $\theta\omega$  (as shown in Figure 18.2.7) into a two-dimensional torus  $T^2$  generated by the product of two circles with radii  $R$  and  $r$ . As illustrated in Figure 18.2.8, this mapping process can be decomposed into two steps: First, a unit square in the plane  $\theta\omega$  is mapped to a rectangle with length  $2\pi R$  and width  $2\pi r$ ; second, this rectangle is mapped to a two-dimensional torus  $T^2$ .

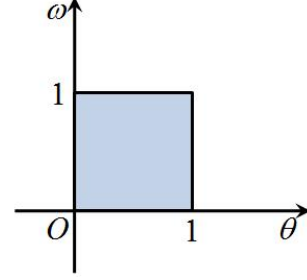


Figure 18.2.7 Unit square

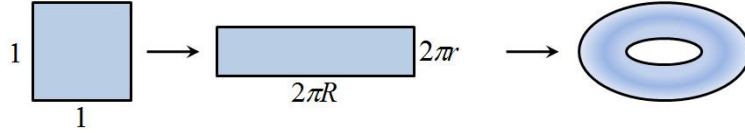


Figure 18.2.8 A square is mapped into a two-dimensional torus in two steps

Assume  $0 \leq \theta \leq 1$  and  $0 \leq \omega \leq 1$ . If  $\theta$  takes a fixed value, denote  $f(\theta, \omega)$  as  $f_\theta(\omega) = f(\theta, \omega)$ . As shown in Figures 18.2.7 and 18.2.8, when  $\omega=0$  and  $\omega=1$ ,  $r e^{i2\pi\omega} = r e^{i2\pi\omega}$ , and  $f_\theta(\omega)$  maps  $(\theta, 0)$  and  $(\theta, 1)$  to the same point, while  $\omega$  takes other values,  $f_\theta(\omega)$  is injective. Thus,  $f_\theta(\omega)$  glues the boundary points of the top and bottom edges of the rectangle together, forming a cylinder. Similarly, assume  $\omega$  takes a fixed value, and denote  $f(\theta, \omega)$  as  $f_\omega(\theta) = f(\theta, \omega)$ . When  $\theta=0$  and  $\theta=1$ ,  $R e^{i2\pi\theta} = R e^{i2\pi\theta}$ , and  $f_\omega(\theta)$  maps  $(0, \omega)$  and  $(1, \omega)$  to the same point, while for other values of  $\theta$ ,  $f_\omega(\theta)$  is injective. Thus,  $f_\omega(\theta)$  glues the boundary points of the left and right edges of the rectangle together, ultimately forming the rectangle into a two-dimensional torus  $T^2$ .

As shown in Figure 18.2.9, if  $0 \leq \theta \leq 1$  and  $0 \leq \omega \leq 1$ , then  $f(\theta, \omega)$  maps square  $A$  to the torus  $T^2$ . If  $1 \leq \theta \leq 2$  and  $2 \leq \omega \leq 3$  again, then  $f(\theta, \omega)$  also maps square  $B$  to the torus  $T^2$ . Both squares  $A$  and  $B$  are mapped to the same torus  $T^2$ . In fact, every square in the figure is mapped by  $f(\theta, \omega)$  to the same torus  $T^2$ .

Points  $(0, \omega), (1, \omega), (2, \omega), \dots, (k, \omega), \dots$ , in the plane  $\theta\omega$  (where  $k$  is an integer) are all mapped by  $f(\theta, \omega)$  to the same point. Similarly, points  $(\theta, 0), (\theta, 1), (\theta, 2), \dots, (\theta, n), \dots$ , (where  $n$  is an integer) are also mapped by  $f(\theta, \omega)$  to the same point. This is the condition under which all squares in the plane  $\theta\omega$  are mapped to the same torus  $T^2$ . Therefore, we define the following equivalence relation  $\sim$  on the plane  $\theta\omega$ :

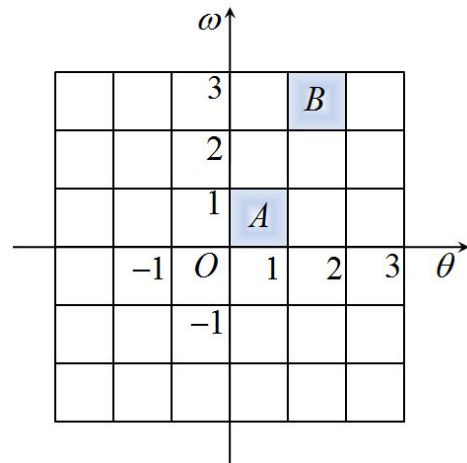


Figure 18.2.9 Every square is mapped to the same torus

Let  $(\theta^1, \omega^1) \in R^2, (\theta^2, \omega^2) \in R^2$ . Then  $(\theta^1, \omega^1) \sim (\theta^2, \omega^2)$  if and only if there exist integers  $k$  and  $n$  such that

$$\theta^2 = \theta^1 + k, \quad \omega^2 = \omega^1 + n. \quad (18.2.14)$$

After defining the equivalence relation  $\sim$  on the plane  $\theta\omega$  (i.e.,  $R^2$ ),  $R^2$  becomes a torus  $T^2$ , denoted as  $T^2 = R^2 / \sim$ .

According to equation (18.2.10),  $\theta^1$  is equivalent to  $\theta^2$  if and only if  $\tilde{x}^1 = R \cos(2\pi\theta^1 + 2k\pi)$  is equivalent to  $\tilde{x}^2 = R \cos(2\pi\theta^2)$  and  $\tilde{y}^1 = R \sin(2\pi\theta^1 + 2k\pi)$  is equivalent to  $\tilde{y}^2 = R \sin(2\pi\theta^2)$ . Similarly,  $\omega^1$  is equivalent to  $\omega^2$  if and only if  $\tilde{z}^1 = r \cos(2\pi\omega^1 + 2n\pi)$  is equivalent to  $\tilde{z}^2 = r \cos(2\pi\omega^2)$  and  $\tilde{w}^1 = r \sin(2\pi\omega^1 + 2n\pi)$  is equivalent to  $\tilde{w}^2 = r \sin(2\pi\omega^2)$ .

Here,  $\tilde{x}^1$  and  $\tilde{y}^1$  represent the circle  $C_{R1}$ , while  $\tilde{z}^1$  and  $\tilde{w}^1$  represent the circle  $C_{r1}$ . According to equation (18.2.9), the product of these two circles  $C_{R1}$  and  $C_{r1}$  forms a torus  $T_1^2$ . Similarly,  $\tilde{x}^2$  and  $\tilde{y}^2$  represent the circle  $C_{R2}$ , and  $\tilde{z}^2$  and  $\tilde{w}^2$  represent the circle  $C_{r2}$ . The product of these two circles  $C_{R2}$  and  $C_{r2}$  also forms a torus  $T_2^2$ . Since  $\tilde{x}^1$  is equivalent to  $\tilde{x}^2$  and  $\tilde{y}^1$  is equivalent to  $\tilde{y}^2$ , the circle  $C_{R1}$  is equivalent to  $C_{R2}$ . Likewise, since  $\tilde{z}^1$  is equivalent to  $\tilde{z}^2$  and  $\tilde{w}^1$  is equivalent to  $\tilde{w}^2$ , the circle  $C_{r1}$  is equivalent to  $C_{r2}$ . Therefore, the torus  $T_1^2$  is the same as the torus  $T_2^2$ .

When  $R$  and  $r$  in equation (18.2.9) take different values, tori of varying sizes are obtained.

If  $\tilde{x}^1$  is equivalent to  $\tilde{x}^2$ , then

$$\Delta\tilde{x} = \tilde{x}^2 - \tilde{x}^1 = 0. \quad (18.2.15)$$

If  $\tilde{y}^1$  is equivalent to  $\tilde{y}^2$ , then

$$\Delta\tilde{y} = \tilde{y}^2 - \tilde{y}^1 = 0. \quad (18.2.16)$$

If  $\tilde{z}^1$  is equivalent to  $\tilde{z}^2$ , then

$$\Delta\tilde{z} = \tilde{z}^2 - \tilde{z}^1 = 0. \quad (18.2.17)$$

If  $\tilde{w}^1$  is equivalent to  $\tilde{w}^2$ , then

$$\Delta\tilde{w} = \tilde{w}^2 - \tilde{w}^1 = 0. \quad (18.2.18)$$

If equation (18.2.8) is modified to

$$|R e^{i4\pi\theta}| |r e^{i4\pi\omega}| = 1,$$

then equation (18.2.13) should be revised to

$$f(\theta, \omega) = (R e^{i4\pi\theta}, r e^{i4\pi\omega}).$$

$f(\theta, \omega)$  maps a square with side length  $\frac{1}{2}$  to the torus  $T^2$ . Therefore, equation (18.2.14) should be rewritten as

$$\theta^2 = \theta^1 + \frac{k}{2}, \quad \omega^2 = \omega^1 + \frac{n}{2},$$

which shortens the distance between two equivalent points in the plane  $\theta\omega$ .

Generally, if equation (18.2.8) is modified to

$$|R e^{i2a\pi\theta}| |r e^{i2a\pi\omega}| = 1, \quad (18.2.19)$$

where  $a$  is a positive real number, then equation (18.2.13) should be revised to

$$f(\theta, \omega) = (R e^{i2a\pi\theta}, r e^{i2a\pi\omega}). \quad (18.2.20)$$

$f(\theta, \omega)$  maps a square with side length  $\frac{1}{a}$  to the torus  $T^2$ . Therefore, equation (18.2.14) should be rewritten as



$$\theta^2 = \theta^1 + \frac{k}{a}, \quad \omega^2 = \omega^1 + \frac{n}{a}. \quad (18.2.21)$$

If  $a$  is sufficiently large, the two equivalent points in the plane  $\theta\omega$  become very close:

$$\Delta\tilde{x} \rightarrow d\tilde{x}, \quad \Delta\tilde{y} \rightarrow d\tilde{y}, \quad \Delta\tilde{z} \rightarrow d\tilde{z}, \quad \Delta\tilde{w} \rightarrow d\tilde{w},$$

so  $f(\theta, \omega)$  maps infinitesimally small squares in  $\theta\omega$  into tori  $T^2$ . From equations (18.2.15) to (18.2.18), we obtain

$$d\tilde{x} = d\tilde{y} = d\tilde{z} = d\tilde{w} = 0,$$

and thus,

$$d\tilde{s}^2 = d\tilde{x}^2 + d\tilde{y}^2 + d\tilde{z}^2 + d\tilde{w}^2 = 0.$$

Let,

$$d\tilde{s} = ds, \quad \tilde{x} = x, \quad \tilde{y} = y, \quad \tilde{z} = z, \quad \tilde{w} = ict,$$

then in  $M^4$ ,

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 = 0, \quad (18.2.22)$$

alternatively,

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 = 0. \quad (18.2.23)$$

From equation (18.2.22) or (18.2.23), we derive  $ds^2 = 0$ , which resembles the metric form of photons in  $M^4$ . This indicates that free neutrinos and antineutrinos also move at the speed of light  $c$  in  $M^4$ .

In summary: By defining the equivalence relation  $\sim$  in the plane  $\theta\omega$  and introducing equation (18.2.10) in  $M^4$ , equation (18.2.2) can be interpreted as describing tori  $T^2$  of varying sizes in  $M^4$ , and equations (18.2.22) or (18.2.23) hold. Therefore, the interior of neutrinos and antineutrinos is composed of tori  $T^2$ , resulting in no electric charge. Additionally, free neutrinos and antineutrinos, like photons, travel at the speed of light  $c$ .

Denote the equivalence class defined by equation (18.2.21) on  $(\theta, \omega) \in R^2$  as  $[\theta, \omega]$ . Define a mapping  $f: T^2 \rightarrow C^2$  such that

$$f([\theta, \omega]) = (R e^{i2a\pi\theta}, r e^{i2b\pi\omega}), \quad \forall [\theta, \omega] \in T^2, \quad (18.2.24)$$

where  $Rr=1$ .  $C^2$  is a two-dimensional complex vector space, four-dimensional in real terms. Like  $R^4$ , it is flat, and  $f$  is an isometric mapping. Thus, the torus  $T^2$  as a Riemannian manifold is flat.

#### 4. Neutrinos and Antineutrinos that Appear and Disappear Intermittently

To generalize our discussion, we replace the second  $a$  in equation (18.2.19) with  $b$ , yielding

$$|R e^{i2a\pi\theta}| |r e^{i2b\pi\omega}| = 1, \quad (18.2.25)$$

where  $a$  and  $b$  are positive real numbers. Equation (18.2.25) maps a rectangle of length  $\frac{1}{a}$  and width  $\frac{1}{b}$  in the plane  $\theta\omega$  to a torus  $T^2$ .

Assuming neutrinos and antineutrinos are in free motion, multiply equation (18.2.25) by  $e^{-\frac{i}{\hbar}Et}$  to obtain

$$|R e^{i2a\pi\theta} e^{-\frac{i}{\hbar}Et}| |r e^{i2b\pi\omega} e^{-\frac{i}{\hbar}Et}| = 1, \quad (18.2.26)$$

that is

$$|R e^{i2a\pi\theta - \frac{i}{\hbar}Et}| |r e^{i2b\pi\omega - \frac{i}{\hbar}Et}| = 1. \quad (18.2.27)$$

Let  $\theta = \theta^1 + \frac{k}{a}$ ,  $\omega = \omega^1 + \frac{n}{b}$ , then

$$| R e^{i2a\pi(\theta^1 + \frac{k}{a}) - \frac{i}{h}Et} | | r e^{i2b\pi(\omega^1 + \frac{n}{b}) - \frac{i}{h}Et} | = 1,$$

$$| R e^{i2a\pi(\theta^1 + \frac{k}{a} - \frac{1}{2a\pi\hbar}Et)} | | r e^{i2b\pi(\omega^1 + \frac{n}{b} - \frac{1}{2b\pi\hbar}Et)} | = 1.$$

Now let

$$\theta^2 = \theta^1 + \frac{k}{a} - \frac{1}{2a\pi\hbar}Et, \quad \omega^2 = \omega^1 + \frac{n}{b} - \frac{1}{2b\pi\hbar}Et. \quad (18.2.28)$$

Substituting into the equation gives

$$| R e^{i2a\pi\theta^2} | | r e^{i2b\pi\omega^2} | = 1.$$

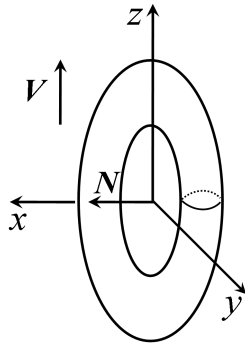
For the point  $(\theta^1, \omega)$  to be identified with  $(\theta^2, \omega)$  and the point  $(\theta, \omega^1)$  to be identified with  $(\theta, \omega^2)$ ,  $\frac{1}{2\pi\hbar}Et$  must be an integer. If  $\frac{1}{2\pi\hbar}Et$  is not an integer, equation (18.2.27) cannot map a rectangle of length  $\frac{1}{a}$  and width  $\frac{1}{b}$  in  $\theta\omega$  to a torus  $T^2$ , meaning neutrinos or antineutrinos would not be produced or would vanish. Since the probability of  $\frac{1}{2\pi\hbar}Et$  being an integer is very low, during flight, neutrinos and antineutrinos would intermittently become part of Minkowski spacetime  $M^4$  for extended periods, rendering them undetectable. They only become visible when  $\frac{1}{2\pi\hbar}Et$  is an integer. This makes it difficult for neutrinos and antineutrinos to collide with other particles or be captured by experimental equipment. However, when neutrinos and antineutrinos are not in free motion,  $\frac{1}{2\pi\hbar}Et = 0$ , which is naturally an integer, and they do not remain hidden.

### 5. Three States of Neutrinos

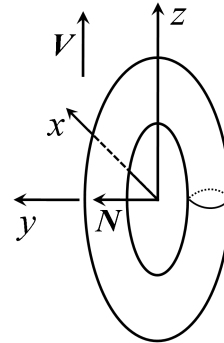
First, it is important to understand that, based on the analysis above, the hole in the torus exists for a finite time after the neutrino is produced. Only when time  $t \rightarrow \infty$  does the hole in the torus disappear.

We take the propagation direction  $V$  of the neutrino as the  $z$ -axis direction in the spatial coordinate system  $(x, y, z)$ . Depending on the relationship between the normal vector direction  $N$  of the hole in the torus  $T^2$  and the propagation direction  $V$ , neutrinos (or antineutrinos) can be classified into three fundamental states.

#### (1) The First Fundamental State



**Figure 18.2.10** The first elementary state  $v_1$



**Figure 18.2.11** The first elementary state  $v_2$

As shown in Figure 18.2.10, if the normal vector direction  $N$  of the hole in the torus  $T^2$  aligns with the  $x$ -axis direction of the spatial coordinate system  $(x, y, z)$  and is perpendicular to the propagation direction  $V$  of the neutrino, then the neutrino is said to be in the first fundamental state, denoted as  $v_1$ .

### (2)The Second Fundamental State

As shown in Figure 18.2.11, if the normal vector direction  $N$  of the hole in the torus  $T^2$  aligns with the  $y$ -axis direction of the spatial coordinate system and is perpendicular to the propagation direction  $V$  of the neutrino, then the neutrino is said to be in the second fundamental state, denoted as  $\nu_2$ .

### (3)The Third Fundamental State

As shown in Figure 18.2.12, if the normal vector direction  $N$  of the hole in the torus  $T^2$  aligns with the  $z$ -axis direction of the spatial coordinate system  $(x, y, z)$  and is the same as the propagation direction  $V$  of the neutrino, then the neutrino is said to be in the third fundamental state, denoted as  $\nu_3$ .

There is no significant difference between the first and second fundamental states, and they can sometimes be regarded as the same state.

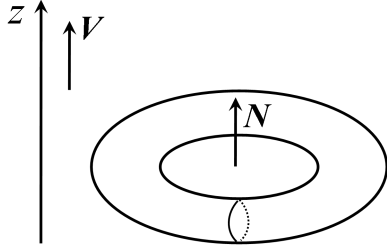


Figure 18.2.12 The first elementary state  $\nu_1$

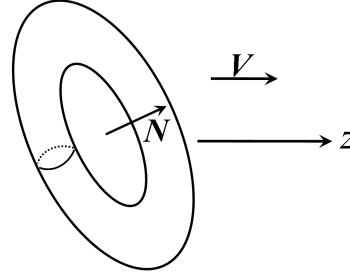


Figure 18.2.13 Mixed state  $\nu$

### (4)Mixed State

A mixed state is a random combination of the three fundamental states mentioned above, expressed as

$$\nu = a_1\nu_1 + a_2\nu_2 + a_3\nu_3, \quad (18.2.29)$$

where the squared moduli of the coefficients  $|a_1|^2$ ,  $|a_2|^2$ , and  $|a_3|^2$  represent the probabilities of the states  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$  occurring, respectively. The mixed state can be regarded as a configuration where the normal vector  $N$  of the hole in the torus is obliquely oriented relative to the propagation direction  $V$  of the neutrino, as shown in Figure 18.2.13.

If the first and second states are considered the same, there are only three states. These three states of neutrinos are analogous to the three polarization states of light waves, as illustrated in Figure 18.2.14.

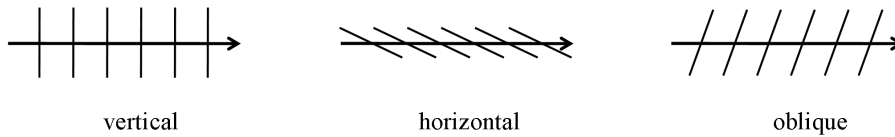


Figure 18.2.14 The three polarization states of light waves

## 6.Types of Neutrinos and Antineutrinos

Extensive experimental observations have revealed that neutrinos and antineutrinos can be categorized into three types: neutrinos  $\nu_e$ , antineutrinos  $\bar{\nu}_e$ ; neutrinos  $\nu_\mu$ , antineutrinos  $\bar{\nu}_\mu$ ; and neutrinos  $\nu_\tau$ , antineutrinos  $\bar{\nu}_\tau$ . We now explore how to distinguish among these three types of neutrinos and antineutrinos, using neutrinos as an example (the following discussion applies equally to the corresponding antineutrinos). For clarity, we begin with the tau neutrino  $\nu_\tau$ .

### (1)Tau Neutrino $\nu_\tau$

In the first and second states  $\nu_1$  and  $\nu_2$  shown in Figures 18.2.10 and 18.2.11, the normal vector direction  $N$  of the hole in the torus  $T^2$  aligns with the  $x$ -axis or  $y$ -axis of the spatial coordinate system  $(x, y, z)$  and is perpendicular to the propagation direction  $V$  of the neutrino. This

state corresponds to the tau neutrino  $\nu_\tau$ . When neutrinos in this state travel toward a detector, they present the side of the torus to the detector, resulting in a very low probability of detection. Experimentally, the detection of  $\nu_\tau$  appears to be the most challenging; in fact,  $\nu_\tau$  was the last neutrino to be discovered (in 2000). Based on these considerations, we identify  $\nu_\tau$  as belonging to either the  $\nu_1$  or  $\nu_2$  state.

Assuming the neutrino is in free motion, we multiply equation (18.2.8) by  $e^{-\frac{i}{\hbar}Et}$ , yielding

$$\begin{aligned} |R e^{i2\pi\theta} e^{-\frac{i}{\hbar}Et}| |r e^{i2\pi\omega} e^{-\frac{i}{\hbar}Et}| &= 1, \\ |R e^{i(2\pi\theta - \frac{1}{\hbar}Et)}| |r e^{i(2\pi\omega - \frac{1}{\hbar}Et)}| &= 1. \end{aligned}$$

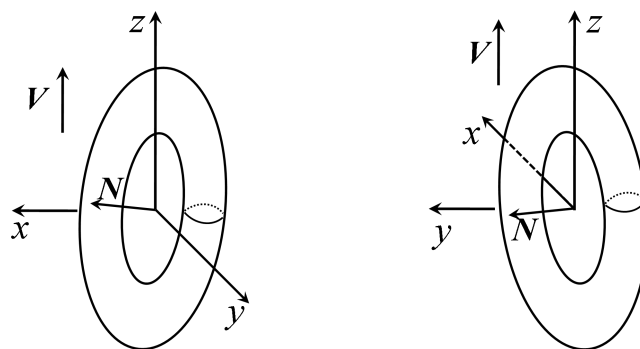
The real part of  $R e^{i(2\pi\theta - \frac{1}{\hbar}Et)}$  is

$$R \cos(2\pi\theta - \frac{1}{\hbar}Et) = R \cos(\frac{1}{\hbar}Et - 2\pi\theta), \quad (18.2.30)$$

and the real part of  $r e^{i(2\pi\omega - \frac{1}{\hbar}Et)}$  is

$$r \cos(2\pi\omega - \frac{1}{\hbar}Et) = r \cos(\frac{1}{\hbar}Et - 2\pi\omega). \quad (18.2.31)$$

From equations (18.2.30) and (18.2.31), it can be observed that for a fixed time  $t$ , a higher energy  $E$  corresponds to a faster rate at which the circle is traced, and consequently, the torus is formed more rapidly. Therefore, as the energy  $E$  of the neutrino increases, the hole in the center of the torus shrinks more quickly, while the thickness of the torus also increases. This results in a higher probability of detecting the neutrino. Hence, one method to enhance the detection probability of such neutrinos is to increase their energy. Experimentally, neutrinos  $\nu_\tau$  are detected when their energy is very high. Based on these considerations, we conclude that neutrinos  $\nu_\tau$  belong to either the  $\nu_1$  or  $\nu_2$  state.



**Figure 18.2.15** Vertical or slightly tilted tori correspond to neutrinos  $\nu_\tau$

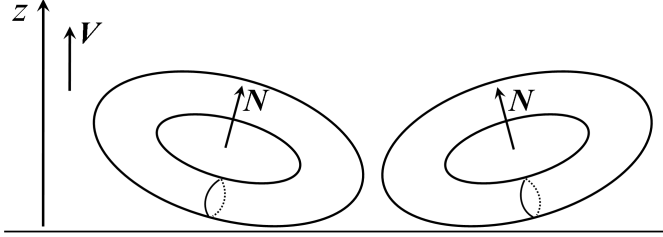
Certainly, when the torus shown in Figure 18.2.10 or Figure 18.2.11 is slightly tilted—provided that the central hole is not visible from a top-down view, as illustrated in Figure 18.2.15—the probability of detecting such neutrinos changes little. Therefore, we also classify this scenario as belonging to neutrinos  $\nu_\tau$ . In this state, the neutrino  $\nu_\tau$  is a linear combination of the three basic states  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ , i.e.,

$$\nu_\tau = U_{\tau 1}\nu_1 + U_{\tau 2}\nu_2 + U_{\tau 3}\nu_3, \quad (18.2.32)$$

where the squared magnitudes of the coefficients  $|U_{\tau 1}|^2$ ,  $|U_{\tau 2}|^2$ , and  $|U_{\tau 3}|^2$  represent the probabilities of observing states  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ , respectively. In this state,  $|U_{\tau 1}|^2$  and  $|U_{\tau 2}|^2$  are relatively large, while  $|U_{\tau 3}|^2$  is relatively small.

## (2) Neutrino $\nu_\mu$

As shown in Figure 18.2.12, the normal vector direction  $N$  of the hole in the torus  $T^2$  aligns with the propagation direction  $V$  of the neutrino. This state corresponds to the neutrino  $\nu_\mu$ . When such a neutrino travels toward a detector, even though its central hole faces the detector, it lies flat and presents a larger effective area toward the detector. Consequently, the probability of detecting this neutrino is higher than that for  $\nu_\tau$ , placing it at a moderate level. Among the three neutrinos, the experimental difficulty in detecting  $\nu_\mu$  also appears to be moderate. Indeed,  $\nu_\mu$  was neither the earliest nor the latest to be discovered (it was detected in 1962). Based on these considerations, we identify  $\nu_\mu$  with this state  $\nu_3$ .



**Figure 18.2.16** A slightly tilted, flat torus also belongs to the neutrino  $\nu_\mu$

Certainly, when the torus depicted in Figure 18.2.12 is slightly tilted, as illustrated in Figure 18.2.16, the probability of detecting such a neutrino changes little. Therefore, we also classify this scenario as belonging to neutrino  $\nu_\mu$ . In this state, the neutrino  $\nu_\mu$  is a linear combination of the three basic states  $\nu_1, \nu_2$ , and  $\nu_3$ , i.e.,

$$\nu_\mu = U_{\mu 1} \nu_1 + U_{\mu 2} \nu_2 + U_{\mu 3} \nu_3, \quad (18.2.33)$$

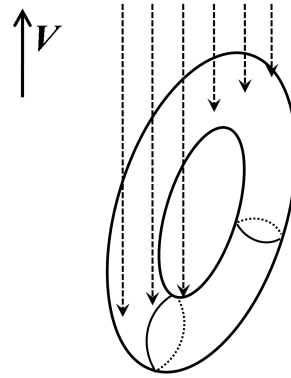
where the squared magnitudes of the coefficients  $|U_{\mu 1}|^2$ ,  $|U_{\mu 2}|^2$ , and  $|U_{\mu 3}|^2$  represent the probabilities of observing states  $\nu_1, \nu_2$ , and  $\nu_3$ , respectively. In this state,  $|U_{\mu 1}|^2$  and  $|U_{\mu 2}|^2$  are relatively small, while  $|U_{\mu 3}|^2$  is relatively large.

### (3) Neutrino $\nu_e$

If the torus in Figure 18.2.15 is tilted at a suitably larger angle such that, when viewed from above, the central hole is not visible, and such that neutrinos in this state present a very large—or even the maximum—torus area toward the detector as they travel toward it, then the probability of detecting such neutrinos becomes very high. As shown in Figure 18.2.17, the dashed arrow indicates the line of sight. When looking from above along this line of sight, the central hole of the torus is not visible, but the torus presents its maximum area toward the line of sight. The neutrino's direction of motion is opposite to the line of sight. Among the three neutrinos, the experimental difficulty in detecting  $\nu_e$  appears to be the lowest; indeed,  $\nu_e$  was the earliest to be discovered (in 1956).

Additionally, even if neutrinos in this state have low energy, detectors can still capture them with high probability. In fact, most experimentally detected  $\nu_e$  neutrinos also have low energy. Based on these considerations, we conclude that neutrino  $\nu_e$  belongs to this state. In this state, the neutrino  $\nu_e$  is a linear combination of the three basic states  $\nu_1, \nu_2$ , and  $\nu_3$ , i.e.,

$$\nu_e = U_{e 1} \nu_1 + U_{e 2} \nu_2 + U_{e 3} \nu_3, \quad (18.2.34)$$



**Figure 18.2.17** Neutrino  $\nu_e$  presents its maximum area toward the detector, but its central hole does not face the detector

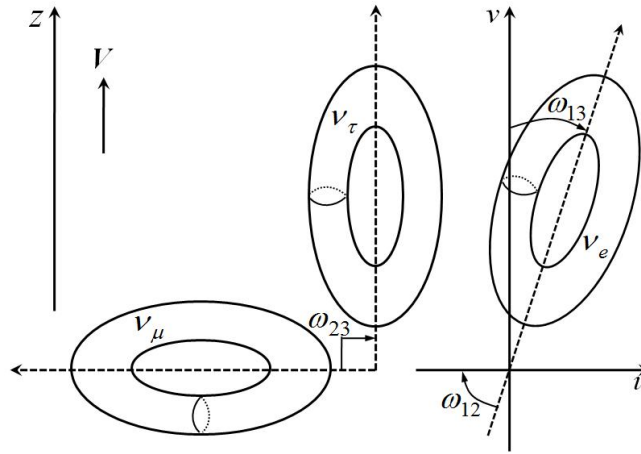
where the squared magnitudes of the coefficients  $|U_{e1}|^2$ ,  $|U_{e2}|^2$ , and  $|U_{e3}|^2$  represent the probabilities of observing states  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ , respectively. In this state,  $|U_{e1}|^2$  and  $|U_{e2}|^2$  are relatively large, while  $|U_{e3}|^2$  is relatively small.

Combining equations (18.2.32) through (18.2.34), we express them in matrix form as

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (18.2.35)$$

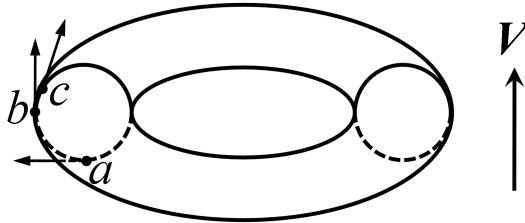
## 7.Oscillations of Neutrinos and Antineutrinos

These three neutrinos  $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$  can interconvert, a phenomenon often referred to as neutrino oscillations. As shown in Figure 18.2.18, when neutrino  $\nu_\mu$  rotates by  $\omega_{23} = 90^\circ$ , it transforms into neutrino  $\nu_\tau$ ; when neutrino  $\nu_e$  rotates by  $\omega_{12}$ , it transforms into neutrino  $\nu_\mu$ , or when it rotates by  $\omega_{13}$ , it transforms into neutrino  $\nu_\tau$ .



**Figure 18.2.18** Interconversion among the three neutrinos

These three neutrinos can interconvert because the neutrino itself is a Lie group  $SU(2)$ . Each group element of the Lie group generates a left or right translation. A left (or right) translation is a smooth homeomorphic transformation acting on the Lie group itself. The homeomorphic transformations of the Lie group inevitably cause the torus to rotate. We do not need to assume that neutrinos have rest mass to explain the cause of neutrino oscillations.



**Figure 18.2.19** Left translation of point elements on the meridian leads to interconversion among the three neutrinos

The group elements  $g \in SU(2)$  generate left or right translations. The left translation  $L_g : SU(2) \rightarrow SU(2)$  induced by group element  $g$  is a diffeomorphic transformation, which induces a tangent mapping  $L_{g*}$  on the  $SU(2)$  group. As shown in Figure 18.2.19, the vertical circle on the torus is called the meridian, and the horizontal circle is called the latitude. On the meridian to the left of the torus, point  $a$  is left-translated to point  $b$ . The tangent mapping induced

by the left translation then maps the tangent vector at point  $a$  to the tangent vector at point  $b$ . This corresponds to the transformation of neutrino  $\nu_\mu$  into neutrino  $\nu_\tau$  as shown in Figure 18.2.18. Similarly, if another left translation moves point  $b$  to point  $c$ , the tangent mapping induced by this left translation maps the tangent vector at point  $b$  to the tangent vector at point  $c$ . This corresponds to the transformation of neutrino  $\nu_\tau$  into neutrino  $\nu_e$  as shown in Figure 18.2.18.

Since the action of group elements is random, the rotation of the torus shown in Figure 18.2.18 is also random. This means the interconversion among the three neutrinos is stochastic. One must not assume, as depicted in Figure 18.2.18, that neutrino  $\nu_\mu$  will continuously transform into neutrino  $\nu_\tau$  and then into neutrino  $\nu_e$  in a clockwise sequence.

We have represented the  $SU(2)$  group of neutrinos using equation (18.2.1), but for convenience, the elements of the neutrino's  $SU(2)$  group are simply expressed as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where  $a, b, c, d$  are complex numbers. On the local region of the torus shown in Figure 18.2.19, we establish a local coordinate system  $(U; u, v)$ . Suppose the coordinates  $(u, v)$  of a point on the torus transform to  $(u', v')$  under the action of the group element. Then we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u' \\ v' \end{pmatrix}. \quad (18.2.36)$$

Thus, the group elements of the  $SU(2)$  group representing neutrinos can generate a linear transformation as shown in equation (18.2.36). Equation (18.2.36) can describe the process in Figure 18.2.19 where point  $a$  is left-translated to point  $b$  and point  $b$  is left-translated to point  $c$ .

We regard the plane shown in Figure 18.2.18 as a two-dimensional complex plane  $(u, v)$ , using the vector  $(u, v)$  to represent the dashed vector in Figure 18.2.18. Considering the  $SU(2)$  group as a linear transformation group on the two-dimensional complex plane  $(u, v)$ , equation (18.2.36) can also be used to describe the interconversion among the three neutrinos illustrated in Figure 18.2.18.

In experimental data analysis, the three-dimensional rotation group is employed to describe such transformations of neutrinos. Current literature on neutrinos assumes the transformation matrix to be

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{23} & \sin \theta_{23} \\ 0 & -\sin \theta_{23} & \cos \theta_{23} \end{pmatrix} \begin{pmatrix} e^{-i\delta/2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{i\delta/2} \end{pmatrix} \begin{pmatrix} \cos \theta_{13} & 0 & \sin \theta_{13} \\ 0 & 1 & 0 \\ -\sin \theta_{13} & 0 & \cos \theta_{13} \end{pmatrix} \\ \times \begin{pmatrix} e^{i\delta/2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-i\delta/2} \end{pmatrix} \begin{pmatrix} \cos \theta_{12} & \sin \theta_{12} & 0 \\ -\sin \theta_{12} & \cos \theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where the angle  $\theta_{12}$  describes the mixing angle between neutrinos  $\nu_e$  and  $\nu_\mu$  (and/or  $\nu_\tau$ ), the angle  $\theta_{13}$  describes the mixing angle between neutrinos  $\nu_e$  and  $\nu_\mu$  (and/or  $\nu_\tau$ ), the angle  $\theta_{23}$  describes the mixing angle between neutrinos  $\nu_\mu$  and  $\nu_\tau$ , and  $\delta$  is a phase angle.

Based on the theoretical analysis above, we can reinterpret the angle  $\theta_{12}$  as describing the rotation required to convert neutrino  $\nu_e$  into neutrino  $\nu_\mu$ , the angle  $\theta_{13}$  as describing the rotation required to convert neutrino  $\nu_e$  into neutrino  $\nu_\tau$ , and the angle  $\theta_{23}$  as describing the rotation required to convert neutrino  $\nu_\mu$  into neutrino  $\nu_\tau$ . This is our geometric interpretation of the mixing angles. Since a rotation by an angle in the  $SU(2)$  group corresponds to a rotation by twice that angle in the  $SO(3)$  group, comparing with Figure 18.2.18 yields:

$$\theta_{23} = \frac{1}{2}\omega_{23} = 45^\circ, \quad \theta_{13} = \frac{1}{2}\omega_{13}, \quad \theta_{12} = \frac{1}{2}\omega_{12},$$

$$\theta_{12} + \theta_{13} = \frac{1}{2}(\omega_{12} + \omega_{13}) = \frac{1}{2} \times 90^\circ = 45^\circ .$$

Experimental measurements give:

$$\theta_{23} = 45^\circ \pm 8^\circ , \quad \theta_{12} = 34^\circ \pm 2^\circ , \quad \theta_{13} = 8.8^\circ \pm 0.8^\circ ,$$

$$\theta_{12} + \theta_{13} \approx 34^\circ + 8.8^\circ = 42.8^\circ .$$

The experimental value of  $\theta_{23}$  fluctuates around  $45^\circ$  with relatively large fluctuations. The reason for this is that, as shown in Figure 18.2.16, a slightly tilted flat torus can still be regarded as neutrino  $\nu_\mu$ , and as shown in Figure 18.2.15, a slightly tilted vertical torus can also be classified as neutrino  $\nu_\tau$ . Similar explanations can be applied to the measurement uncertainties of angles  $\theta_{12}$  and  $\theta_{13}$ .

## §18.3 The Reason Why Photons Carry No Charge

### 1. First Explanation

Each photon is an  $SO(3)$  group and also a real projective space  $RP^3$ . Since there exists a homomorphism  $f$  from  $SU(2)$  to  $SO(3)$ , the kernel of  $f$  is  $N(E, -E)$ , where  $E$  is the identity element of  $SU(2)$ . That is, for each element in  $SU(2)$ ,  $\forall g \in SU(2)$ , there exist two elements  $(Eg, -Eg)$  in  $SU(2)$  corresponding to the same element in  $SO(3)$ . That is, for each element in  $SO(3)$ , there are two corresponding elements in  $SU(2)$ : one is  $Eg$  and the other is  $-Eg$ . These two elements are antipodal points in  $SU(2)$  (or  $S^3$ ), one positive and one negative. Therefore, for the  $SO(3)$  group, we can add equations (18.1.12) and (18.1.18) to obtain

$$\oint_{S^2} \mathbf{E} \cdot d\boldsymbol{\sigma} + \oint_{S^2} \mathbf{E} \cdot d\boldsymbol{\sigma} = \frac{Q}{\epsilon_0} - \frac{Q}{\epsilon_0} = 0 ,$$

that is,

$$\oint_{S^2} \mathbf{E} \cdot d\boldsymbol{\sigma} = 0 .$$

This equation holds for any closed surface  $M$  as well, so we write it as

$$\oint_M \mathbf{E} \cdot d\boldsymbol{\sigma} = 0 . \quad (18.3.1)$$

Although the right-hand side of equation (18.3.1) is zero, we cannot conclude that  $\mathbf{E}$  on the left-hand side is zero, because  $\mathbf{E}$  is not constant. Therefore, because a photon is an  $SO(3)$  group, it carries no net charge but still possesses an electric field.



## 2.Second Explanation

Equation (18.3.1) can also be derived through another approach. We know that a photon is a three-dimensional real projective space  $RP^3$ .  $RP^3$  has several equivalent models, one of which views it as a smooth manifold in four-dimensional Euclidean space  $R^4$  consisting of infinitely many lines passing through the origin. However, the origin of  $R^4$  must be removed, and all points on the same line are identified as a single point. As shown in Figure 18.3.1, points  $p_1$  and  $p_2$  represent the same point, points  $p_3$  and  $p_4$  also represent the same point, and so on. Our physical space is the three-dimensional Euclidean space  $R^3$ , which is a subspace of  $R^4$ . These lines

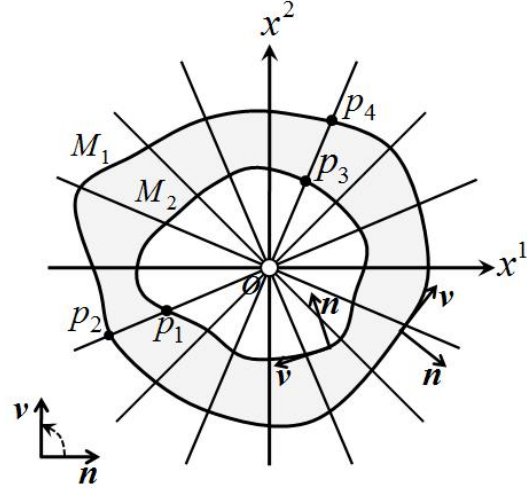


Figure 18.3.1 Integration surfaces in  $RP^3$

passing through the origin appear as straight lines to us living in  $R^3$ , but for  $RP^3$ , they are considered as points. Therefore, for us living in  $R^3$ , the surfaces  $M_1$  and  $M_2$  enclosing the origin appear as distinct surfaces. However, for  $RP^3$ , they are considered the same surface. Both surfaces  $M_1$  and  $M_2$  are closed, compact, and orientable. According to the Gauss-Bonnet theorem, for surface  $M_1$ ,

$$\oint_{M_1} K_1 d\sigma_1 = 2\pi\chi(M_1), \quad (18.3.2)$$

where  $K_1$  is the Gaussian curvature on  $M_1$ ,  $d\sigma_1$  is the oriented area element on  $M_1$ , and  $\chi$  is the Euler characteristic of  $M_1$ , with  $\chi = 2$ . The outward normal vector  $\mathbf{n}$  of  $M_1$  is taken as the positive orientation. As shown in Figure 18.3.1, the positive orientation is defined as rotating the normal vector counterclockwise to the direction of the tangent vector  $\mathbf{v}$ . Similarly, for surface  $M_2$ ,

$$-\oint_{M_2} K_2 d\sigma_2 = 2\pi\chi, \quad (18.3.3)$$

where  $K_2$  is the Gaussian curvature on  $M_2$ ,  $d\sigma_2$  is the oriented area element on  $M_2$ , and  $\chi$  is the Euler characteristic of  $M_2$ , with  $\chi = 2$ . Since the outward normal direction of  $M_1$  is taken as the positive orientation, the positive orientation of  $M_2$  should be the inward normal direction. Because the orientations of  $M_1$  and  $M_2$  are opposite, a negative sign should be added to the left-hand side of equation (18.3.3). Since the Euler characteristics of the two surfaces are equal, equations (18.3.2) and (18.3.3) yield

$$\oint_{M_1} K_1 d\sigma_1 = -\oint_{M_2} K_2 d\sigma_2.$$

Rearranging gives

$$\oint_{M_1} K_1 d\sigma_1 + \oint_{M_2} K_2 d\sigma_2 = 0.$$

Since  $M_1$  and  $M_2$  represent the same surface in  $RP^3$ , we have  $M_1 = M_2$  and  $K_1 = K_2$ . Thus, the above equation can be combined as

$$2 \oint_{M_1} K_1 d\sigma_1 = 0,$$

that is

$$\oint_{M_1} K_1 d\sigma_1 = 0.$$

This holds for any closed surface  $M$  enclosing the origin. Therefore, rewriting the equation as

$$\oint_M K d\sigma = 0.$$

Taking the outward normal vector  $\mathbf{n}$  of  $M$  as the positive orientation, we express it in vector form:

$$\oint_M \eta K \mathbf{n} \cdot \mathbf{n} d\sigma = 0. \quad (18.3.4)$$

Here,  $\eta$  is a proportionality constant. Using a method similar to the previous one, let

$$\mathbf{E} = \eta K \mathbf{n}, \quad (18.3.5)$$

and

$$d\sigma = \mathbf{n} d\sigma,$$

we obtain

$$\oint_M \mathbf{E} \cdot d\sigma = 0. \quad (18.3.6)$$

This equation is identical to equation (18.3.1).

### 3. Why Light Waves Are Transverse Waves

Waves are categorized into longitudinal waves and transverse waves. It is well known that light waves are transverse, but why is that the case? The answer lies in the fact that a photon is a  $RP^3$  space. As shown in Figure 18.3.2, each ray can be viewed as the direction of light wave propagation. However, all points along a given ray are identical and indistinguishable—for example, points  $p_1$ ,  $p_2$ , and  $p_3$  are all the same. Consequently, the electric field intensity and magnetic field intensity do not vary along the direction of any ray, which prevents light waves from being longitudinal. On the other hand, points on different rays are distinct, allowing light waves to vary in directions perpendicular to each ray. This forces light waves to be transverse.

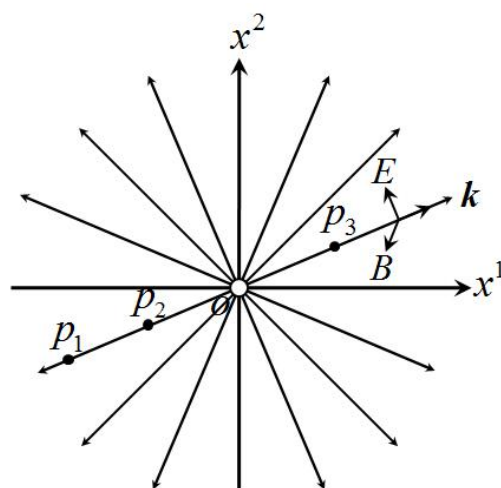


Figure 18.3.2 Light waves are transverse

## §18.4 Derivation of Maxwell's Equations

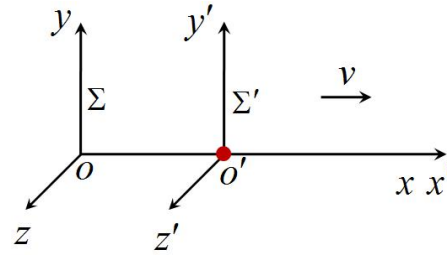
The main content of this section is derived from the article "Other Approaches to Establishing Maxwell's Equations" by Chen Ximou and Shu Yousheng<sup>3</sup>.

Maxwell's equations describe the variation laws of electromagnetic fields, and a vast body of experimental evidence confirms their correctness. Using the Lorentz transformation, Maxwell's equations can be derived from equation (18.1.22).

Consider two coordinate systems  $\Sigma$  and  $\Sigma'$ . Initially, their  $x$ -axes coincide,  $y$ -axes coincide, and  $z$ -axes coincide, with the origins  $O$  and  $O'$  also coinciding. The directions of the three

<sup>3</sup>Chen Ximou, Shu Yousheng. Other Approaches to Establishing Maxwell's Equations [J]. University Physics, 1984, (2): 7–12.

coordinate axes are the same. When the origins of the two coordinate systems coincide, their coordinates are both  $(0,0,0)$ , as shown in Figure 18.4.1. Suppose a particle with charge  $Q$  is located at the origin of coordinate system  $\Sigma'$ , and system  $\Sigma'$  moves with velocity  $v$  along the positive  $x$ -direction. Based on observation, the charged particle produces a static electric field in system  $\Sigma'$ , while in system  $\Sigma$ , it generates both electric and magnetic fields. We derive the transformation laws of electromagnetic fields between these two coordinate systems—Maxwell's equations.



**Figure 18.4.1** The charged particle is located at the origin of the  $\Sigma'$  frame and moves rightward with velocity  $v$  along with the  $\Sigma'$  frame.

### 1. The First Equation

The Lorentz transformation formulas between coordinate systems  $\Sigma'$  and  $\Sigma$  are

$$\begin{cases} x' = \gamma(x - vt), \\ y' = y, \\ z' = z, \\ t' = \gamma(t - \frac{v}{c^2}x), \end{cases} \quad (18.4.1)$$

and the inverse transformation formulas are

$$\begin{cases} x = \gamma(x' + vt'), \\ y = y', \\ z = z', \\ t = \gamma(t' + \frac{v}{c^2}x'), \end{cases} \quad (18.4.2)$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad \frac{1 - \gamma^2}{\gamma^2} = -\frac{v^2}{c^2}.$$

In coordinate system  $\Sigma'$ , from equation (18.1.16), we have

$$\mathbf{E}' = \frac{Q\mathbf{r}'}{4\pi\epsilon_0 r'^3}. \quad (18.4.3)$$

And from equation (18.1.22),

$$\nabla' \cdot \mathbf{E}' = \frac{\rho'}{\epsilon_0}. \quad (18.4.4)$$

From equation (18.4.3), we get

$$\nabla' \times \mathbf{E}' = 0, \quad (18.4.5)$$

which in component form is

$$\frac{\partial E_{z'}}{\partial y'} - \frac{\partial E_{y'}}{\partial z'} = 0, \quad \frac{\partial E_{x'}}{\partial z'} - \frac{\partial E_{z'}}{\partial x'} = 0, \quad \frac{\partial E_{y'}}{\partial x'} - \frac{\partial E_{x'}}{\partial y'} = 0.$$

In coordinate system  $\Sigma$ , the charged particle is moving, and its volume contracts. Let the volume element in  $\Sigma$  be  $dV$ , the charge density be  $\rho$ , the volume element in  $\Sigma'$  be  $dV'$ , and the charge density be  $\rho'$ . Then

$$dV' = \gamma dV, \quad \rho' = \frac{\rho}{\gamma}.$$

In coordinate system  $\Sigma'$ ,

$$\frac{\partial}{\partial x'} = \frac{\partial}{\partial x} \frac{\partial x}{\partial x'} + \frac{\partial}{\partial t} \frac{\partial t}{\partial x'} = \gamma \frac{\partial}{\partial x} + \gamma \frac{v}{c^2} \frac{\partial}{\partial t}, \quad \frac{\partial}{\partial y'} = \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial z'} = \frac{\partial}{\partial z};$$

$$r' = \sqrt{x'^2 + y'^2 + z'^2} = \sqrt{\gamma^2(x - vt)^2 + y^2 + z^2}.$$

Therefore,

$$\begin{aligned} \nabla' \cdot \frac{\mathbf{r}'}{r'^3} &= \left( \frac{\partial}{\partial x'}, \frac{\partial}{\partial y'}, \frac{\partial}{\partial z'} \right) \cdot \frac{\mathbf{r}'}{r'^3} = \frac{\partial}{\partial x'} \frac{x'}{r'^3} + \frac{\partial}{\partial y'} \frac{y'}{r'^3} + \frac{\partial}{\partial z'} \frac{z'}{r'^3} \\ &= \left( \gamma \frac{\partial}{\partial x} + \gamma \frac{v}{c^2} \frac{\partial}{\partial t} \right) \frac{x'}{r'^3} + \frac{\partial}{\partial y} \frac{y'}{r'^3} + \frac{\partial}{\partial z} \frac{z'}{r'^3} = \gamma \frac{\partial}{\partial x} \frac{x'}{r'^3} + \frac{\partial}{\partial y} \frac{y'}{r'^3} + \frac{\partial}{\partial z} \frac{z'}{r'^3} + \gamma \frac{v}{c^2} \frac{\partial}{\partial t} \frac{x'}{r'^3} \\ &= \gamma \frac{\partial}{\partial x} \frac{x'}{r'^3} + \frac{\partial}{\partial y} \frac{y'}{r'^3} + \frac{\partial}{\partial z} \frac{z'}{r'^3} + \gamma \frac{v}{c^2} \frac{\partial}{\partial t} \frac{x'}{r'^3} + \gamma \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} - \gamma \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} + \gamma \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} - \gamma \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} \\ &= \gamma \frac{\partial}{\partial x} \frac{x'}{r'^3} + \gamma \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} + \gamma \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} + \frac{\partial}{\partial y} \frac{y'}{r'^3} - \gamma \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} + \frac{\partial}{\partial z} \frac{z'}{r'^3} - \gamma \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} + \gamma \frac{v}{c^2} \frac{\partial}{\partial t} \frac{x'}{r'^3} \\ &= \gamma \left[ \left( \frac{\partial}{\partial x} \frac{x'}{r'^3} + \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} + \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} \right) + \frac{1 - \gamma^2}{\gamma^2} \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} + \frac{1 - \gamma^2}{\gamma^2} \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} + \frac{v}{c^2} \frac{\partial}{\partial t} \frac{x'}{r'^3} \right] \\ &= \gamma \left[ \left( \frac{\partial}{\partial x} \frac{x'}{r'^3} + \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} + \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} \right) - \frac{v^2}{c^2} \frac{\partial}{\partial y} \frac{\eta y'}{r'^3} - \frac{v^2}{c^2} \frac{\partial}{\partial z} \frac{\eta z'}{r'^3} + \frac{v}{c^2} \frac{\partial}{\partial t} \frac{x'}{r'^3} \right] \\ &\quad \left( \because \frac{1 - \gamma^2}{\gamma^2} = -\frac{v^2}{c^2} \right). \end{aligned}$$

Using the above formulas, we calculate:

$$\begin{aligned} \nabla' \cdot \mathbf{E}' &= \nabla' \cdot \frac{Q}{4\pi\epsilon_0} \frac{\mathbf{r}'}{r'^3} \\ &= \gamma \left[ \left( \frac{\partial}{\partial x} \frac{Qx'}{4\pi\epsilon_0 r'^3} + \frac{\partial}{\partial y} \frac{\eta Qy'}{4\pi\epsilon_0 r'^3} + \frac{\partial}{\partial z} \frac{\eta Qz'}{4\pi\epsilon_0 r'^3} \right) \right. \\ &\quad \left. - \frac{v^2}{c^2} \frac{\partial}{\partial y} \frac{Q\eta y'}{4\pi\epsilon_0 r'^3} - \frac{v^2}{c^2} \frac{\partial}{\partial z} \frac{Q\eta z'}{4\pi\epsilon_0 r'^3} + \frac{v}{c^2} \frac{\partial}{\partial t} \frac{Qx'}{4\pi\epsilon_0 r'^3} \right] \\ &= \gamma \left[ \left( \frac{\partial}{\partial x} \frac{Qx'}{4\pi\epsilon_0 r'^3} + \frac{\partial}{\partial y} \frac{\eta Qy'}{4\pi\epsilon_0 r'^3} + \frac{\partial}{\partial z} \frac{\eta Qz'}{4\pi\epsilon_0 r'^3} \right) - \frac{\partial}{\partial y} \left( \frac{\eta^2}{c^2} \frac{Qy'}{4\pi\epsilon_0 r'^3} \right) - \frac{\partial}{\partial z} \left( \frac{\eta^2}{c^2} \frac{Qz'}{4\pi\epsilon_0 r'^3} \right) \right] \\ &= \gamma \left[ \left( \frac{\partial}{\partial x} \frac{Qx'}{4\pi\epsilon_0 r'^3} + \frac{\partial}{\partial y} \frac{\eta Qy'}{4\pi\epsilon_0 r'^3} + \frac{\partial}{\partial z} \frac{\eta Qz'}{4\pi\epsilon_0 r'^3} \right) + v \left[ -\frac{\partial}{\partial y} \left( \frac{v\eta}{c^2} \frac{Qy'}{4\pi\epsilon_0 r'^3} \right) - \frac{\partial}{\partial z} \left( \frac{v\eta}{c^2} \frac{Qz'}{4\pi\epsilon_0 r'^3} \right) \right] \right] \\ &= \frac{\rho'}{\epsilon_0} \quad \left( \because \frac{v}{c^2} \frac{\partial}{\partial t} \frac{Qx'}{4\pi\epsilon_0 r'^3} = 0 \right). \end{aligned}$$

Define:

$$\begin{aligned} E_{x'} &= \frac{Qx'}{4\pi\epsilon_0 r'^3}, & E_{y'} &= \frac{Qy'}{4\pi\epsilon_0 r'^3}, & E_{z'} &= \frac{Qz'}{4\pi\epsilon_0 r'^3}; \\ E_x &= E_{x'}, & E_y &= \eta E_{y'}, & E_z &= \eta E_{z'}; \\ B_x &= 0, & B_y &= -\frac{v}{c^2} \gamma \frac{Qz'}{4\pi\epsilon_0 r'^3} = -\frac{v}{c^2} \eta E_{z'}, & B_z &= \frac{v}{c^2} \gamma \frac{Qy'}{4\pi\epsilon_0 r'^3} = \frac{v}{c^2} \eta E_{y'}. \end{aligned}$$

Furthermore, define:

$$\begin{aligned} E_x, E_y, E_z &\text{ form the electric field intensity vector } \mathbf{E} = (E_x, E_y, E_z); \\ E_{x'}, E_{y'}, E_{z'} &\text{ form the electric field intensity vector } \mathbf{E}' = (E_{x'}, E_{y'}, E_{z'}); \\ B_x, B_y, B_z &\text{ form the magnetic field intensity vector } \mathbf{B} = (B_x, B_y, B_z); \\ B_{x'}, B_{y'}, B_{z'} &\text{ form the magnetic field intensity vector } \mathbf{B}' = (B_{x'}, B_{y'}, B_{z'}). \end{aligned}$$

In coordinate system  $\Sigma$ ,

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x'} \frac{\partial x'}{\partial x} + \frac{\partial}{\partial t'} \frac{\partial t'}{\partial x} = \gamma \frac{\partial}{\partial x'} - \gamma \frac{v}{c^2} \frac{\partial}{\partial t'}, \quad \frac{\partial}{\partial y} = \frac{\partial}{\partial y'}, \quad \frac{\partial}{\partial z} = \frac{\partial}{\partial z'};$$

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t'} \frac{\partial t'}{\partial t} + \frac{\partial}{\partial x'} \frac{\partial x'}{\partial t} = \gamma \frac{\partial}{\partial t'} - \gamma v \frac{\partial}{\partial x'} ;$$

$$\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) = \left( \gamma \frac{\partial}{\partial x'} - \gamma \frac{v}{c^2} \frac{\partial}{\partial t'}, \frac{\partial}{\partial y'}, \frac{\partial}{\partial z'} \right).$$

Calculate:

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) (E_x, E_y, E_z) = \left( \gamma \frac{\partial}{\partial x'} - \gamma \frac{v}{c^2} \frac{\partial}{\partial t'}, \frac{\partial}{\partial y'}, \frac{\partial}{\partial z'} \right) (\gamma E_{x'}, \gamma E_{y'}, \gamma E_{z'}) \\ &= \gamma \left( \frac{\partial}{\partial x'} - \frac{v}{c^2} \frac{\partial}{\partial t'} \right) E_{x'} + \frac{\partial}{\partial y'} (\gamma E_{y'}) + \frac{\partial}{\partial z'} (\gamma E_{z'}) \\ &= \gamma \left( \frac{\partial E_{x'}}{\partial x'} + \frac{\partial E_{y'}}{\partial y'} + \frac{\partial E_{z'}}{\partial z'} \right) \quad \left( \because \frac{v}{c^2} \frac{\partial}{\partial t'} E_{x'} = 0 \right) \\ &= \gamma \nabla' \cdot \mathbf{E}' = \gamma \frac{\rho'}{\epsilon_0} = \gamma \frac{\rho}{\gamma \epsilon_0} = \frac{\rho}{\epsilon_0}. \end{aligned}$$

Thus, we obtain the first equation,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}.$$

## 2.The Second Equation

Next, calculate:

$$\begin{aligned} \nabla \cdot \mathbf{B} &= \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) (B_x, B_y, B_z) = \left( \gamma \frac{\partial}{\partial x'} - \gamma \frac{v}{c^2} \frac{\partial}{\partial t'}, \frac{\partial}{\partial y'}, \frac{\partial}{\partial z'} \right) \left( 0, -\frac{v}{c^2} \gamma E_{z'}, \frac{v}{c^2} \gamma E_{y'} \right) \\ &= -\frac{v}{c^2} \gamma \frac{\partial E_{z'}}{\partial y'} + \frac{v}{c^2} \gamma \frac{\partial E_{y'}}{\partial z'} = -\frac{v}{c^2} \gamma \left( \frac{\partial E_{z'}}{\partial y'} - \frac{\partial E_{y'}}{\partial z'} \right) = 0. \end{aligned}$$

Thus, we derive the second equation,

$$\nabla \cdot \mathbf{B} = 0.$$

## 3.The Third Equation

Further calculate:

$$\begin{aligned} \nabla \times \mathbf{E} &= \left( \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \mathbf{i} + \left( \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) \mathbf{j} + \left( \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \mathbf{k} \\ &= \left[ \frac{\partial}{\partial y'} (\gamma E_{z'}) - \frac{\partial}{\partial z'} (\gamma E_{y'}) \right] \mathbf{i} + \left[ \frac{\partial E_{x'}}{\partial z'} - \gamma \left( \frac{\partial}{\partial x'} - \frac{v}{c^2} \frac{\partial}{\partial t'} \right) \gamma E_{z'} \right] \mathbf{j} + \left[ \gamma \left( \frac{\partial}{\partial x'} - \frac{v}{c^2} \frac{\partial}{\partial t'} \right) \gamma E_{y'} - \frac{\partial E_{x'}}{\partial y'} \right] \mathbf{k} \\ &= \gamma \left( \frac{\partial E_{z'}}{\partial y'} - \frac{\partial E_{y'}}{\partial z'} \right) \mathbf{i} + \left( \frac{\partial E_{x'}}{\partial z'} - \gamma^2 \frac{\partial E_{z'}}{\partial x'} \right) \mathbf{j} + \left( \gamma^2 \frac{\partial E_{y'}}{\partial x'} - \frac{\partial E_{x'}}{\partial y'} \right) \mathbf{k} \\ &= \left( \frac{\partial E_{x'}}{\partial z'} - \gamma^2 \frac{\partial E_{z'}}{\partial x'} \right) \mathbf{j} + \left( \gamma^2 \frac{\partial E_{y'}}{\partial x'} - \frac{\partial E_{x'}}{\partial y'} \right) \mathbf{k}, \end{aligned} \tag{18.4.6}$$

$$\begin{aligned} \frac{\partial \mathbf{B}}{\partial t} &= \left( \gamma \frac{\partial}{\partial t'} - \gamma v \frac{\partial}{\partial x'} \right) (B_x \mathbf{i} + B_y \mathbf{j} + B_z \mathbf{k}) \\ &= \gamma \left( \frac{\partial}{\partial t'} - v \frac{\partial}{\partial x'} \right) \left( 0 \mathbf{i} - \frac{v}{c^2} \gamma E_{z'} \mathbf{j} + \frac{v}{c^2} \gamma E_{y'} \mathbf{k} \right) \\ &= \frac{v^2}{c^2} \gamma^2 \frac{\partial E_{z'}}{\partial x'} \mathbf{j} - \frac{v^2}{c^2} \gamma^2 \frac{\partial E_{y'}}{\partial x'} \mathbf{k} = \left( -\frac{1-\gamma^2}{\gamma^2} \right) \gamma^2 \frac{\partial E_{z'}}{\partial x'} \mathbf{j} - \left( -\frac{1-\gamma^2}{\gamma^2} \right) \gamma^2 \frac{\partial E_{y'}}{\partial x'} \mathbf{k} \\ &= (\gamma^2 - 1) \frac{\partial E_{z'}}{\partial x'} \mathbf{j} + (1 - \gamma^2) \frac{\partial E_{y'}}{\partial x'} \mathbf{k}. \end{aligned} \tag{18.4.7}$$

From equations (18.4.6), (18.4.7), and (18.4.5), we find

$$\begin{aligned}\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= \left( \frac{\partial E_{x'}}{\partial z'} - \gamma^2 \frac{\partial E_{z'}}{\partial x'} + (\gamma^2 - 1) \frac{\partial E_{z'}}{\partial x'} \right) \mathbf{j} + \left( \gamma^2 \frac{\partial E_{y'}}{\partial x'} - \frac{\partial E_{x'}}{\partial y'} + (1 - \gamma^2) \frac{\partial E_{y'}}{\partial x'} \right) \mathbf{k} \\ &= \left( \frac{\partial E_{x'}}{\partial z'} - \frac{\partial E_{z'}}{\partial x'} \right) \mathbf{j} + \left( \frac{\partial E_{y'}}{\partial x'} - \frac{\partial E_{x'}}{\partial y'} \right) \mathbf{k} = 0.\end{aligned}$$

Thus, we obtain the third equation,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (18.4.8)$$

#### 4. The Fourth Equation

Calculate:

$$\begin{aligned}\nabla \times \mathbf{B} &= \left( \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \right) \mathbf{i} + \left( \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} \right) \mathbf{j} + \left( \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right) \mathbf{k} \\ &= \left[ \frac{\partial}{\partial y'} \left( \frac{v\gamma}{c^2} E_{y'} \right) - \frac{\partial}{\partial z'} \left( -\frac{v\gamma}{c^2} E_{z'} \right) \right] \mathbf{i} - \left[ \gamma \left( \frac{\partial}{\partial x'} - \frac{v}{c^2} \frac{\partial}{\partial t'} \right) \frac{v\gamma}{c^2} E_{y'} \right] \mathbf{j} \\ &\quad + \left[ \gamma \left( \frac{\partial}{\partial x'} - \frac{v}{c^2} \frac{\partial}{\partial t'} \right) \left( -\frac{v\gamma}{c^2} E_{z'} \right) \right] \mathbf{k} \\ &= \frac{v\gamma}{c^2} \left( \frac{\partial E_{y'}}{\partial y'} + \frac{\partial E_{z'}}{\partial z'} \right) \mathbf{i} - \frac{v\gamma^2}{c^2} \frac{\partial E_{y'}}{\partial x'} \mathbf{j} - \frac{v\gamma^2}{c^2} \frac{\partial E_{z'}}{\partial x'} \mathbf{k} \\ &= \frac{v\gamma}{c^2} \left( \frac{\partial E_{x'}}{\partial x'} + \frac{\partial E_{y'}}{\partial y'} + \frac{\partial E_{z'}}{\partial z'} \right) \mathbf{i} - \frac{v\gamma}{c^2} \frac{\partial E_{x'}}{\partial x'} \mathbf{i} - \frac{v\gamma^2}{c^2} \frac{\partial E_{y'}}{\partial x'} \mathbf{j} - \frac{v\gamma^2}{c^2} \frac{\partial E_{z'}}{\partial x'} \mathbf{k} \\ &= \frac{v\gamma}{c^2} (\nabla' \cdot \mathbf{E}') \mathbf{i} - \frac{v\gamma}{c^2} \frac{\partial E_{x'}}{\partial x'} \mathbf{i} - \frac{v\gamma^2}{c^2} \frac{\partial E_{y'}}{\partial x'} \mathbf{j} - \frac{v\gamma^2}{c^2} \frac{\partial E_{z'}}{\partial x'} \mathbf{k} \\ &= \frac{v\gamma}{c^2} \frac{\rho'}{\varepsilon_0} \mathbf{i} - \frac{v\gamma}{c^2} \frac{\partial E_{x'}}{\partial x'} \mathbf{i} - \frac{v\gamma^2}{c^2} \frac{\partial E_{y'}}{\partial x'} \mathbf{j} - \frac{v\gamma^2}{c^2} \frac{\partial E_{z'}}{\partial x'} \mathbf{k}.\end{aligned}$$

Finally, we obtain

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} - \frac{v\gamma}{c^2} \frac{\partial E_{x'}}{\partial x'} \mathbf{i} - \frac{v\gamma^2}{c^2} \frac{\partial E_{y'}}{\partial x'} \mathbf{j} - \frac{v\gamma^2}{c^2} \frac{\partial E_{z'}}{\partial x'} \mathbf{k}, \quad (18.4.9)$$

where

$$\frac{v\gamma}{c^2} \frac{\rho'}{\varepsilon_0} \mathbf{i} = \varepsilon_0 \mu_0 \gamma v \frac{\rho'}{\varepsilon_0} \mathbf{i} = \mu_0 v \rho \mathbf{i} = \mu_0 \mathbf{J}, \quad \mathbf{J} = v \rho \mathbf{i}, \quad c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.$$

Calculate:

$$\begin{aligned}\frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} &= \frac{1}{c^2} \left( \gamma \frac{\partial}{\partial t'} - \gamma v \frac{\partial}{\partial x'} \right) (E_x \mathbf{i} + E_y \mathbf{j} + E_z \mathbf{k}) = \frac{\gamma}{c^2} \left( \frac{\partial}{\partial t'} - v \frac{\partial}{\partial x'} \right) (E_x \mathbf{i} + \gamma E_{y'} \mathbf{j} + \gamma E_{z'} \mathbf{k}) \\ &= -\frac{v\gamma}{c^2} \frac{\partial E_{x'}}{\partial x'} \mathbf{i} - \frac{v\gamma^2}{c^2} \frac{\partial E_{y'}}{\partial x'} \mathbf{j} - \frac{v\gamma^2}{c^2} \frac{\partial E_{z'}}{\partial x'} \mathbf{k}.\end{aligned} \quad (18.4.10)$$

Subtract equation (18.4.10) from equation (18.4.9) to get

$$\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J},$$

which yields the fourth equation,

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}.$$

Summarizing the above results, we obtain the following Maxwell's equations:

$$\begin{cases} \nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}, \\ \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \\ \nabla \cdot \mathbf{B} = 0, \\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}. \end{cases} \quad (18.4.11)$$

## 5. Tensor Form of the Equations

Maxwell's equations (18.4.11) are not form-invariant under Lorentz transformations. However, since tensor equations maintain their form under coordinate transformations, we rewrite equations (18.4.11) in tensor form.

Let the coordinates of four-dimensional Euclidean space be

$$x^\mu = (x^1, x^2, x^3, x^4), \quad \mu, \nu = 1, 2, 3, 4.$$

The metric form is

$$ds^2 = d(x^1)^2 + d(x^2)^2 + d(x^3)^2 + d(x^4)^2,$$

and the metric tensor is

$$g_{11} = g_{22} = g_{33} = g_{44} = 1, \quad g_{\mu\nu} = 0 (\mu \neq \nu),$$

so

$$\sqrt{G} = 1. \quad (18.4.12)$$

If we let

$$x^\mu = (x^1, x^2, x^3, x^4) = (x, y, z, ict),$$

then

$$ds^2 = d(x^1)^2 + d(x^2)^2 + d(x^3)^2 + d(x^4)^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2.$$

Therefore, if we take the coordinates of Minkowski spacetime as  $x^\mu = (x, y, z, ict)$ , the metric tensor becomes

$$g_{11} = g_{22} = g_{33} = 1, \quad g_{44} = -c^2, \quad g_{\mu\nu} = 0 (\mu \neq \nu),$$

so

$$\sqrt{G} = \sqrt{-c^2} = ic.$$

The following discussion of the tensor form of Maxwell's equations takes place in four-dimensional Euclidean space, while also accommodating Minkowski spacetime coordinates. That is, during calculations, we set  $x^1 = x, x^2 = y, x^3 = z, x^4 = ict$ .

First, construct a second-order covariant antisymmetric tensor

$$F_{\mu\nu} = \begin{pmatrix} 0 & B_3 & -B_2 & -\frac{i}{c}E_1 \\ -B_3 & 0 & B_1 & -\frac{i}{c}E_2 \\ B_2 & -B_1 & 0 & -\frac{i}{c}E_3 \\ \frac{i}{c}E_1 & \frac{i}{c}E_2 & \frac{i}{c}E_3 & 0 \end{pmatrix}, \quad (18.4.13)$$

where  $i$  is the imaginary unit. Next, construct a four-dimensional covariant current density vector

$$J_\mu = (J_1, J_2, J_3, J_4) = (J_1, J_2, J_3, ic\rho).$$

In four-dimensional Euclidean space, the second-order contravariant antisymmetric tensor is the same as the second-order covariant antisymmetric tensor, i.e.,

$$F_{\mu\nu} = F^{\mu\nu};$$

Similarly, the four-dimensional contravariant current density vector  $J^\mu$  is also the same as the four-dimensional covariant current density vector  $J_\mu$ , i.e.,

$$J^\mu = J_\mu,$$

which means

$$J^1 = J_1, \quad J^2 = J_2, \quad J^3 = J_3, \quad J^4 = J_4.$$

The transformation matrix of equation (18.4.1) is

$$a = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix}, \quad (18.4.14)$$

where  $\beta = \frac{v}{c}$ . The transformation of  $F_{\mu\nu}$  is

$$F'_{\mu\nu} = a^\lambda_\mu a^\tau_\nu F_{\lambda\tau},$$

and its component transformation relations correspond to the electromagnetic field transformation relations:

$$\begin{cases} E'_1 = E_1, \\ E'_2 = \gamma(E_2 - vB_3), \\ E'_3 = \gamma(E_3 + vB_2), \end{cases} \quad (18.4.15)$$

$$B'_1 = B_1, \quad B'_2 = \gamma(B_2 + \frac{v}{c^2}E_3), \quad B'_3 = \gamma(B_3 - \frac{v}{c^2}E_2);$$

Alternatively,

$$\begin{aligned} E_1 &= E'_1, & E_2 &= \gamma(E'_2 + vB'_3), & E_3 &= \gamma(E'_3 - vB'_2), \\ B_1 &= B'_1, & B_2 &= \gamma(B'_2 - \frac{v}{c^2}E'_3), & B_3 &= \gamma(B'_3 + \frac{v}{c^2}E'_2). \end{aligned}$$

We express Maxwell's equations in the following tensor form:

$$dF = 0, \quad (18.4.16)$$

$$d(*F) = \mu_0(*J), \quad (18.4.17)$$

where

$$F = E_1 dx \wedge dt + E_2 dy \wedge dt + E_3 dz \wedge dt + B_1 dy \wedge dz + B_2 dz \wedge dx + B_3 dx \wedge dy,$$

$$*F = B_1 dx \wedge d(ict) + B_2 dy \wedge d(ict) + B_3 dz \wedge d(ict) - \frac{i}{c} E_1 dy \wedge dz - \frac{i}{c} E_2 dz \wedge dx - \frac{i}{c} E_3 dx \wedge dy,$$

$$*J = J^1 dy \wedge dz \wedge d(ict) + J^2 dz \wedge dx \wedge d(ict) + J^3 dx \wedge dy \wedge d(ict) - J^4 dx \wedge dy \wedge dz.$$

The system consisting of equations (18.4.16) and (18.4.17) encompasses all four Maxwell's equations and remains form-invariant under arbitrary coordinate transformations. Below, we detail the construction process of this system.

First, following equation (11.13.4), we construct the following 2-form differential form:

$$\begin{aligned} F &= \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu = \frac{1}{2} F_{1\nu} dx^1 \wedge dx^\nu + \frac{1}{2} F_{2\nu} dx^2 \wedge dx^\nu + \frac{1}{2} F_{3\nu} dx^3 \wedge dx^\nu + \frac{1}{2} F_{4\nu} dx^4 \wedge dx^\nu \\ &= \frac{1}{2} F_{11} dx^1 \wedge dx^1 + \frac{1}{2} F_{12} dx^1 \wedge dx^2 + \frac{1}{2} F_{13} dx^1 \wedge dx^3 + \frac{1}{2} F_{14} dx^1 \wedge dx^4 \\ &\quad + \frac{1}{2} F_{21} dx^2 \wedge dx^1 + \frac{1}{2} F_{22} dx^2 \wedge dx^2 + \frac{1}{2} F_{23} dx^2 \wedge dx^3 + \frac{1}{2} F_{24} dx^2 \wedge dx^4 \\ &\quad + \frac{1}{2} F_{31} dx^3 \wedge dx^1 + \frac{1}{2} F_{32} dx^3 \wedge dx^2 + \frac{1}{2} F_{33} dx^3 \wedge dx^3 + \frac{1}{2} F_{34} dx^3 \wedge dx^4 \\ &\quad + \frac{1}{2} F_{41} dx^4 \wedge dx^1 + \frac{1}{2} F_{42} dx^4 \wedge dx^2 + \frac{1}{2} F_{43} dx^4 \wedge dx^3 + \frac{1}{2} F_{44} dx^4 \wedge dx^4 \\ &= \frac{1}{2} B_3 dx \wedge dy + \frac{1}{2} (-B_2) dx \wedge dz + \frac{1}{2} E_1 dx \wedge dt \\ &\quad + \frac{1}{2} (-B_3) dy \wedge dx + \frac{1}{2} B_1 dy \wedge dz + \frac{1}{2} E_2 dy \wedge dt \\ &\quad + \frac{1}{2} B_2 dz \wedge dx + \frac{1}{2} (-B_1) dz \wedge dy + \frac{1}{2} E_3 dz \wedge dt \\ &\quad + \frac{1}{2} (-E_1) dt \wedge dx + \frac{1}{2} (-E_2) dt \wedge dy + \frac{1}{2} (-E_3) dt \wedge dz, \end{aligned}$$

thus,



$$F = E_1 dx \wedge dt + E_2 dy \wedge dt + E_3 dz \wedge dt + B_1 dy \wedge dz + B_2 dz \wedge dx + B_3 dx \wedge dy .$$

Taking the exterior derivative of  $F$ , we obtain

$$\begin{aligned} dF &= \frac{\partial E_1}{\partial y} dy \wedge dx \wedge dt + \frac{\partial E_1}{\partial z} dz \wedge dx \wedge dt + \frac{\partial E_2}{\partial x} dx \wedge dy \wedge dt + \frac{\partial E_2}{\partial z} dz \wedge dy \wedge dt \\ &\quad + \frac{\partial E_3}{\partial x} dx \wedge dz \wedge dt + \frac{\partial E_3}{\partial y} dy \wedge dz \wedge dt \\ &\quad + \frac{\partial B_1}{\partial x} dx \wedge dy \wedge dz + \frac{\partial B_1}{\partial t} dt \wedge dy \wedge dz + \frac{\partial B_2}{\partial y} dy \wedge dz \wedge dx + \frac{\partial B_2}{\partial t} dt \wedge dz \wedge dx \\ &\quad + \frac{\partial B_3}{\partial z} dz \wedge dx \wedge dy + \frac{\partial B_3}{\partial t} dt \wedge dx \wedge dy \\ &= \left( \frac{\partial E_3}{\partial y} - \frac{\partial E_2}{\partial z} + \frac{\partial B_1}{\partial t} \right) dy \wedge dz \wedge dt + \left( \frac{\partial E_1}{\partial z} - \frac{\partial E_3}{\partial x} + \frac{\partial B_2}{\partial t} \right) dz \wedge dx \wedge dt \\ &\quad + \left( \frac{\partial E_2}{\partial x} - \frac{\partial E_1}{\partial y} + \frac{\partial B_3}{\partial t} \right) dx \wedge dy \wedge dt + \left( \frac{\partial B_1}{\partial x} + \frac{\partial B_2}{\partial y} + \frac{\partial B_3}{\partial z} \right) dx \wedge dy \wedge dz . \end{aligned}$$

Setting  $dF = 0$  yields:

$$\begin{aligned} \frac{\partial E_3}{\partial y} - \frac{\partial E_2}{\partial z} + \frac{\partial B_1}{\partial t} &= 0, & \frac{\partial E_1}{\partial z} - \frac{\partial E_3}{\partial x} + \frac{\partial B_2}{\partial t} &= 0, \\ \frac{\partial E_2}{\partial x} - \frac{\partial E_1}{\partial y} + \frac{\partial B_3}{\partial t} &= 0, & \frac{\partial B_1}{\partial x} + \frac{\partial B_2}{\partial y} + \frac{\partial B_3}{\partial z} &= 0, \end{aligned}$$

which gives two of Maxwell's equations:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (18.4.18)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (18.4.19)$$

According to equation (11.13.9), we apply the Hodge  $*$  operator to  $F$ :

$$\begin{aligned} *F &= \frac{1}{(4-2)!} \frac{\sqrt{G}}{2!} \delta_{i_1 i_2 i_3 i_4}^{1234} F^{i_1 i_2} dx^{i_3} \wedge dx^{i_4} \quad (\sqrt{G} = 1) \\ &= \frac{1}{4} \delta_{12 i_3 i_4}^{1234} F^{1 i_2} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{2 i_2 i_3 i_4}^{1234} F^{2 i_2} dx^{i_3} \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{3 i_2 i_3 i_4}^{1234} F^{3 i_2} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{4 i_2 i_3 i_4}^{1234} F^{4 i_2} dx^{i_3} \wedge dx^{i_4} \\ &= \frac{1}{4} \delta_{12 i_3 i_4}^{1234} F^{12} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{13 i_3 i_4}^{1234} F^{13} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{14 i_3 i_4}^{1234} F^{14} dx^{i_3} \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{21 i_3 i_4}^{1234} F^{21} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{23 i_3 i_4}^{1234} F^{23} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{24 i_3 i_4}^{1234} F^{24} dx^{i_3} \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{31 i_3 i_4}^{1234} F^{31} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{32 i_3 i_4}^{1234} F^{32} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{34 i_3 i_4}^{1234} F^{34} dx^{i_3} \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{41 i_3 i_4}^{1234} F^{41} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{42 i_3 i_4}^{1234} F^{42} dx^{i_3} \wedge dx^{i_4} + \frac{1}{4} \delta_{43 i_3 i_4}^{1234} F^{43} dx^{i_3} \wedge dx^{i_4} \\ &= \frac{1}{4} \delta_{123 i_4}^{1234} F^{12} dx^3 \wedge dx^{i_4} + \frac{1}{4} \delta_{124 i_4}^{1234} F^{12} dx^4 \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{132 i_4}^{1234} F^{13} dx^2 \wedge dx^{i_4} + \frac{1}{4} \delta_{134 i_4}^{1234} F^{13} dx^4 \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{142 i_4}^{1234} F^{14} dx^2 \wedge dx^{i_4} + \frac{1}{4} \delta_{143 i_4}^{1234} F^{14} dx^3 \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{213 i_4}^{1234} F^{21} dx^3 \wedge dx^{i_4} + \frac{1}{4} \delta_{214 i_4}^{1234} F^{21} dx^4 \wedge dx^{i_4} \\ &\quad + \frac{1}{4} \delta_{231 i_4}^{1234} F^{23} dx^1 \wedge dx^{i_4} + \frac{1}{4} \delta_{234 i_4}^{1234} F^{23} dx^4 \wedge dx^{i_4} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{4} \delta_{241i_4}^{1234} F^{24} dx^1 \wedge dx^{i_4} + \frac{1}{4} \delta_{243i_4}^{1234} F^{24} dx^3 \wedge dx^{i_4} \\
& + \frac{1}{4} \delta_{312i_4}^{1234} F^{31} dx^2 \wedge dx^{i_4} + \frac{1}{4} \delta_{314i_4}^{1234} F^{31} dx^4 \wedge dx^{i_4} \\
& + \frac{1}{4} \delta_{321i_4}^{1234} F^{32} dx^1 \wedge dx^{i_4} + \frac{1}{4} \delta_{324i_4}^{1234} F^{32} dx^4 \wedge dx^{i_4} \\
& + \frac{1}{4} \delta_{341i_4}^{1234} F^{34} dx^1 \wedge dx^{i_4} + \frac{1}{4} \delta_{342i_4}^{1234} F^{34} dx^2 \wedge dx^{i_4} \\
& + \frac{1}{4} \delta_{412i_4}^{1234} F^{41} dx^2 \wedge dx^{i_4} + \frac{1}{4} \delta_{413i_4}^{1234} F^{41} dx^3 \wedge dx^{i_4} \\
& + \frac{1}{4} \delta_{421i_4}^{1234} F^{42} dx^1 \wedge dx^{i_4} + \frac{1}{4} \delta_{423i_4}^{1234} F^{42} dx^3 \wedge dx^{i_4} \\
& + \frac{1}{4} \delta_{431i_4}^{1234} F^{43} dx^1 \wedge dx^{i_4} + \frac{1}{4} \delta_{432i_4}^{1234} F^{43} dx^2 \wedge dx^{i_4} \\
& = \frac{1}{4} \delta_{1234}^{1234} F^{12} dx^3 \wedge dx^4 + \frac{1}{4} \delta_{1243}^{1234} F^{12} dx^4 \wedge dx^3 + \frac{1}{4} \delta_{1324}^{1234} F^{13} dx^2 \wedge dx^4 + \frac{1}{4} \delta_{1342}^{1234} F^{13} dx^4 \wedge dx^2 \\
& + \frac{1}{4} \delta_{1423}^{1234} F^{14} dx^2 \wedge dx^3 + \frac{1}{4} \delta_{1432}^{1234} F^{14} dx^3 \wedge dx^2 \\
& + \frac{1}{4} \delta_{2134}^{1234} F^{21} dx^3 \wedge dx^4 + \frac{1}{4} \delta_{2143}^{1234} F^{21} dx^4 \wedge dx^3 + \frac{1}{4} \delta_{2314}^{1234} F^{23} dx^1 \wedge dx^4 + \frac{1}{4} \delta_{2341}^{1234} F^{23} dx^4 \wedge dx^1 \\
& + \frac{1}{4} \delta_{2413}^{1234} F^{24} dx^1 \wedge dx^3 + \frac{1}{4} \delta_{2431}^{1234} F^{24} dx^3 \wedge dx^1 \\
& + \frac{1}{4} \delta_{3124}^{1234} F^{31} dx^2 \wedge dx^4 + \frac{1}{4} \delta_{3142}^{1234} F^{31} dx^4 \wedge dx^2 + \frac{1}{4} \delta_{3214}^{1234} F^{32} dx^1 \wedge dx^4 + \frac{1}{4} \delta_{3241}^{1234} F^{32} dx^4 \wedge dx^1 \\
& + \frac{1}{4} \delta_{3412}^{1234} F^{34} dx^1 \wedge dx^2 + \frac{1}{4} \delta_{3421}^{1234} F^{34} dx^2 \wedge dx^1 \\
& + \frac{1}{4} \delta_{4123}^{1234} F^{41} dx^2 \wedge dx^3 + \frac{1}{4} \delta_{4132}^{1234} F^{41} dx^3 \wedge dx^2 + \frac{1}{4} \delta_{4213}^{1234} F^{42} dx^1 \wedge dx^3 + \frac{1}{4} \delta_{4231}^{1234} F^{42} dx^3 \wedge dx^1 \\
& + \frac{1}{4} \delta_{4312}^{1234} F^{43} dx^1 \wedge dx^2 + \frac{1}{4} \delta_{4321}^{1234} F^{43} dx^2 \wedge dx^1 \\
& = \frac{1}{2} F^{12} dx^3 \wedge dx^4 - \frac{1}{2} F^{13} dx^2 \wedge dx^4 + \frac{1}{2} F^{14} dx^2 \wedge dx^3 \\
& - \frac{1}{2} F^{21} dx^3 \wedge dx^4 + \frac{1}{2} F^{23} dx^1 \wedge dx^4 - \frac{1}{2} F^{24} dx^1 \wedge dx^3 \\
& + \frac{1}{2} F^{31} dx^2 \wedge dx^4 - \frac{1}{2} F^{32} dx^1 \wedge dx^4 + \frac{1}{2} F^{34} dx^1 \wedge dx^2 \\
& - \frac{1}{2} F^{41} dx^2 \wedge dx^3 + \frac{1}{2} F^{42} dx^1 \wedge dx^3 - \frac{1}{2} F^{43} dx^1 \wedge dx^2 \\
& = F^{12} dx^3 \wedge dx^4 - F^{13} dx^2 \wedge dx^4 + F^{14} dx^2 \wedge dx^3 + F^{23} dx^1 \wedge dx^4 - F^{24} dx^1 \wedge dx^3 + F^{34} dx^1 \wedge dx^2 \\
& = B_3 dz \wedge d(ict) + B_2 dy \wedge d(ict) - \frac{i}{c} E_1 dy \wedge dz + B_1 dx \wedge d(ict) + \frac{i}{c} E_2 dx \wedge dz - \frac{i}{c} E_3 dx \wedge dy,
\end{aligned}$$

that is

$$*F = B_1 dx \wedge d(ict) + B_2 dy \wedge d(ict) + B_3 dz \wedge d(ict) - \frac{i}{c} E_1 dy \wedge dz - \frac{i}{c} E_2 dz \wedge dx - \frac{i}{c} E_3 dx \wedge dy.$$

Taking the exterior derivative of  $*F$ , we obtain

$$\begin{aligned}
d(*F) &= \frac{\partial B_1}{\partial y} dy \wedge dx \wedge d(ict) + \frac{\partial B_1}{\partial z} dz \wedge dx \wedge d(ict) + \frac{\partial B_2}{\partial x} dx \wedge dy \wedge d(ict) + \frac{\partial B_2}{\partial z} dz \wedge dy \wedge d(ict) \\
&+ \frac{\partial B_3}{\partial x} dx \wedge dz \wedge d(ict) + \frac{\partial B_3}{\partial y} dy \wedge dz \wedge d(ict) - \frac{i}{c} \frac{\partial E_1}{\partial x} dx \wedge dy \wedge dz - \frac{i}{c} \frac{\partial E_1}{\partial (ict)} d(ict) \wedge dy \wedge dz
\end{aligned}$$

$$\begin{aligned}
& -\frac{i}{c} \frac{\partial E_2}{\partial y} dy \wedge dz \wedge dx - \frac{i}{c} \frac{\partial E_2}{\partial(ict)} d(ict) \wedge dz \wedge dx - \frac{i}{c} \frac{\partial E_3}{\partial z} dz \wedge dx \wedge dy - \frac{i}{c} \frac{\partial E_3}{\partial(ict)} d(ict) \wedge dx \wedge dy \\
& = \left( \frac{\partial B_3}{\partial y} - \frac{\partial B_2}{\partial z} - \frac{i}{c} \frac{\partial E_1}{\partial(ict)} \right) dy \wedge dz \wedge d(ict) + \left( \frac{\partial B_1}{\partial z} - \frac{\partial B_3}{\partial x} - \frac{i}{c} \frac{\partial E_2}{\partial(ict)} \right) dz \wedge dx \wedge d(ict) \\
& \quad + \left( \frac{\partial B_2}{\partial x} - \frac{\partial B_1}{\partial y} - \frac{i}{c} \frac{\partial E_3}{\partial(ict)} \right) dx \wedge dy \wedge d(ict) - \left( \frac{i}{c} \frac{\partial E_1}{\partial x} + \frac{i}{c} \frac{\partial E_2}{\partial y} + \frac{i}{c} \frac{\partial E_3}{\partial z} \right) dx \wedge dy \wedge dz.
\end{aligned}$$

From the four-dimensional covariant current density vector  $J_\mu$ , we construct a 1-form:

$$\begin{aligned}
J &= J_1 dx^1 + J_2 dx^2 + J_3 dx^3 + J_4 dx^4 = J_1 dx^1 + J_2 dx^2 + J_3 dx^3 + ic\rho d(ict) \\
&= J_1 dx + J_2 dy + J_3 dz - \rho c^2 dt.
\end{aligned}$$

According to equation (11.13.9), we apply the Hodge \* operator to  $J$ :

$$\begin{aligned}
*J &= \frac{1}{(4-1)!} \frac{\sqrt{G}}{1!} \delta_{i_1 i_2 i_3 i_4}^{1234} J^{i_1} dx^{i_2} \wedge dx^{i_3} \wedge dx^{i_4} \quad (\sqrt{G}=1) \\
&= \frac{1}{6} \delta_{1i_2 i_3 i_4}^{1234} J^1 dx^{i_2} \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{2i_2 i_3 i_4}^{1234} J^2 dx^{i_2} \wedge dx^{i_3} \wedge dx^{i_4} \\
&\quad + \frac{1}{6} \delta_{3i_2 i_3 i_4}^{1234} J^3 dx^{i_2} \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{4i_2 i_3 i_4}^{1234} J^4 dx^{i_2} \wedge dx^{i_3} \wedge dx^{i_4} \\
&= \frac{1}{6} \delta_{12i_3 i_4}^{1234} J^1 dx^2 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{13i_3 i_4}^{1234} J^1 dx^3 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{14i_3 i_4}^{1234} J^1 dx^4 \wedge dx^{i_3} \wedge dx^{i_4} \\
&\quad + \frac{1}{6} \delta_{21i_3 i_4}^{1234} J^2 dx^1 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{23i_3 i_4}^{1234} J^2 dx^3 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{24i_3 i_4}^{1234} J^2 dx^4 \wedge dx^{i_3} \wedge dx^{i_4} \\
&\quad + \frac{1}{6} \delta_{31i_3 i_4}^{1234} J^3 dx^1 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{32i_3 i_4}^{1234} J^3 dx^2 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{34i_3 i_4}^{1234} J^3 dx^4 \wedge dx^{i_3} \wedge dx^{i_4} \\
&\quad + \frac{1}{6} \delta_{41i_3 i_4}^{1234} J^4 dx^1 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{42i_3 i_4}^{1234} J^4 dx^2 \wedge dx^{i_3} \wedge dx^{i_4} + \frac{1}{6} \delta_{43i_3 i_4}^{1234} J^4 dx^3 \wedge dx^{i_3} \wedge dx^{i_4} \\
&= \frac{1}{6} \delta_{1234}^{1234} J^1 dx^2 \wedge dx^3 \wedge dx^4 + \frac{1}{6} \delta_{1243}^{1234} J^1 dx^2 \wedge dx^4 \wedge dx^3 \\
&\quad + \frac{1}{6} \delta_{1324}^{1234} J^1 dx^3 \wedge dx^2 \wedge dx^4 + \frac{1}{6} \delta_{1342}^{1234} J^1 dx^3 \wedge dx^4 \wedge dx^2 \\
&\quad + \frac{1}{6} \delta_{1423}^{1234} J^1 dx^4 \wedge dx^2 \wedge dx^3 + \frac{1}{6} \delta_{1432}^{1234} J^1 dx^4 \wedge dx^3 \wedge dx^2 \\
&\quad + \frac{1}{6} \delta_{2134}^{1234} J^2 dx^1 \wedge dx^3 \wedge dx^4 + \frac{1}{6} \delta_{2143}^{1234} J^2 dx^1 \wedge dx^4 \wedge dx^3 \\
&\quad + \frac{1}{6} \delta_{2314}^{1234} J^2 dx^3 \wedge dx^1 \wedge dx^4 + \frac{1}{6} \delta_{2341}^{1234} J^2 dx^3 \wedge dx^4 \wedge dx^1 \\
&\quad + \frac{1}{6} \delta_{2413}^{1234} J^2 dx^4 \wedge dx^1 \wedge dx^3 + \frac{1}{6} \delta_{2431}^{1234} J^2 dx^4 \wedge dx^3 \wedge dx^1 \\
&\quad + \frac{1}{6} \delta_{3124}^{1234} J^3 dx^1 \wedge dx^2 \wedge dx^4 + \frac{1}{6} \delta_{3142}^{1234} J^3 dx^1 \wedge dx^4 \wedge dx^2 \\
&\quad + \frac{1}{6} \delta_{3214}^{1234} J^3 dx^2 \wedge dx^1 \wedge dx^4 + \frac{1}{6} \delta_{3241}^{1234} J^3 dx^2 \wedge dx^4 \wedge dx^1 \\
&\quad + \frac{1}{6} \delta_{3412}^{1234} J^3 dx^4 \wedge dx^1 \wedge dx^2 + \frac{1}{6} \delta_{3421}^{1234} J^3 dx^4 \wedge dx^2 \wedge dx^1 \\
&\quad + \frac{1}{6} \delta_{4132}^{1234} J^4 dx^1 \wedge dx^3 \wedge dx^2 + \frac{1}{6} \delta_{4123}^{1234} J^4 dx^1 \wedge dx^2 \wedge dx^3 \\
&\quad + \frac{1}{6} \delta_{4213}^{1234} J^4 dx^2 \wedge dx^1 \wedge dx^3 + \frac{1}{6} \delta_{4231}^{1234} J^4 dx^2 \wedge dx^3 \wedge dx^1 \\
&\quad + \frac{1}{6} \delta_{4321}^{1234} J^4 dx^3 \wedge dx^2 \wedge dx^1 + \frac{1}{6} \delta_{4312}^{1234} J^4 dx^3 \wedge dx^1 \wedge dx^2
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{3} J^1 dx^2 \wedge dx^3 \wedge dx^4 - \frac{1}{3} J^1 dx^3 \wedge dx^2 \wedge dx^4 + \frac{1}{3} J^1 dx^4 \wedge dx^2 \wedge dx^3 \\
&\quad - \frac{1}{3} J^2 dx^1 \wedge dx^3 \wedge dx^4 + \frac{1}{3} J^2 dx^3 \wedge dx^1 \wedge dx^4 - \frac{1}{3} J^2 dx^4 \wedge dx^1 \wedge dx^3 \\
&\quad + \frac{1}{3} J^3 dx^1 \wedge dx^2 \wedge dx^4 - \frac{1}{3} J^3 dx^2 \wedge dx^1 \wedge dx^4 + \frac{1}{3} J^3 dx^4 \wedge dx^1 \wedge dx^2 \\
&\quad + \frac{1}{3} J^4 dx^1 \wedge dx^3 \wedge dx^2 + \frac{1}{3} J^4 dx^2 \wedge dx^1 \wedge dx^3 + \frac{1}{3} J^4 dx^3 \wedge dx^2 \wedge dx^1 \\
&= J^1 dx^2 \wedge dx^3 \wedge dx^4 + J^2 dx^3 \wedge dx^1 \wedge dx^4 + J^3 dx^1 \wedge dx^2 \wedge dx^4 + J^4 dx^1 \wedge dx^3 \wedge dx^2,
\end{aligned}$$

thus, we obtain

$$*J = J^1 dy \wedge dz \wedge d(ict) + J^2 dz \wedge dx \wedge d(ict) + J^3 dx \wedge dy \wedge d(ict) - J^4 dx \wedge dy \wedge dz.$$

$*J$  is a 3-form, and  $d(*F)$  is also a 3-form, so we set

$$d(*F) = \mu_0(*J),$$

which yields:

$$\begin{aligned}
\frac{\partial B_3}{\partial y} - \frac{\partial B_2}{\partial z} - \frac{i}{c} \frac{\partial E_1}{\partial(ict)} &= \mu_0 J^1, & \frac{\partial B_1}{\partial z} - \frac{\partial B_3}{\partial x} - \frac{i}{c} \frac{\partial E_2}{\partial(ict)} &= \mu_0 J^2, \\
\frac{\partial B_2}{\partial x} - \frac{\partial B_1}{\partial y} - \frac{i}{c} \frac{\partial E_3}{\partial(ict)} &= \mu_0 J^3, & \frac{i}{c} \frac{\partial E_1}{\partial x} + \frac{i}{c} \frac{\partial E_2}{\partial y} + \frac{i}{c} \frac{\partial E_3}{\partial z} &= \mu_0 J^4.
\end{aligned}$$

Simplifying, we obtain:

$$\begin{aligned}
\frac{\partial B_3}{\partial y} - \frac{\partial B_2}{\partial z} &= \frac{1}{c^2} \frac{\partial E_1}{\partial t} + \mu_0 J^1 = \mu_0 \epsilon_0 \frac{\partial E_1}{\partial t} + \mu_0 J^1, \\
\frac{\partial B_1}{\partial z} - \frac{\partial B_3}{\partial x} &= \frac{1}{c^2} \frac{\partial E_2}{\partial t} + \mu_0 J^2 = \mu_0 \epsilon_0 \frac{\partial E_2}{\partial t} + \mu_0 J^2, \\
\frac{\partial B_2}{\partial x} - \frac{\partial B_1}{\partial y} &= \frac{1}{c^2} \frac{\partial E_3}{\partial t} + \mu_0 J^3 = \mu_0 \epsilon_0 \frac{\partial E_3}{\partial t} + \mu_0 J^3, \\
\frac{\partial E_1}{\partial x} + \frac{\partial E_2}{\partial y} + \frac{\partial E_3}{\partial z} &= -ic\mu_0 J^4 = -ic\mu_0(ic\rho) = c^2 \mu_0 \rho = \frac{1}{\mu_0 \epsilon_0} \mu_0 \rho = \frac{\rho}{\epsilon_0}.
\end{aligned}$$

Therefore, we obtain the other two equations of Maxwell's equations:

$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J}, \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}.$$

Equations (18.4.16) and (18.4.17) are tensor equations expressed in terms of 3-forms. Hence, these equations are form-invariant under arbitrary coordinate transformations, including Lorentz transformations. Although we derived these equations in Euclidean space, since they are expressed as tensor equations using 3-forms, they remain form-invariant in any spacetime geometry.

We assume a charged particle is located at the origin of the  $\Sigma'$  coordinate system and is at rest relative to  $\Sigma'$ , while  $\Sigma'$  moves uniformly relative to the  $\Sigma$  coordinate system. From this, we derive Maxwell's equations (18.4.11). Then, we define an electromagnetic field tensor  $F_{\mu\nu}$  and obtain the transformation relations for the electromagnetic fields based on the transformation rules for second-order tensors. Finally, we express Maxwell's equations in the tensor form of equations (18.4.16) and (18.4.17). Since tensor equations (18.4.16) and (18.4.17) remain form-invariant under arbitrary coordinate transformations, the laws described by Maxwell's equations hold unchanged in any coordinate system. Therefore, although we derived the tensor equations (18.4.16) and (18.4.17) under the special case where  $\Sigma'$  moves uniformly relative to  $\Sigma$  with velocity  $v$  along the  $x$ -axis, the conclusions we obtain, due to the nature of tensor equations, are universal. They apply not only when  $\Sigma'$  moves in any direction but also when  $\Sigma'$  is stationary and  $\Sigma$  moves uniformly or even non-uniformly relative to  $\Sigma'$ . Although we assumed the charged particle to be at rest in  $\Sigma'$ , so that  $\Sigma'$  has only an electric field and no magnetic field, the conclusions we draw, owing to the characteristics of tensor equations, also apply to situations where  $\Sigma'$  has both electric and magnetic fields.

## §18.5 Coulomb Force

According to equation (18.4.19), the magnetic field  $\mathbf{B}$  is source-free. Based on the source-free nature of  $\mathbf{B}$ , we can introduce a vector potential  $\mathbf{A}$  such that

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (18.5.1)$$

and require  $\mathbf{A}$  to satisfy the Coulomb gauge:

$$\nabla \cdot \mathbf{A} = 0.$$

Substituting equation (18.5.1) into equation (18.4.18) gives

$$\nabla \times \left( \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0.$$

This indicates that the vector  $\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}$  is irrotational, so it can be described by a scalar potential  $\varphi$ :

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \varphi.$$

Rearranging this equation yields

$$\mathbf{E} = -\nabla \varphi - \frac{\partial \mathbf{A}}{\partial t}, \quad (18.5.2)$$

which is the general expression for the electric field.

According to Example 2.2.2 in Chapter 2, the gradient is a first-order covariant tensor. A first-order covariant tensor is also called a 1-form. Since the electric field is a gradient field or a 1-form field, a charge in the electric field experiences a force. The force acting on a charge  $e$  in an electric field  $\mathbf{E}$  is defined as

$$\mathbf{F} = e\mathbf{E}, \quad (18.5.3)$$

and  $\mathbf{F}$  is called the Coulomb force.

The electric field intensity produced by a particle with charge  $Q$  is determined by equation (18.1.16). The force experienced by a particle with charge  $e$  in this electric field is

$$\mathbf{F} = e\mathbf{E} = \frac{eQ\mathbf{r}}{4\pi\epsilon_0 r^3}, \quad (18.5.4)$$

which is Coulomb's law.

## §18.6 Lorentz Force

In the coordinate system  $\Sigma'$ , the charged particle  $Q$  is at rest relative to  $\Sigma'$ , producing only an electric field  $\mathbf{E}'$ . Thus, it exerts only an electric force  $\mathbf{F}'$  on other particles with charge  $e$ , given by

$$\mathbf{F}' = e\mathbf{E}'. \quad (18.6.1)$$

However, when observed in the coordinate system  $\Sigma$ , the charged particle  $Q$ , which is at rest in  $\Sigma'$ , produces not only an electric field but also a magnetic field, exerting a magnetic force on other particles. This magnetic force can be derived by performing a coordinate transformation on equation (18.6.1). Below, we derive the Lorentz force formula.

Writing equation (18.6.1) in component form and using equation (18.4.15), we obtain:

$$\begin{cases} F'_1 = eE'_1 = eE_1, \\ F'_2 = eE'_2 = \gamma e(E_2 - vB_3), \\ F'_3 = eE'_3 = \gamma e(E_3 + vB_2). \end{cases} \quad (18.6.2)$$

However,  $\mathbf{F}$  or  $\mathbf{F}'$  is the force defined in classical mechanics—a three-dimensional force that does not possess Lorentz transformation covariance. We need to define a four-dimensional force that is covariant under Lorentz transformations.

The four-momentum vector  $p_\mu = (p_1, p_2, p_3, \frac{i}{c}w)$  is covariant under Lorentz

transformations, where  $w$  represents the particle's energy. The four-force acting on a particle moving with velocity  $\mathbf{V}$  is defined as

$$K_\mu = \frac{dp_\mu}{d\tau} = \left( \mathbf{K}, \frac{i}{c} \mathbf{K} \cdot \mathbf{V} \right), \quad (18.6.3)$$

where  $\mathbf{K} = (K_1, K_2, K_3)$ ,  $\mathbf{K} \cdot \mathbf{V} = K_1 v_1 + K_2 v_2 + K_3 v_3$ ,  $d\tau$  is the proper time, and  $\gamma d\tau = dt$ .  $K_\mu$  is a first-order tensor—a force that is covariant under Lorentz transformations.

The relationship between the force  $\mathbf{F}$  and  $\mathbf{K}$  is

$$\mathbf{F} = (F_1, F_2, F_3) = \frac{d\mathbf{p}}{dt} = \frac{\mathbf{K}}{\gamma}. \quad (18.6.4)$$

Since charge  $Q$  is at rest in  $\Sigma'$ ,  $\gamma = 1$ , so in  $\Sigma'$ ,

$$\mathbf{K}' = \mathbf{F}'.$$

Thus, the first three components of  $\mathbf{K}'$  from equation (18.6.2) are

$$\begin{cases} K'_1 = eE_1, \\ K'_2 = \gamma e(E_2 - vB_3), \\ K'_3 = \gamma e(E_3 + vB_2). \end{cases} \quad (18.6.5)$$

We now determine the first three components of  $\mathbf{K}$  in  $\Sigma$ . Since

$$a\mathbf{K} = \mathbf{K}',$$

the transformation matrix  $a$  is given by equation (18.4.14). Thus, we obtain:

$$\begin{pmatrix} \gamma & 0 & 0 & i\frac{v}{c}\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\frac{v}{c}\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} K_1 \\ K_2 \\ K_3 \\ \frac{i}{c} \mathbf{K} \cdot \mathbf{V} \end{pmatrix} = \begin{pmatrix} K'_1 \\ K'_2 \\ K'_3 \\ K'_4 \end{pmatrix}. \quad (18.6.6)$$

In  $\Sigma$ , since  $\Sigma'$  moves only along the  $x$ -axis, the velocity  $\mathbf{V}$  of charge  $Q$  has only an  $x$ -component:  $\mathbf{V} = (v, 0, 0)$ . Thus, from equation (18.6.3), we have

$$K_4 = \frac{i}{c} \mathbf{K} \cdot \mathbf{V} = \frac{i}{c} (K_1, K_2, K_3)(v, 0, 0) = \frac{i}{c} K_1 v.$$

From equation (18.6.6), we get

$$\begin{aligned} K'_1 &= \gamma K_1 - \frac{v}{c^2} \gamma \mathbf{K} \cdot \mathbf{V} = \gamma K_1 - \frac{v}{c^2} \gamma K_1 v = K_1 \gamma \left( 1 - \frac{v^2}{c^2} \right) \\ &= K_1 \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \left( 1 - \frac{v^2}{c^2} \right) = K_1 \sqrt{1 - \frac{v^2}{c^2}} = \frac{K_1}{\gamma}, \end{aligned}$$

$$K'_2 = K_2,$$

$$K'_3 = K_3.$$

Using equation (18.6.5), we find

$$K_1 = \gamma eE_1, \quad K_2 = \gamma e(E_2 - vB_3), \quad K_3 = \gamma e(E_3 + vB_2).$$

Since  $\mathbf{F} = \frac{\mathbf{K}}{\gamma}$ , we have

$$F_1 = eE_1, \quad F_2 = e(E_2 - vB_3), \quad F_3 = e(E_3 + vB_2),$$

Combining these into vector form yields

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}),$$

which is the Lorentz force formula.

## §18.7 Like Charges Repel and Opposite Charges Attract

It is well known that, macroscopically, when two objects carrying like charges (both positive or both negative) approach each other, they experience a repulsive force. Conversely, when one object carrying a positive charge and another carrying a negative charge approach each other, they experience an attractive force. This phenomenon is described as "like charges repel, opposite charges attract."

If the linear dimensions of macroscopic charged objects are much smaller than the distance over which they interact, such that the influence of their shape and volume on the interaction force can be neglected, we approximately treat the macroscopic charged objects as point charges. Below, we analyze the reason why like point charges repel and opposite point charges attract on a macroscopic scale.

Based on the preceding discussion, we can macroscopically view the electrostatic field around each point charge as a geometric structure composed of infinitely many two-dimensional concentric spherical surfaces  $S^2$ . For a positively charged point charge, the orientation of each concentric spherical surface  $S^2$  is defined by its outward normal vector direction. For a negatively charged point charge, the orientation of each concentric spherical surface  $S^2$  is defined by its inward normal vector direction.

### 1. Repulsive Force Between Positively Charged Particles

Assume point charge  $Q_1^+$  carries a positive charge. Its orientation is determined by the outward normal vector  $n_1$  of its concentric spherical surface  $S_{Q_1^+}^2$ , as shown in the left diagram of Figure 18.7.1, where the circle represents a concentric spherical surface  $S_{Q_1^+}^2$ . Let  $(U; x^1, y^1)$  be the local coordinate system on the concentric spherical surface  $S_{Q_1^+}^2$  of point charge  $Q_1^+$ . In this coordinate system, the 2-form differential  $\varphi_1 = dx^1 \wedge dy^1$  can also determine the orientation of  $S_{Q_1^+}^2$ . Together,  $dx^1, dy^1$ , and the outward normal vector  $n_1$  follow an orientation that can be chosen as the right-hand rule orientation, as illustrated in Figure 18.7.3.

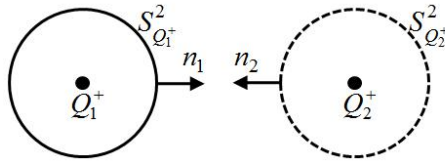


Figure 18.7.1 Before interaction

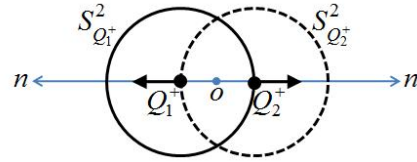


Figure 18.7.2 During interaction

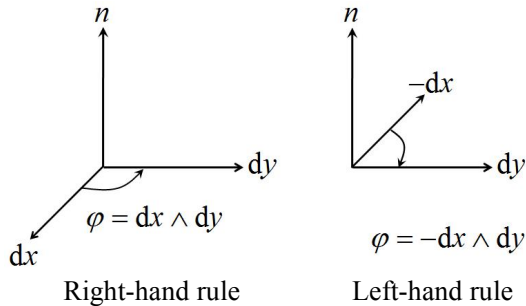


Figure 18.7.3 Right-hand and left-hand rules

determine the orientation of  $S_{Q_2^+}^2$ . Together,  $dx^2, dy^2$ , and the outward normal vector  $n_2$  also follow the right-hand rule orientation.

As shown in Figure 18.7.2, when these two positively charged point charges approach each

Similarly, point charge  $Q_2^+$  also carries a positive charge. Its orientation is determined by the outward normal vector  $n_2$  of its concentric spherical surface, as shown in the right diagram of Figure 18.7.1. Let  $(V; x^2, y^2)$  be the local coordinate system on the concentric spherical surface  $S_{Q_2^+}^2$  of point charge  $Q_2^+$ . In this coordinate system, the 2-form differential  $\varphi_2 = dx^2 \wedge dy^2$  can also

other, the concentric spherical surfaces of the left positive point charge  $Q_1^+$  intersect or are tangent to the concentric spherical surfaces of the right positive point charge  $Q_2^+$ . Due to the interaction between point charges  $Q_1^+$  and  $Q_2^+$ , each of their concentric spherical surfaces forms a product manifold  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ . Let  $(U \times V; x^1, y^1, x^2, y^2)$  be a local coordinate system on  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ . The natural projections are defined as

$$\pi_1 : S_{Q_1^+}^2 \times S_{Q_2^+}^2 \rightarrow S_{Q_1^+}^2, \quad \pi_2 : S_{Q_1^+}^2 \times S_{Q_2^+}^2 \rightarrow S_{Q_2^+}^2,$$

i.e., for  $(x^1, y^1, x^2, y^2) \in S_{Q_1^+}^2 \times S_{Q_2^+}^2$ , we have

$$\pi_1(x^1, y^1, x^2, y^2) \rightarrow (x^1, y^1), \quad \pi_2(x^1, y^1, x^2, y^2) \rightarrow (x^2, y^2).$$

The natural projection  $\pi_1$  can be expressed as a function:

$$x^1 = x^1, \quad y^1 = y^1,$$

where the left-hand side  $(x^1, y^1)$  are coordinates on  $S_{Q_1^+}^2$ , and the right-hand side  $(x^1, y^1)$  are coordinates on  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ . Since

$$\pi_1^*(dx^1 \wedge dy^1) = \pi_1^*(dx^1) \wedge \pi_1^*(dy^1) = d(\pi_1^*(x^1)) \wedge d(\pi_1^*(y^1)) = dx^1 \wedge dy^1,$$

the mapping  $\pi_1^*$  pulls back the form  $dx^1 \wedge dy^1$  on  $S_{Q_1^+}^2$  to  $dx^1 \wedge dy^1$  on  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ .

The natural projection  $\pi_2$  can be expressed as a function:

$$x^2 = x^2, \quad y^2 = y^2,$$

where the left-hand side  $(x^2, y^2)$  are coordinates on  $S_{Q_2^+}^2$ , and the right-hand side  $(x^2, y^2)$  are coordinates on  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ . Since

$$\pi_2^*(dx^2 \wedge dy^2) = \pi_2^*(dx^2) \wedge \pi_2^*(dy^2) = d(\pi_2^*(x^2)) \wedge d(\pi_2^*(y^2)) = dx^2 \wedge dy^2,$$

the mapping  $\pi_2^*$  pulls back the form  $dx^2 \wedge dy^2$  on  $S_{Q_2^+}^2$  to  $dx^2 \wedge dy^2$  on  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ .

Since

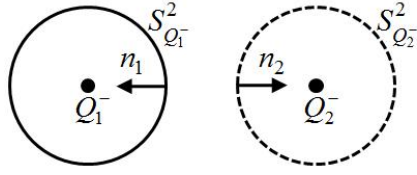
$$\pi_1^*(dx^1 \wedge dy^1) \wedge \pi_2^*(dx^2 \wedge dy^2) = dx^1 \wedge dy^1 \wedge dx^2 \wedge dy^2,$$

according to Theorem 9.5.2,  $dx^1 \wedge dy^1 \wedge dx^2 \wedge dy^2$  determines the orientation of the product manifold  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ , which is consistent with the orientations of  $S_{Q_1^+}^2$  and  $S_{Q_2^+}^2$ , also determined by the outward normal vector direction. **We take the direction of the Coulomb force as the orientation of the product manifold.** Therefore, as shown in Figure 18.7.2, the motion directions of point charges  $S_{Q_1^+}^2$  and  $S_{Q_2^+}^2$  both align with the outward normal vector  $n$  of the product manifold  $S_{Q_1^+}^2 \times S_{Q_2^+}^2$ . This causes these two positively charged point charges to repel each other and move apart.

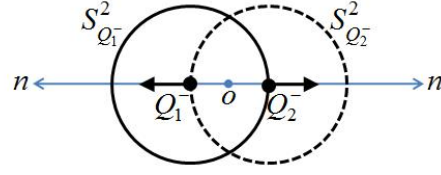
## 2. Repulsive Force Between Negatively Charged Particles

If point charge  $Q_1^-$  carries a negative charge, its orientation is determined by the inward normal vector  $n_1$  of its concentric spherical surface  $S_{Q_1^-}^2$ , as shown in the left diagram of Figure 18.7.4. Let  $(U; x^1, y^1)$  be the local coordinate system on the concentric spherical surface  $S_{Q_1^-}^2$  of point charge  $Q_1^-$ . In this coordinate system, the 2-form differential  $\varphi_1 = -dx^1 \wedge dy^1$  can also determine the orientation of  $S_{Q_1^-}^2$ . Together,  $-dx^1, dy^1$ , and the inward normal vector  $n_1$  follow the left-hand rule orientation.





**Figure 18.7.4** Before interaction



**Figure 18.7.5** During interaction

Similarly, if point charge  $Q_2^-$  also carries a negative charge, its orientation is determined by the inward normal vector  $n_2$  of its concentric spherical surface, as shown in the right diagram of Figure 18.7.4. Let  $(V; x^2, y^2)$  be the local coordinate system on the concentric spherical surface  $S^2_{Q_2^-}$  of point charge  $Q_2^-$ . In this coordinate system, the 2-form differential  $\varphi_2 = -dx^2 \wedge dy^2$  can also determine the orientation of  $S^2_{Q_2^-}$ . Together,  $-dx^2, dy^2$ , and the inward normal vector  $n_2$  also follow the left-hand rule orientation.

As shown in Figure 18.7.5, when these two negatively charged point charges approach each other, the concentric spherical surfaces of the left negative point charge  $Q_1^-$  intersect or are tangent to the concentric spherical surfaces of the right negative point charge  $Q_2^-$ . Due to the interaction between point charges  $Q_1^-$  and  $Q_2^-$ , each of their concentric spherical surfaces forms a product manifold  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ . Let  $(U \times V; x^1, y^1, x^2, y^2)$  be a local coordinate system on  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ . The natural projections are

$$\pi_1 : S^2_{Q_1^-} \times S^2_{Q_2^-} \rightarrow S^2_{Q_1^-}, \quad \pi_2 : S^2_{Q_1^-} \times S^2_{Q_2^-} \rightarrow S^2_{Q_2^-},$$

i.e., for  $(x^1, y^1, x^2, y^2) \in S^2_{Q_1^-} \times S^2_{Q_2^-}$ , we have

$$\pi_1(x^1, y^1, x^2, y^2) \rightarrow (x^1, y^1), \quad \pi_2(x^1, y^1, x^2, y^2) \rightarrow (x^2, y^2).$$

The natural projection  $\pi_1$  can be expressed as a function:

$$x^1 = x^1, \quad y^1 = y^1,$$

where the left-hand side  $(x^1, y^1)$  are coordinates on  $S^2_{Q_1^-}$ , and the right-hand side  $(x^1, y^1)$  are coordinates on  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ . Since

$$\pi_1^*(-dx^1 \wedge dy^1) = \pi_1^*(-dx^1) \wedge \pi_1^*(dy^1) = d(\pi_1^*(-x^1)) \wedge d(\pi_1^*(y^1)) = -dx^1 \wedge dy^1,$$

the mapping  $\pi_1^*$  pulls back the form  $-dx^1 \wedge dy^1$  on  $S^2_{Q_1^-}$  to  $-dx^1 \wedge dy^1$  on  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ .

The natural projection  $\pi_2$  can be expressed as a function:

$$x^2 = x^2, \quad y^2 = y^2,$$

where the left-hand side  $(x^2, y^2)$  are coordinates on  $S^2_{Q_2^-}$ , and the right-hand side  $(x^2, y^2)$  are coordinates on  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ . Since

$$\pi_2^*(-dx^2 \wedge dy^2) = \pi_2^*(-dx^2) \wedge \pi_2^*(dy^2) = d(\pi_2^*(-x^2)) \wedge d(\pi_2^*(y^2)) = -dx^2 \wedge dy^2,$$

the mapping  $\pi_2^*$  pulls back the form  $-dx^2 \wedge dy^2$  on  $S^2_{Q_2^-}$  to  $-dx^2 \wedge dy^2$  on  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ .

Since

$$\begin{aligned} \pi_1^*(-dx^1 \wedge dy^1) \wedge \pi_2^*(-dx^2 \wedge dy^2) &= (-dx^1 \wedge dy^1) \wedge (-dx^2 \wedge dy^2) \\ &= dx^1 \wedge dy^1 \wedge dx^2 \wedge dy^2, \end{aligned}$$

this 4-form  $dx^1 \wedge dy^1 \wedge dx^2 \wedge dy^2$  determines the orientation of the product manifold  $S^2_{Q_1^-} \times S^2_{Q_2^-}$ , which is consistent with the orientation of the product manifold formed by the two positively charged point charges in the previous case—namely, the right-hand rule orientation. Therefore, as

shown in Figure 18.7.5, the motion directions of point charges  $S_{Q_1^-}^2$  and  $S_{Q_2^-}^2$  both align with the outward normal vector  $n$  of the product manifold  $S_{Q_1^-}^2 \times S_{Q_2^-}^2$ . This causes these two negatively charged point charges to repel each other and move apart. This aligns with observed behavior, confirming that taking the direction of the Coulomb force as the orientation of the product manifold is correct.

### 3. Attractive Force Between Positive and Negative Charged Particles

Assume point charge  $Q_1^+$  carries a positive charge. Its orientation is determined by the outward normal vector  $n_1$  of its concentric spherical surface  $S_{Q_1^+}^2$ , as shown in the left diagram of Figure 18.7.6. Let  $(U; x^1, y^1)$  be the local coordinate system on the concentric spherical surface  $S_{Q_1^+}^2$  of point charge  $Q_1^+$ . In this coordinate system, the 2-form differential  $\varphi_1 = dx^1 \wedge dy^1$  can also determine the orientation of  $S_{Q_1^+}^2$ . Together,  $dx^1, dy^1$ , and the outward normal vector  $n_1$  follow the right-hand rule orientation.

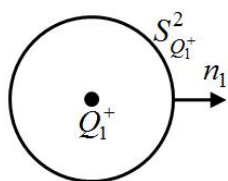


Figure 18.7.6 Before interaction

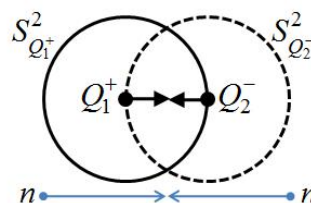
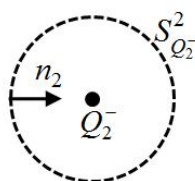


Figure 18.7.7 During interaction

Assume point charge  $Q_2^-$  carries a negative charge. Its orientation is determined by the inward normal vector  $n_2$  of its concentric spherical surface, as shown in the right diagram of Figure 18.7.6. Let  $(V; x^2, y^2)$  be the local coordinate system on the concentric spherical surface  $S_{Q_2^-}^2$  of point charge  $Q_2^-$ . In this coordinate system, the 2-form differential

$$\varphi_2 = -dx^2 \wedge dy^2 = dy^2 \wedge dx^2$$

can also determine the orientation of  $S_{Q_2^-}^2$ . Together,  $-dx^2, dy^2$ , and the outward normal vector  $n_2$  follow the left-hand rule orientation.

As shown in Figure 18.7.7, when these positively and negatively charged point charges approach each other, the concentric spherical surfaces of the left positive point charge  $Q_1^+$  intersect or are tangent to the concentric spherical surfaces of the right negative point charge  $Q_2^-$ . Due to the interaction between point charges  $Q_1^+$  and  $Q_2^-$ , each of their concentric spherical surfaces forms a product manifold  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ . Let  $(U \times V; x^1, y^1, x^2, y^2)$  be a local coordinate system on  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ . The natural projections are

$$\pi_1 : S_{Q_1^+}^2 \times S_{Q_2^-}^2 \rightarrow S_{Q_1^+}^2, \quad \pi_2 : S_{Q_1^+}^2 \times S_{Q_2^-}^2 \rightarrow S_{Q_2^-}^2,$$

i.e., for  $(x^1, y^1, x^2, y^2) \in S_{Q_1^+}^2 \times S_{Q_2^-}^2$ , we have

$$\pi_1(x^1, y^1, x^2, y^2) \rightarrow (x^1, y^1), \quad \pi_2(x^1, y^1, x^2, y^2) \rightarrow (x^2, y^2).$$

The natural projection  $\pi_1$  can be expressed as a function:

$$x^1 = x^1, \quad y^1 = y^1,$$

where the left-hand side  $(x^1, y^1)$  are coordinates on  $S_{Q_1^+}^2$ , and the right-hand side  $(x^1, y^1)$  are

coordinates on  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ . Since

$$\pi_1^*(dx^1 \wedge dy^1) = dx^1 \wedge dy^1,$$

the mapping  $\pi_1^*$  pulls back the form  $dx^1 \wedge dy^1$  on  $S_{Q_1^+}^2$  to  $dx^1 \wedge dy^1$  on  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ .

The natural projection  $\pi_2$  can be expressed as a function:

$$x^2 = x^2, \quad y^2 = y^2,$$

where the left-hand side  $(x^2, y^2)$  are coordinates on  $S_{Q_2^-}^2$ , and the right-hand side  $(x^2, y^2)$  are coordinates on  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ . Since

$$\pi_2^*(-dx^2 \wedge dy^2) = -dx^2 \wedge dy^2,$$

the mapping  $\pi_2^*$  pulls back the form  $-dx^2 \wedge dy^2$  on  $S_{Q_2^-}^2$  to  $-dx^2 \wedge dy^2$  on  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ .

Since

$$\begin{aligned} \pi_1^*(dx^1 \wedge dy^1) \wedge \pi_2^*(dx^2 \wedge dy^2) &= dx^1 \wedge dy^1 \wedge (-dx^2 \wedge dy^2) \\ &= -dx^1 \wedge dy^1 \wedge dx^2 \wedge dy^2, \end{aligned}$$

this 4-form  $-dx^1 \wedge dy^1 \wedge dx^2 \wedge dy^2$  determines the orientation of the product manifold  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ , which is opposite to the orientations of the product manifolds in the previous two cases—specifically, it follows the left-hand rule orientation. Because the direction of the Coulomb force is taken as the orientation of the product manifold, as shown in Figure 18.7.7, the motion directions of point charges  $S_{Q_1^+}^2$  and  $S_{Q_2^-}^2$  both align with the inward normal vector  $n$  of the product manifold  $S_{Q_1^+}^2 \times S_{Q_2^-}^2$ . This causes these positive and negative point charges to attract each other and move closer. This aligns with observed behavior, further confirming that taking the direction of the Coulomb force as the orientation of the product manifold is correct.

## Chapter 19 Gravitational Field

### §19.1 Characteristics of the Gravitational Field

#### 1.The Gravitational Field as a Generalized Riemannian Manifold

To compute the length of a curve on a smooth manifold, the length of a tangent vector, and the angle between two tangent vectors, it is necessary to introduce a metric tensor field on the smooth manifold. As inhabitants of Earth, we can measure the length of a road, the angle between two straight segments, and the speed of a car. This indicates that if we regard Earth and its surrounding space as a smooth manifold, there must exist a metric tensor field on this smooth manifold. According to Definition 11.1.3, if a smooth metric tensor field is given on a smooth manifold, then this smooth manifold is called a generalized Riemannian manifold.

According to Theorem 11.1.2 (the fundamental theorem of Riemannian geometry), on an  $m$ -dimensional generalized Riemannian manifold, there exists a unique torsion-free compatible connection. This connection is called the Levi-Civita connection, or Riemannian connection, of the generalized Riemannian manifold. With the Riemannian connection, one can compute the differentials of tensor fields on the smooth manifold.

If a person stands in an elevator, at the moment the elevator starts ascending, they feel heavier or as if a downward force is pulling at their feet. Conversely, if a person stands in an elevator when it begins descending, they suddenly feel lighter or as if a force is supporting them from below, making them feel as if they are about to float. Einstein believed that when a person in an elevator suddenly feels heavier or lighter, it is because the elevator experiences acceleration at the moment of starting. This acceleration enhances or diminishes the Earth's gravitational field, meaning that this acceleration is equivalent to the gravitational field strength. In particular, when the elevator is in free fall, the person experiences complete weightlessness because the acceleration fully cancels out the gravitational field strength, making them feel as if they are in a space without a gravitational field. Therefore, Einstein proposed that, locally, gravitational field strength is equivalent to acceleration. This is the equivalence principle.

If a person stands vertically still in an elevator while it ascends or descends, this process corresponds to the parallel transport of a tangent vector along a curve on a smooth manifold. The formula for parallel transport is given by equation (10.2.5):

$$A^i(p \rightarrow q) = A^i(p) - \Gamma_{jk}^i(p) A^j(p) dx^k,$$

where  $\Gamma_{jk}^i(p)$  are the connection coefficients. The reason a person in an elevator suddenly feels heavier or lighter is precisely because  $\Gamma_{jk}^i(p)$  is non-zero. Using the connection coefficients

$\Gamma_{jk}^i(p)$ , we can compute the Riemann curvature tensor  $R_{ikl}^j$  according to equation (10.3.5).

Based on the geometric interpretation of the Riemann curvature tensor  $R_{ikl}^j$  discussed in Chapter 10, a non-zero curvature tensor  $R_{ikl}^j$  indicates that the local region is curved. Therefore, Einstein proposed that a gravitational field is equivalent to curved space, or, more precisely, it is the curvature of space that gives rise to a gravitational field. This curved space is precisely the generalized Riemannian manifold.

#### 2.Bi-invariant Metric Tensor

Both the  $SO(3)$  group and the  $SU(2)$  group are Lie groups. Below, we use  $G$  to denote a Lie group (including  $SO(3)$  and  $SU(2)$ ). According to Definition 12.8.1, we introduce a bi-invariant metric tensor on these two Lie groups. The reasons are twofold:

1) If the metric tensor introduced on a Lie group is not invariant under left and right translations, then due to the Lie group being a continuous group with infinitely many group

elements, each element can induce left and right translation transformations. Under the infinitely many left and right translations by these group elements, the metric tensor would undergo infinitely many variations. The physical effects described by such a metric tensor would be immeasurable and thus meaningless.

2) After introducing a bi-invariant metric tensor, according to equation (12.8.1), for any left-invariant (or right-invariant) vector field  $\hat{X}$  on the Lie group, we have

$$D_{\hat{X}}\hat{X} = 0,$$

which indicates that the integral curve  $C$  of any left-invariant (or right-invariant) vector field  $\hat{X}$  is a geodesic. That is, the transport of  $\hat{X}$  along the geodesic  $C$  is parallel transport. Furthermore, any left-invariant (or right-invariant) tangent vector field  $\hat{X}$  remains invariant under left (or right) translations. Therefore, the transport of  $\hat{X}$  along a geodesic is consistent with its left (or right) translation along the geodesic—both remain unchanged.

Finally, it should be noted that, according to Theorem 12.8.2, every compact connected Lie group admits a bi-invariant metric tensor. Since  $SO(3)$  and  $SU(2)$  are both compact connected Lie groups, they each possess a bi-invariant metric tensor. For example,  $S^3$  as a submanifold of  $R^4$  inherits an induced metric tensor from  $R^4$ , and this metric tensor is bi-invariant. Therefore, elementary particles and particles or matter composed of elementary particles all possess a bi-invariant metric tensor.

### 3. Sectional Curvature

A basis for the Lie algebra of the  $SU(2)$  group can be chosen as

$$\hat{X}'_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \hat{X}'_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \hat{X}'_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Their commutation relations, as shown in equation (12.10.14), are

$$[\hat{X}'_1, \hat{X}'_2] = 2\hat{X}'_3, \quad [\hat{X}'_2, \hat{X}'_3] = 2\hat{X}'_1, \quad [\hat{X}'_3, \hat{X}'_1] = 2\hat{X}'_2.$$

We may regard  $\hat{X}'_1$ ,  $\hat{X}'_2$ , and  $\hat{X}'_3$  as orthonormal tangent vectors. Then the sectional curvature  $K(\hat{X}'_1, \hat{X}'_2)$  along the two-dimensional section determined by  $\hat{X}'_1$  and  $\hat{X}'_2$  is given by equation (12.8.5):

$$K(\hat{X}'_1, \hat{X}'_2) = \frac{1}{4} |[\hat{X}'_1, \hat{X}'_2]|^2 = \frac{1}{4} |2\hat{X}'_3|^2 = |\hat{X}'_3|^2 = 1.$$

Similarly, we obtain

$$K(\hat{X}'_2, \hat{X}'_3) = K(\hat{X}'_3, \hat{X}'_1) = 1.$$

Another basis for the Lie algebra of  $SU(2)$  can be chosen as:

$$\hat{X}_1 = \frac{1}{2} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \hat{X}_2 = \frac{1}{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \hat{X}_3 = \frac{1}{2} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Their commutation relations, as shown in equation (12.10.11), are:

$$[\hat{X}_1, \hat{X}_2] = \hat{X}_3, \quad [\hat{X}_2, \hat{X}_3] = \hat{X}_1, \quad [\hat{X}_3, \hat{X}_1] = \hat{X}_2.$$

$\hat{X}_1$ ,  $\hat{X}_2$ , and  $\hat{X}_3$  can be viewed as linear combinations of  $\hat{X}'_1$ ,  $\hat{X}'_2$ , and  $\hat{X}'_3$ :

$$\hat{X}_1 = \frac{1}{2} \hat{X}'_1, \quad \hat{X}_2 = \frac{1}{2} \hat{X}'_2, \quad \hat{X}_3 = \frac{1}{2} \hat{X}'_3,$$

Therefore, according to the definition of sectional curvature, the sectional curvature along the two-dimensional section determined by  $\hat{X}_i$  and  $\hat{X}_j$  ( $i \neq j; i, j = 1, 2, 3$ ) is equal to that along the section determined by  $\hat{X}'_i$  and  $\hat{X}'_j$ , both equal to 1.

Yet another basis for the Lie algebra of  $SU(2)$  can be chosen as

$$\hat{S}_1 = -\frac{1}{2} i\hbar \hat{X}'_1 = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_2 = \frac{1}{2} i\hbar \hat{X}'_2 = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{S}_3 = -\frac{1}{2} i\hbar \hat{X}'_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Their commutation relations, as shown in equation (16.11.3), are

$$[\hat{S}_1, \hat{S}_2] = i\hbar\hat{S}_3, \quad [\hat{S}_2, \hat{S}_3] = i\hbar\hat{S}_1, \quad [\hat{S}_3, \hat{S}_1] = i\hbar\hat{S}_2.$$

$\hat{S}_1$ ,  $\hat{S}_2$ , and  $\hat{S}_3$  can also be regarded as linear combinations of  $X'_1$ ,  $X'_2$ , and  $X'_3$ . Hence, according to the definition of sectional curvature, the sectional curvature along the two-dimensional section determined by  $\hat{S}_i$  and  $\hat{S}_j$  ( $i \neq j; i, j = 1, 2, 3$ ) is equal to that along the section determined by  $X'_i$  and  $X'_j$ , both equal to 1.

Since  $\hat{X}'_1$ ,  $\hat{X}'_2$ , and  $\hat{X}'_3$  are left-invariant (or right-invariant) tangent vector fields, we have everywhere on  $SU(2)$ :

$$K(\hat{X}'_1, \hat{X}'_2) = K(\hat{X}'_2, \hat{X}'_3) = K(\hat{X}'_3, \hat{X}'_1) = 1.$$

Therefore, as an  $SU(2)$  group, the elementary particle is a constant-curvature manifold with sectional curvature 1, implying it is curved and possesses a gravitational field. Matter composed of such particles naturally also exhibits a gravitational field.

In the space  $R^4$ , the unit sphere  $S^3$  is a constant-curvature manifold with sectional curvature 1. Its equation is

$$x^2 + y^2 + z^2 + w^2 = 1,$$

where  $x, y, z, w$  are spatial coordinates. If we set  $w = ict$ , then the space  $R^4$  transforms into Minkowski spacetime  $M^4$ , and the equation for  $S^3$  becomes

$$x^2 + y^2 + z^2 - c^2t^2 = 1.$$

Thus, the unit sphere  $S^3$  transforms into a three-dimensional hyperboloid of one sheet.

The three-dimensional hyperboloid of one sheet is non-compact. According to Theorem 11.9.2, it can be endowed with a Riemannian metric of positive or negative sectional curvature. However, in three-dimensional space  $R^3$ , the sectional curvature at each point on a two-dimensional hyperboloid of one sheet is negative,  $K < 0$ . The geometry of Einstein's rotating disk corresponds to a geometry with negative sectional curvature.

In Chapter 15, however, we interpreted the unit sphere  $S^3$  as an expanding geometric object in Minkowski spacetime  $M^4$  over time—this geometric object is not a three-dimensional hyperboloid of one sheet.

A similar analysis can be performed for the  $SO(3)$  group.

#### 4. Gravitational Fields of Particles and Antiparticles Are Identical

Consider the Euclidean plane  $R^2$  with a counterclockwise orientation. Let  $\mathbf{a}$  and  $\mathbf{b}$  be two vectors in  $R^2$  with an included angle  $\theta$ . Their inner product (dot product) is defined as

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta.$$

If the orientation of the plane  $R^2$  is changed to clockwise, the included angle between  $\mathbf{a}$  and  $\mathbf{b}$  becomes  $-\theta$ , and their inner product is

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos(-\theta) = |\mathbf{a}| |\mathbf{b}| \cos \theta.$$

The result shows that the inner product of vectors  $\mathbf{a}$  and  $\mathbf{b}$  is independent of the orientation of the plane  $R^2$ .

Each point of a smooth manifold  $M$  has a tangent space. Similar to defining an inner product in  $R^2$ , we define an inner product in the tangent space  $T_p M$  at each point  $p$  of the smooth manifold  $M$ . Specifically, for any  $\hat{X}, \hat{Y} \in T_p M$ , we set

$$\hat{X} \cdot \hat{Y} = g(\hat{X}, \hat{Y}) = g_{ij}(p) X^i Y^j. \quad (19.1.1)$$

Once an inner product is defined on the smooth manifold  $M$ , it becomes a generalized Riemannian manifold  $(M, g)$ , and  $g(\hat{X}, \hat{Y})$  becomes the metric tensor field on  $M$ .

The inner product is symmetric, i.e.,

$$\hat{X} \cdot \hat{Y} = g(\hat{X}, \hat{Y}) = \hat{Y} \cdot \hat{X} = g(\hat{Y}, \hat{X}). \quad (19.1.2)$$

In the definition of the inner product, the orientation of the smooth manifold is not considered because the inner product is symmetric and independent of orientation. That is, the metric tensor

field is independent of orientation. Therefore, for two diffeomorphic connected smooth manifolds, even though they are assigned different orientations, they can still have the same inner product and the same metric tensor field.

The bi-invariant metric tensor defined on a Lie group  $G$  is also an inner product defined in the tangent space  $T_p G$  at each point  $p$  of the Lie group. Naturally, it is independent of the orientation of the Lie group. Consequently, particles and antiparticles produce identical gravitational fields.

### 5. Conditions for Spatial Symmetry

Both the  $SO(3)$  group and the  $SU(2)$  group are compact connected Lie groups. According to Theorem 12.8.5, every compact connected Lie group is a Riemannian symmetric space. Therefore, elementary particles are all Riemannian symmetric spaces. Since elementary particles combine via inner direct products to form other particles, these other particles are also isomorphic to either the  $SO(3)$  group or the  $SU(2)$  group. Consequently, particles (and matter) composed of elementary particles are also Riemannian symmetric spaces. According to Theorem 11.5.1, a necessary and sufficient condition for a Riemannian manifold to be a Riemannian symmetric space is

$$R_{\mu\lambda\sigma}^{\rho}{}_{;\alpha} = 0. \quad (19.1.3)$$

Since

$$R_{\mu\nu\lambda\sigma} = R_{\mu\lambda\sigma}^{\rho} g_{\rho\nu}, \quad g_{\mu\nu}{}_{;\alpha} = 0,$$

it follows that

$$R_{\mu\nu\lambda\sigma}{}_{;\alpha} = 0. \quad (19.1.4)$$

Moreover, because the Ricci tensor is given by

$$R_{\mu\nu} = R_{\mu}^{\lambda}{}_{\lambda\nu} = g^{\lambda\sigma} R_{\mu\sigma\lambda\nu},$$

we have

$$R_{\mu\nu}{}_{;\alpha} = 0. \quad (19.1.5)$$

Conversely, if  $R_{\mu\nu}{}_{;\alpha} = 0$ , then

$$R_{\mu\nu}{}_{;\alpha} = R_{\mu}^{\lambda}{}_{\lambda\nu}{}_{;\alpha} = (g^{\lambda\sigma} R_{\mu\sigma\lambda\nu})_{;\alpha} = g^{\lambda\sigma} R_{\mu\sigma\lambda\nu}{}_{;\alpha} = 0,$$

and therefore,

$$R_{\mu\sigma\lambda\nu}{}_{;\alpha} = 0.$$

Furthermore, since the scalar curvature is

$$R = g^{\mu\nu} R_{\mu\nu},$$

we obtain

$$R_{;\alpha} = R_{,\alpha} = 0, \quad (19.1.6)$$

meaning the scalar curvature  $R$  of the Lie group is a constant. However, for a system of particles composed of multiple particles, if the number of particles changes continuously, the scalar curvature  $R$  of the particle system is not constant. Alternatively, if the Lie group is acted upon by a non-isometric transformation group causing  $R_{\mu\sigma\lambda\nu}{}_{;\alpha} \neq 0$ , then the scalar curvature  $R$  of the Lie group may not necessarily be constant.

## §19.2 Equations of Motion in a Gravitational Field

Let  $G_1$  and  $G_2$  be two three-dimensional compact connected Lie groups. Their inner direct product group is  $G_1 \otimes G_2$ . The dimension of the inner direct product group  $G_1 \otimes G_2$  equals the dimension of  $G_1$  and  $G_2$ . We use the inner direct product of two Lie groups to describe a composite particle consisting of two particles. Let  $\hat{X} \in G_1$  be a left-invariant tangent vector field on  $G_1$  and  $\hat{Y} \in G_2$  a left-invariant tangent vector field on  $G_2$ , then  $(\hat{X} + \hat{Y}) \in G_1 \otimes G_2$  is a



left-invariant tangent vector field on  $G_1 \otimes G_2$ . According to equation (12.8.1), we have

$$D_{\hat{X}+\hat{Y}}(\hat{X}+\hat{Y})=0. \quad (19.2.1)$$

Expanding this gives

$$D_{\hat{X}}\hat{X}+D_{\hat{X}}\hat{Y}+D_{\hat{Y}}\hat{X}+D_{\hat{Y}}\hat{Y}=0.$$

Further, from equation (12.8.1), we obtain

$$D_{\hat{X}}\hat{X}=0, \quad (19.2.2)$$

$$D_{\hat{Y}}\hat{Y}=0. \quad (19.2.3)$$

Additionally, from equation (12.12.5), we have

$$D_{\hat{X}}\hat{Y}=0. \quad (19.2.4)$$

Equation (19.2.1) is the geodesic equation for the inner direct product Lie group  $G_1 \otimes G_2$ , which can be used to describe the motion of the point-like particles corresponding to these two Lie groups. Equation (19.2.2) is the geodesic equation for  $G_1$ , and equation (19.2.3) is the geodesic equation for  $G_2$ . Equation (19.2.4) indicates that the left-invariant tangent vector field  $\hat{Y}$  of  $G_2$  undergoes parallel transport along the integral curves of the left-invariant tangent vector field  $\hat{X}$  of  $G_1$ . Since the integral curves of  $\hat{X}$  are also geodesics of  $G_1$ ,  $\hat{Y}$  moves along the geodesics of  $G_1$ .

When  $G_1$  and  $G_2$  form the inner direct product group  $G_1 \otimes G_2$ , suppose point  $p \in G_1$  and point  $q \in G_2$  combine to form a point  $(p,q) \in G_1 \otimes G_2$ . The tangent space  $T_{(p,q)}G_1 \otimes G_2$  at this point  $(p,q)$  is the sum of the tangent spaces  $T_pG_1$  and  $T_qG_2$ . Therefore, at point  $(p,q)$ , the tangent vectors in the tangent spaces  $T_pG_1$  and  $T_qG_2$  can be linearly represented by each other.

The tangent spaces  $T_pG_1$ ,  $T_qG_2$ , and  $T_{(p,q)}G_1 \otimes G_2$  are all three-dimensional. According to Theorem 7.1.1, they are isomorphic. Let  $(e_1, e_2, e_3)$  be a basis of  $T_{(p,q)}G_1 \otimes G_2$ , and let  $(X_1, X_2, X_3)$  be three linearly independent tangent vectors in  $T_pG_1$ . Then

$$X = AE, \quad (19.2.5)$$

where

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad E = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}.$$

Similarly, let  $(Y_1, Y_2, Y_3)$  be three linearly independent tangent vectors in  $T_qG_2$ . Then

$$Y = BE, \quad (19.2.6)$$

where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

Both  $A$  and  $B$  are invertible matrices. From equation (19.2.5), we have

$$A^{-1}X = E,$$

Substituting into equation (19.2.6) yields

$$Y = BA^{-1}X.$$

Thus, tangent vectors in  $T_qG_2$  can be linearly represented by tangent vectors in  $T_pG_1$ . Consequently, in the inner direct product Lie group  $G_1 \otimes G_2$ , the left-invariant (or right-invariant) tangent vector field  $\hat{Y}$  of  $G_2$  can be linearly expressed by the three linearly independent left-invariant (or right-invariant) tangent vector fields  $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$  of  $G_1$ ,



$$\hat{Y} = c_1 \hat{X}_1 + c_2 \hat{X}_2 + c_3 \hat{X}_3,$$

so  $\hat{Y}$  becomes a left-invariant tangent vector field on  $G_1$ . Therefore, according to equation (12.8.1), on  $G_1$ , the following holds:

$$D_{\hat{Y}} \hat{Y} = 0. \quad (19.2.7)$$

Thus, macroscopically, the trajectory of a point-like particle of  $G_2$  moving in the gravitational field of  $G_1$  with a definite macroscopic orbit is a geodesic of  $G_1$ , and its equation is given by (19.2.7). Equation (19.2.7) can be written as

$$\frac{Du^\mu}{d\tau} = 0, \quad (19.2.8)$$

where

$$d\tau = \frac{ids}{c},$$

$\tau$  is proper time,  $i$  is the imaginary unit,  $ds$  is the differential of arc length, and  $c$  is the speed of light.

If  $\hat{X}$  is a smooth tangent vector field on  $G_1$  or a general smooth manifold  $M_1$ , and  $\hat{Y}$  is a smooth tangent vector field on  $G_2$  or a general smooth manifold  $M_2$ , then according to equation (11.10.5), equation (19.2.4) also holds. Therefore, equations (19.2.7) and (19.2.8) similarly remain valid.

On a microscopic scale, a point-like particle requires specific conditions to move continuously along a geodesic with a definite trajectory. These conditions can be found in the content of Chapter 16.

### §19.3 Curvature of Composite Particles

After introducing a bi-invariant metric tensor on the  $SO(3)$  and  $SU(2)$  groups, according to Theorem 12.8.4, the Ricci curvature  $Ric(\hat{X})$  of these two Lie groups has a strict positive lower bound  $a$ . Since electrons, protons, neutrinos, and their antiparticles are all  $SU(2)$  groups, and photons are  $SO(3)$  groups, the Ricci curvature  $Ric(\hat{X})$  of these elementary particles—electrons, protons, neutrinos, photons, and their antiparticles—also has a strict positive lower bound.

We have also pointed out that elementary particles combine via inner direct products to form other particles. The inner direct product group  $G_1 \otimes G_2$  is a subgroup of the direct product group  $G_1 \times G_2$ . According to equation (12.12.11), the Ricci curvature of this inner direct product group  $G_1 \otimes G_2$  at a point  $p$  is the sum of the Ricci curvatures of the two Lie groups forming the inner direct product at that point. Let the Ricci curvature of  $G_1$  at point  $p$  be  $Ric(\hat{X}')|_p$ , the Ricci curvature of  $G_2$  at point  $q$  be  $Ric(\hat{X}'')|_q$ , and the Ricci curvature of the inner direct product Lie group  $G_1 \otimes G_2$  at point  $(p, q)$  be  $Ric(X)|_{(p,q)}$ . Then

$$Ric(\hat{X})|_{(p,q)} = Ric(\hat{X}')|_p + Ric(\hat{X}'')|_q. \quad (19.3.1)$$

Since the Lie groups representing particles are all compact, the Ricci curvature of each compact Lie group is positive. Therefore, the Ricci curvature  $Ric(v)$  of particles or matter composed of these elementary particles also has a strict positive lower bound. Moreover, matter composed of more elementary particles has a larger Ricci curvature  $Ric(v)$ .

Similarly, let the scalar curvature of  $G_1$  at point  $p$  be  $R_1$ , the Ricci curvature of  $G_2$  at point  $q$  be  $R_2$ , and the scalar curvature of the inner direct product group  $G_1 \otimes G_2$  at point  $(p, q)$  be  $R$ . Then

$$R = R_1 + R_2. \quad (19.3.2)$$

Since the Ricci curvature of the compact Lie groups representing particles is positive, their scalar

curvature is also positive. Thus, the scalar curvature of particles or matter composed of these elementary particles is positive, and matter composed of more elementary particles has a larger scalar curvature. Therefore, consider a particle system composed of many elementary particles. Let  $R$  denote the scalar curvature of this system, and let the trace  $T$  of the energy-momentum tensor  $T^{\mu\nu}$  represent the density of the number of particles in this system:

$$T = g_{\mu\nu} T^{\mu\nu} = T^\mu_\mu,$$

then  $R$  and  $T$  are linearly related. This linear relationship can be assumed as

$$R = -\kappa T, \quad (19.3.3)$$

where  $\kappa$  is a constant. The negative sign in front of  $\kappa$  is for convenience in later derivations. Based on this assumption, Einstein's gravitational field equations will be derived later.

## §19.4 Time as a One-Parameter Transformation Group

Let  $t$  denote time. According to Definition 6.6.1, we introduce a one-parameter differentiable transformation on a Lie group  $G$ , i.e., a smooth mapping  $\varphi: R \times G \rightarrow G$ , where  $(t, p)$  is an arbitrary point in  $R \times G$ , so  $(t, p) \in R \times G$ .

Denote  $\varphi_t(p) = \varphi(t, p)$ . When  $t = 0$ ,  $\varphi_0(p) = \varphi(0, p) = p$ , and  $\varphi_0(p_1) = \varphi(0, p_1) = p_1$ , and so on. Thus,  $\varphi_0(p)$  is the identity map.  $\varphi_0(p)$  represents the state of  $G$  at the start of observation, i.e., at time  $t = 0$ . When  $t = t_1$ ,  $\varphi_{t_1}(p) = \varphi(t_1, p)$ .  $\varphi(t_1, p)$  represents the state of  $G$  at time  $t = t_1$ . When  $t = t_2$ ,  $\varphi_{t_2}(p) = \varphi(t_2, p)$ .  $\varphi(t_2, p)$  represents the state of  $G$  at time  $t = t_2$ . After a time interval  $t = t_1 + t_2$ , the state of  $G$  becomes  $\varphi_{t_2} \circ \varphi_{t_1} = \varphi_{t_2+t_1} = \varphi(t_2 + t_1, p)$ . When  $t = -t_1$ ,  $\varphi_{-t_1}(p) = \varphi(-t_1, p) = (\varphi_{t_1})^{-1}$ . The inverse transformation  $\varphi(-t_1, p)$  denotes the state of  $G$  when tracing back to time  $t = -t_1$ . We call  $\varphi_t$  a one-parameter differentiable transformation group acting on  $G$ , or say  $R$  acts smoothly on  $G$  from the left. For any fixed  $t$ ,  $\varphi_t$  is a diffeomorphism from  $G$  to itself.

For example, let  $x, y, z, w$  be an arbitrary real number, and set  $a = w + iz$ ,  $b = -y + ix$ ,  $i$  be the imaginary unit, and define

$$g(x, y, z, w) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} = \begin{pmatrix} w + iz & -y + ix \\ y + ix & w - iz \end{pmatrix}. \quad (19.4.1)$$

Denote the set of all matrices  $g(x, y, z, w)$  as  $M(4, R^4)$ . At this stage,  $M(4, R^4)$  is not yet a Lie group. Referring to equation (12.9.5), compute the infinitesimal generators of  $g(x, y, z, w)$  at the identity element  $E$  ( $x = y = z = 0, w = 1$ ):

$$\hat{X}_1 = \frac{\partial g}{\partial x} \Big|_{\substack{x=y=z=0 \\ w=1}} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad (19.4.2)$$

$$\hat{X}_2 = \frac{\partial g}{\partial y} \Big|_{\substack{x=y=z=0 \\ w=1}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad (19.4.3)$$

$$\hat{X}_3 = \frac{\partial g}{\partial z} \Big|_{\substack{x=y=z=0 \\ w=1}} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad (19.4.4)$$

$$\hat{X}_4 = \frac{\partial g}{\partial w} \Big|_{\substack{x=y=z=0 \\ w=1}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (19.4.5)$$

Denote the set of all invertible matrices  $g(x, y, z, w)$  as  $GL(4, R^4)$ , i.e.,

$$GL(4, R^4) = \{Q \in M(4, R^4) ; \det Q \neq 0\}. \quad (19.4.6)$$

$GL(4, R^4)$  is a 4-dimensional real Lie group. Its Lie algebra is spanned by  $\hat{X}_1, \hat{X}_2, \hat{X}_3$ , and  $\hat{X}_4$ .

Thus,  $GL(4, R^4)$  can be generated by  $\hat{X}_1, \hat{X}_2, \hat{X}_3$ , and  $\hat{X}_4$ :

$$GL(4, R^4) = \exp[\hat{X}_1\alpha + \hat{X}_2\beta + \hat{X}_3\gamma + \hat{X}_4s], \quad (19.4.7)$$

where  $s \in (-\infty, +\infty) = R$ . Since  $\hat{X}_4$  is the identity matrix and commutes with  $\hat{X}_1, \hat{X}_2$ , and  $\hat{X}_3$ , i.e.,

$$[\hat{X}_1, \hat{X}_4] = 0, \quad [\hat{X}_2, \hat{X}_4] = 0, \quad [\hat{X}_3, \hat{X}_4] = 0,$$

equation (19.4.7) can be rewritten as

$$GL(4, R^4) = \exp[\hat{X}_4s] \exp[\hat{X}_1\alpha + \hat{X}_2\beta + \hat{X}_3\gamma]. \quad (19.4.8)$$

If we set  $s = ict$ , equation (19.4.8) becomes

$$GL(4, R^4) = \exp[ic\hat{X}_4t] \exp[\hat{X}_1\alpha + \hat{X}_2\beta + \hat{X}_3\gamma]. \quad (19.4.9)$$

Define a subset of  $GL(4, R^4)$ :

$$SU(2) = \{Q \in GL(4, R^4) ; Q^+ = Q^{-1}, |Q| = 1\}, \quad (19.4.10)$$

where  $+$  denotes the conjugate transpose (Hermitian conjugate). The set  $SU(2)$  is a group and a Lie subgroup of the general linear group  $GL(4, R^4)$ . It is well known that  $SU(2)$  is homeomorphic to the three-dimensional unit sphere  $S^3$  of radius 1.

We know that  $SU(2)$  can be generated by  $\hat{X}_1, \hat{X}_2$ , and  $\hat{X}_3$ :

$$SU(2) = \exp[\hat{X}_1\theta + \hat{X}_2\varphi + \hat{X}_3\omega]. \quad (19.4.11)$$

Since  $SU(2)$  is a Lie subgroup of  $GL(4, R^4)$ , from equation (19.4.8), we obtain

$$GL(4, R^4)' = \exp[\hat{X}_4s] \exp[\hat{X}_1\theta + \hat{X}_2\varphi + \hat{X}_3\omega], \quad (19.4.12)$$

or from equation (19.4.9), we obtain

$$GL(4, R^4)' = \exp[ic\hat{X}_4t] \exp[\hat{X}_1\theta + \hat{X}_2\varphi + \hat{X}_3\omega]. \quad (19.4.13)$$

The one-parameter subgroup  $\exp[\hat{X}_4s]$  or  $\exp[ic\hat{X}_4t]$  generated by  $\hat{X}_4$  is isomorphic to  $R$ , so it corresponds to  $R$ . The product of  $\exp[ic\hat{X}_4t]$  and  $\exp[\hat{X}_1\theta + \hat{X}_2\varphi + \hat{X}_3\omega]$  corresponds to the previously mentioned  $R \times G$ .

Therefore, we can view time as a curve  $M_1$ , and a particle as a three-dimensional smooth manifold  $M_2$ . Their product forms a four-dimensional product manifold  $M_1 \times M_2$ . Referring to equation (11.10.1), we can introduce a metric  $g = g_1 \times g_2$  on the product manifold  $M_1 \times M_2$  such that for any  $(p, q) \in M_1 \times M_2$ ,

$$g(X, Y) = g_1(X_1, X_1) + g_2(X_2, Y_2),$$

where

$$X_1 \in T_p M_1, \quad X_2, Y_2 \in T_q M_2, \quad X, Y \in T_{(p,q)} M_1 \times M_2.$$

In the local coordinates  $(U; ict, x^1, x^2, x^3)$  of the product manifold  $M_1 \times M_2$ , the general form of the metric is

$$ds^2 = A(t, x^1, x^2, x^3) dt^2 + B_{ij}(t, x^1, x^2, x^3) dx^i dx^j, \quad i, j = 1, 2, 3.$$

This way of describing the gravitational field is independent of Minkowski spacetime.

Due to the limiting nature of the speed of light and the requirement that measurable quantities (e.g., distance, time) must be real, the following restrictions are imposed on the metric tensor  $g_{\mu\nu}$  describing the gravitational field of particles.

**Theorem 19.5.1<sup>4</sup>** For the coordinate transformation  $\tilde{x}^\mu = \tilde{x}^\mu(\nu)$  to yield four new coordinate variables  $(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)$  where the first represents the time coordinate and the last three represent spatial coordinates, the new metric tensor  $g_{\mu\nu}$  must satisfy the following necessary and sufficient conditions:

<sup>4</sup>Zhao Zheng, Liu Wenbiao. *Foundations of General Relativity* [M]. Beijing: Tsinghua University Press, 2010:47~48.

$$g_{00} < 0, \quad \begin{vmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{vmatrix} < 0, \quad \begin{vmatrix} g_{00} & g_{01} & g_{02} \\ g_{10} & g_{11} & g_{12} \\ g_{20} & g_{21} & g_{22} \end{vmatrix} < 0, \quad \begin{vmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33} \end{vmatrix} < 0.$$

## §19.5 Describing the Gravitational Field in Minkowski Spacetime

If particles are embedded in Minkowski spacetime, the gravitational field of particles can also be described within Minkowski spacetime.

Let  $R_+$  denote the set of positive real numbers, i.e.,  $R_+ = [0, +\infty)$ , and let  $S^m$  denote the unit sphere in  $R^{m+1}$ . Then there exists a one-to-one correspondence:

$$f: R^{m+1} - \{0\} \rightarrow R_+ \times S^m,$$

defined by

$$f(x^1, \dots, x^{m+1}) = \left( \|x\|, \left( \frac{x^1}{\|x\|}, \dots, \frac{x^{m+1}}{\|x\|} \right) \right), \quad (19.5.1)$$

Here  $\|x\| = \sqrt{(x^1)^2 + \dots + (x^{m+1})^2}$ .  $R_+$  and  $S^m$  are subspaces of the topological spaces  $R^1$  and  $R^{m+1}$ , respectively. We equip  $R^{m+1} - \{0\}$  with the induced topology from  $R^{m+1}$ , and  $R_+ \times S^m$  with the product topology. Then  $f$  is a homeomorphism. Clearly, both  $R^{m+1} - \{0\}$  and  $R_+ \times S^m$  are smooth manifolds. From equation (19.5.1), it can be seen that the mapping  $f$  is smooth, so  $f$  is a diffeomorphism.

When  $m=3$ , both  $R^4 - \{0\}$  and  $R_+ \times S^3$  are smooth manifolds, and the mapping  $f$  is also a diffeomorphism. The behavior of spin-1/2 elementary particles, which are three-dimensional unit spheres  $S^3$ , over time can be described by the smooth manifold  $R_+ \times S^3$ , where  $R_+$  is identified as the time interval. We define a metric tensor  $g$  on the smooth manifold  $R_+ \times S^3$ , making it a generalized Riemannian manifold. The pullback mapping  $f^*$  induces a second-order covariant tensor  $h = f^*g$  as the metric tensor on the smooth manifold  $R^4 - \{0\}$ , thus making  $R^4 - \{0\}$  a generalized Riemannian manifold as well. Since  $f$  is a diffeomorphism,  $f: R^4 - \{0\} \rightarrow R_+ \times S^3$  is an isometry, meaning the two generalized Riemannian manifolds  $R^4 - \{0\}$  and  $R_+ \times S^3$  are isometric to each other. Therefore, we can use the metric tensor, curvature tensor, and other tensors on  $R^4 - \{0\}$  to describe the gravitational field produced by spin-1/2 elementary particles.

We have previously discussed that a photon is an  $RP^3$  space, which can be described using a real projective model in  $R^4$  (see Figure 15.2.1). Define an equivalence relation  $\sim$  on  $R^4 - \{0\}$  as follows:

$$\forall (x^1, \dots, x^4), (y^1, \dots, y^4) \in R^4 - \{0\},$$

$x \sim y$  if and only if there exists a nonzero real number  $\lambda$  such that  $y^\alpha = \lambda x^\alpha$  ( $1 \leq \alpha \leq 4$ ). After defining this equivalence relation  $\sim$ ,  $RP^3$  is identified as the quotient space of  $R^4 - \{0\}$  under this equivalence relation:

$$RP^3 = (R^4 - \{0\}) / \sim.$$

Let  $[x] = [(x^1, \dots, x^4)]$  denote the equivalence class of  $x = (x^1, \dots, x^4) \in R^4 - \{0\}$ . Then

$$RP^3 = \{[x] = [(x^1, \dots, x^4)] \mid (x^1, \dots, x^4) \in R^4 - \{0\}\}.$$

Thus, we can also use the metric tensor, curvature tensor, and other tensors on  $R^4 - \{0\}$  to

describe the gravitational field produced by photons.

Since we can use the metric tensor, curvature tensor, and other tensors on  $R^4 - \{0\}$  to describe the gravitational fields produced by spin-1/2 elementary particles and photons, we can similarly use these tensors on  $R^4 - \{0\}$  to describe the gravitational field produced by composite particles formed from spin-1/2 elementary particles and photons.

To align with reality, we should transform the space  $R^4$  into spacetime  $M^4$ . This is done by establishing coordinates  $(x^0 = ict, x^1, x^2, x^3)$  in  $R^4$ , where  $t, x^1, x^2, x^3$  are real numbers,  $i$  is the imaginary unit, and  $c$  is the speed of light. Since the origin in  $R^4$  does not belong to the solution of the gravitational field equations, naturally, the origin in  $M^4$  also does not belong to the solution of the gravitational field equations.

It is also possible to establish a purely real coordinate system in  $R^4$ , turning it into spacetime  $M^4$ . The method is as follows: In the local region of the gravitational field, establish a local inertial frame using Cartesian coordinates:

$$x^\mu = (x^0, x^1, x^2, x^3), \quad x^0 = ct,$$

where  $t$  is the time coordinate in the local inertial frame. The metric form is

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad \mu, \nu = 0, 1, 2, 3,$$

where

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

If the Jacobian determinant of the coordinate transformation  $\tilde{x}^\mu = \tilde{x}^\mu(\nu)$  satisfies

$$\det\left(\frac{\partial \tilde{x}^\mu}{\partial x^\nu}\right) \neq 0,$$

then the transformed metric tensor is obtained:

$$g_{\mu\nu} = \eta_{\alpha\beta} \frac{\partial x^\alpha}{\partial \tilde{x}^\mu} \frac{\partial x^\beta}{\partial \tilde{x}^\nu}.$$

Here,  $g_{\mu\nu}$  becomes the metric tensor in spacetime  $M^4$ .

## §19.6 Einstein's Field Equations of Gravitation

From equation (19.3.3)

$$R = -\kappa T,$$

we have

$$R - \frac{1}{2} \times 4R = \kappa T.$$

Since  $R = R^\mu_\mu$ ,  $T = g_{\mu\nu} T^{\mu\nu} = T^\mu_\mu$ ,  $g^{\nu\lambda} g_{\lambda\nu} = 4$  (as the dimension of spacetime  $M^4$  is 4, i.e.,  $\mu, \nu = 0, 1, 2, 3$ ), the above equation can be transformed into

$$R^\mu_\mu - \frac{1}{2} R g^{\nu\lambda} g_{\lambda\nu} = \kappa T^\mu_\mu.$$

Multiplying both sides by  $g_{\mu\nu}$ , we obtain

$$g_{\mu\nu} R^\mu_\mu - \frac{1}{2} R g_{\mu\nu} g^{\nu\lambda} g_{\lambda\nu} = \kappa g_{\mu\nu} T^\mu_\mu,$$

that is

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \kappa T_{\mu\nu}.$$

Finally, we obtain

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu}. \quad (19.6.1)$$

Equation (19.6.1) is precisely **Einstein's field equation of gravitation**.

Substituting  $R = -\kappa T$  into equation (19.6.1) yields:

$$\begin{aligned} R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}(-\kappa T) &= \kappa T_{\mu\nu}, \\ R_{\mu\nu} + \frac{1}{2}\kappa g_{\mu\nu}T &= \kappa T_{\mu\nu}, \end{aligned}$$

that is

$$R_{\mu\nu} = \kappa(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T), \quad (19.6.2)$$

which is another form of Einstein's field equation. Following the same method used to derive equation (19.6.1) from  $R = -\kappa T$ , one can directly derive equation (19.6.2) from  $R = -\kappa T$ . The next section will determine the constant:

$$\kappa = \frac{8\pi G}{c^4}.$$

In vacuum,  $T_{\mu\nu} = 0$  and  $T = 0$ , so the gravitational field equation (19.6.2) becomes

$$R_{\mu\nu} = 0. \quad (19.6.3)$$

Each particle is a Lie group, and a Lie group is a smooth manifold. Introducing a metric tensor on the smooth manifold describes its curvature, and the curvature of the smooth manifold gives rise to gravity. By embedding the smooth manifold representing the particle into  $M^4$  (the origin of  $M^4$  must be excluded from the solution of the gravitational field equations), we establish Einstein's field equation (19.6.1) to describe this gravity. Therefore, in essence, it is the curvature of the particle as a smooth manifold that produces gravity, not the curvature of a purely empty four-dimensional spacetime. Of course, since the parts of the Lie group other than the point-like particles constitute space (or a field), one could also say that gravity arises from the curvature of space—but here, the space referred to is an integral part of the particle.

## §19.7 Determination of the Coefficient $\kappa$

The method for determining the magnitude of the coefficient  $\kappa$  is well-known. For completeness, this section is included.

To determine the value of  $\kappa$ , we make the following assumptions and apply them to approximate Einstein's gravitational field equations under weak-field conditions. By comparing the resulting equation with the Poisson equation of Newtonian gravitational theory,  $\nabla^2\Phi = 4\pi G\rho$ , we can determine  $\kappa$ .

### 1. Linear Approximation of the Metric Tensor

1) The gravitational field is weak,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu},$$

where

$$(\eta_{\mu\nu}) = (\eta^{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$|h_{\mu\nu}| \ll 1.$$

Here,  $\eta_{\mu\nu}$  is the metric tensor of Minkowski spacetime  $M^4$ . Since  $\eta_{\mu\nu}$  is assumed to be

the metric of Minkowski spacetime, the coordinates describing the gravitational field must be chosen as

$$x^0 = ct, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z,$$

and not as

$$x^0 = ict, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z.$$

2) The gravitational field is static, meaning  $g_{\mu\nu}$  does not vary with time. Thus, the ordinary time derivatives of  $g_{\mu\nu}$  vanish:

$$g_{\mu\nu,0} = h_{\mu\nu,0} = 0.$$

3) The gravitational field varies slowly in space:

$$|g_{\mu\nu,i}| = |h_{\mu\nu,i}| \ll 1, \quad i = 1, 2, 3.$$

4) Point-like particles moving in the gravitational field are non-relativistic:

$$\left| \frac{dx^i}{dx^0} \right| \ll 1, \quad \left| \frac{dx^i}{d\tau} \right| \ll \left| \frac{dx^0}{d\tau} \right|, \quad \left| \frac{dx^i}{d\tau} \right| \ll c.$$

## 2. Linear Approximation of the Gravitational Field Equations

Assume the gravitational field is produced by a quasi-static distribution of dust. The energy-momentum tensor for such dust is

$$T^{\mu\nu} = \rho u^\mu u^\nu,$$

where  $\rho$  is the mass density measured by a local inertial observer at rest relative to the dust. In coordinates comoving with the dust:

$$u^\mu = \frac{dx^\mu}{d\tau} = \left( \frac{dx^0}{d\tau}, 0, 0, 0 \right) = \left( \frac{d(ct)}{d\tau}, 0, 0, 0 \right) = \frac{1}{\sqrt{-g_{00}}} (c, 0, 0, 0),$$

where

$$d\tau = \frac{ids}{c} = \frac{1}{c} \sqrt{-g_{\mu\nu} dx^\mu dx^\nu} = \frac{1}{c} \sqrt{-g_{00}} dx^0 = \sqrt{-g_{00}} dt.$$

The only non-zero component of the energy-momentum tensor is

$$T^{00} = \rho u^0 u^0 = \rho \frac{dx^0}{d\tau} \frac{dx^0}{d\tau} = \rho \frac{c}{\sqrt{-g_{00}}} \frac{c}{\sqrt{-g_{00}}} = -\frac{\rho c^2}{g_{00}}.$$

The trace of the energy-momentum tensor is

$$T = g_{\mu\nu} T^{\mu\nu} = g_{00} T^{00} = -\rho c^2.$$

The covariant form of the energy-momentum tensor is

$$T_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} T^{\alpha\beta},$$

so the only non-zero covariant component is

$$T_{00} = g_{0\alpha} g_{0\beta} T^{\alpha\beta} = g_{00} g_{00} T^{00} = g_{00} g_{00} \left( -\frac{\rho c^2}{g_{00}} \right) = -g_{00} \rho c^2 \approx \rho c^2,$$

that is

$$T_{00} \approx \rho c^2, \quad (19.7.1)$$

since using the first approximation condition,  $g_{00} \approx \eta_{00} = -1$ . Finally we have

$$T_{\mu\nu} = \begin{pmatrix} -g_{00} \rho c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \approx \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since only one component of the energy-momentum tensor is non-zero, we can determine  $\kappa$  by considering a single component of equation (19.6.1). Write the (00)-component of equation (19.6.1):

$$R_{00} - \frac{1}{2} g_{00} R = \kappa T_{00}. \quad (19.7.2)$$

Given that only  $T_{00}$  is non-zero, we have

$$R_{ij} - \frac{1}{2} g_{ij} R = 0, \quad R_{ij} = \frac{1}{2} g_{ij} R, \quad i, j = 1, 2, 3.$$

The scalar curvature is

$$R = g^{\mu\nu} R_{\mu\nu} = g^{00} R_{00} + g^{ij} R_{ij} = g^{00} R_{00} + \frac{1}{2} g^{ij} g_{ij} R.$$

Since  $g^{ij} g_{ij} = 3$ ,

$$\begin{aligned} R &= g^{00} R_{00} + \frac{3}{2} R, \\ R &= -2g^{00} R_{00}. \end{aligned} \quad (19.7.3)$$

The Christoffel symbols can be approximated as

$$\begin{aligned} \Gamma_{\alpha\beta}^{\mu} &\approx \frac{1}{2} \eta^{\mu\rho} (h_{\rho\alpha, \beta} + h_{\rho\beta, \alpha} - h_{\alpha\beta, \rho}), \\ \Gamma_{00}^i &\approx \frac{1}{2} \eta^{i\rho} (h_{\rho 0, 0} + h_{\rho 0, 0} - h_{00, \rho}) = -\frac{1}{2} \eta^{ii} h_{00, i} = -\frac{1}{2} h_{00, i}, \\ \Gamma_{00}^0 &\approx \frac{1}{2} h_{00, 0} = 0, \\ \Gamma_{0j}^i &\approx \frac{1}{2} \eta^{i\rho} (h_{\rho 0, j} + h_{\rho j, 0} - h_{0j, \rho}) = \frac{1}{2} (h_{i0, j} - h_{0j, i}), \\ \Gamma_{\mu\nu}^{\beta} \Gamma_{\beta\alpha}^{\alpha} &\approx 0, \quad \Gamma_{\mu\alpha}^{\beta} \Gamma_{\beta\nu}^{\alpha} \approx 0. \end{aligned} \quad (19.7.4)$$

The Ricci tensor can be approximated as

$$\begin{aligned} R_{\mu\nu} &= \partial_{\alpha} \Gamma_{\mu\nu}^{\alpha} - \partial_{\nu} \Gamma_{\mu\alpha}^{\alpha} + \Gamma_{\mu\nu}^{\beta} \Gamma_{\beta\alpha}^{\alpha} - \Gamma_{\mu\alpha}^{\beta} \Gamma_{\beta\nu}^{\alpha} \approx \partial_{\alpha} \Gamma_{\mu\nu}^{\alpha} - \partial_{\nu} \Gamma_{\mu\alpha}^{\alpha}, \\ R_{00} &\approx \partial_{\alpha} \Gamma_{00}^{\alpha} - \partial_0 \Gamma_{0\alpha}^{\alpha}. \end{aligned}$$

Since  $\partial_0 g_{\mu\nu} = 0$ ,  $\partial_0 \Gamma_{\mu\nu}^{\lambda} = 0$ , we have

$$R_{00} \approx \partial_{\alpha} \Gamma_{00}^{\alpha} = \partial_0 \Gamma_{00}^0 + \partial_i \Gamma_{00}^i = \partial_i \Gamma_{00}^i = -\frac{1}{2} h_{00, i, i}. \quad (19.7.5)$$

Substituting equations (19.7.1), (19.7.3), and (19.7.5) into equation (19.7.2),

$$R_{00} - \frac{1}{2} g_{00} R = \kappa T_{00},$$

we obtain

$$\begin{aligned} R_{00} - \frac{1}{2} g_{00} (-2g^{00} R_{00}) &= 2R_{00} = 2 \left( -\frac{1}{2} h_{00, i, i} \right) = \kappa \rho c^2, \\ h_{00, i, i} &= -\kappa \rho c^2. \end{aligned} \quad (19.7.6)$$

### 3. Determining $h_{00}$

The motion of a point-like particle in the gravitational field produced by another particle follows the geodesic equation (19.2.8):

$$\frac{Du^{\mu}}{d\tau} = 0,$$

that is

$$m \frac{Du^{\mu}}{d\tau} = 0.$$

The specific form is

$$\frac{d(mu^{\mu})}{d\tau} + m \Gamma_{\alpha\beta}^{\mu} u^{\alpha} u^{\beta} = 0. \quad (19.7.7)$$

That is, the trajectory of a point-like particle moving in the gravitational field produced by another



particle is a geodesic. Using the fourth approximation condition, the component form of the equation of motion (19.7.7) is

$$\frac{d}{d\tau} \left( m \frac{dx^0}{d\tau} \right) = 0, \quad (19.7.8)$$

$$\frac{d}{d\tau} \left( m \frac{dx^i}{d\tau} \right) + m \Gamma_{00}^i \left( \frac{dx^0}{d\tau} \right)^2 = 0, \quad i = 1, 2, 3. \quad (19.7.9)$$

Integrating equation (19.7.8) once gives  $dx^0 = a(d\tau)$  ( $a$  is a constant). Substituting into equation (19.7.9), we obtain

$$m \frac{d^2 x^i}{(dx^0)^2} = -m \Gamma_{00}^i.$$

Substituting equation (19.7.4) into the above equation, we obtain

$$m \frac{d^2 x^i}{(dx^0)^2} = \frac{m}{2} h_{00, i}.$$

Substituting  $x^0 = ct$  into the equation yields

$$m \frac{d^2 x^i}{dt^2} = \frac{m}{2} c^2 h_{00, i}. \quad (19.7.10)$$

In Newtonian gravity, according to Newton's second law, we obtain

$$m \frac{d^2 x^i}{dt^2} = m \left( -\frac{\partial \Phi}{\partial x^i} \right), \quad (19.7.11)$$

where  $\Phi = -\frac{GM}{r}$  is the Newtonian gravitational potential. Comparing equations (19.7.10) and (19.7.11), we have

$$\Phi = -\frac{c^2}{2} h_{00} + \text{constant}.$$

Assuming the gravitational field vanishes at infinity,  $\Phi = 0$ , and the spacetime metric becomes flat at infinity,  $h_{00} = 0$ , the constant must be zero. Thus

$$h_{00} = -\frac{2\Phi}{c^2}. \quad (19.7.12)$$

#### 4. Determining $\kappa$ by Comparison with the Poisson Equation

Substituting equation (19.7.12) into equation (19.7.6), we obtain

$$\nabla^2 \left( \frac{2\Phi}{c^2} \right) = \kappa \rho c^2, \quad \nabla^2 \Phi = \frac{1}{2} \kappa \rho c^4.$$

Comparing with the Poisson equation  $\nabla^2 \Phi = 4\pi G \rho$ , we obtain

$$\frac{1}{2} \kappa \rho c^4 = 4\pi G \rho,$$

so

$$\kappa = \frac{8\pi G}{c^4}. \quad (19.7.13)$$

Thus, the final form of Einstein's gravitational field equation (19.6.1) is

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (19.7.14)$$

and the final form of equation (19.6.2) is

$$R_{\mu\nu} = \frac{8\pi G}{c^4} (T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T). \quad (19.7.15)$$

Finally, let us compare gravitational and electromagnetic fields. The study of gravitational fields begins with the equivalence principle, which requires the use of connection coefficients and the metric tensor field. Thus, describing gravitational fields involves the metric tensor field, and

finding its explicit form. In other words, when studying the gravity of particles (or matter), particles are treated as generalized Riemannian manifolds. On the other hand, treating particles as principal bundles and introducing connections on them describes electromagnetic fields. Therefore, gravitational and electromagnetic fields correspond to distinct geometric structures.

# Chapter 20 Spin Magnetic Moments of Electrons, Protons, and Neutrons

## §20.1 Integral Curves of the $SU(2)$ Group Spin Operators

### 1. Integral Curves of the Spin Operator $\hat{S}_x$ for Protons and Antiprotons

We have concluded that the spin operator primarily responsible for the properties of protons and antiprotons is the  $SU(2)$  group operator  $\hat{S}_x$ . Now, we aim to derive the integral curves of the spin operator  $\hat{S}_x$ . From the expression of  $\hat{S}_x$ ,

$$\hat{S}_x = \frac{1}{2}\hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right),$$

the system of equations satisfied by the integral curves is

$$\begin{cases} \frac{du}{dt} = \frac{1}{2}\hbar v, \\ \frac{dv}{dt} = \frac{1}{2}\hbar u. \end{cases} \quad (20.1.1)$$

The characteristic equation of system (20.1.1) is

$$\begin{vmatrix} -\lambda & \frac{1}{2}\hbar \\ \frac{1}{2}\hbar & -\lambda \end{vmatrix} = 0,$$

i.e.,

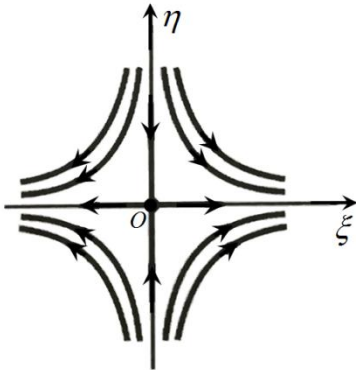
$$\lambda^2 - \frac{1}{4}\hbar^2 = 0. \quad (20.1.2)$$

Solving, we obtain

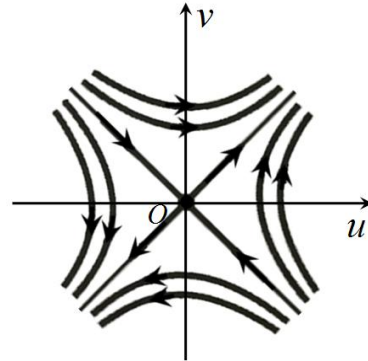
$$\lambda_1 = \frac{1}{2}\hbar, \quad \lambda_2 = -\frac{1}{2}\hbar.$$

These are two real roots with opposite signs. Therefore, system (20.1.1) can be transformed into the following standard form:

$$\frac{d\xi}{dt} = \frac{1}{2}\hbar\xi, \quad \frac{d\eta}{dt} = -\frac{1}{2}\hbar\eta. \quad (20.1.3)$$



**Figure 20.1.1** Integral curves of proton  $\hat{S}_x$  in the plane  $\xi\eta$



**Figure 20.1.2** Integral curves of proton  $\hat{S}_x$  in the plane  $uv$

Since the characteristic roots are real with opposite signs, the type of singular point is a saddle point. In the phase plane  $\xi\eta$ , this is illustrated in Figure 20.1.1. However, we intend to depict the shape of the singular point in the phase plane  $uv$  (the phase plane can be regarded as a plane within the tangent space at the identity element of the Lie group).

According to the coordinate transformation (6.9.8), we have

$$\begin{aligned}\xi &= (0 - \frac{1}{2}\hbar)u - \frac{1}{2}\hbar v, \\ \eta &= (0 + \frac{1}{2}\hbar)u - \frac{1}{2}\hbar v.\end{aligned}$$

In matrix form, this is

$$\hbar \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \xi \\ \eta \end{pmatrix}. \quad (20.1.4)$$

Ignoring the constant factor  $-\hbar$ , we obtain

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \quad (20.1.5)$$

which is

$$\begin{cases} \xi = \frac{1}{2}u + \frac{1}{2}v, \\ \eta = -\frac{1}{2}u + \frac{1}{2}v. \end{cases} \quad (20.1.6)$$

The inverse transformation of equation (20.1.5) is

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad (20.1.7)$$

that is

$$\begin{cases} u = \xi - \eta, \\ v = \xi + \eta. \end{cases} \quad (20.1.8)$$

We verify that equation (20.1.5) or (20.1.6) indeed transforms equation (20.1.3) into equation (20.1.1). From equation (20.1.6), we have

$$\begin{cases} \frac{d\xi}{dt} = \frac{1}{2} \frac{du}{dt} + \frac{1}{2} \frac{dv}{dt} = \frac{1}{2} \hbar \left( \frac{1}{2}u + \frac{1}{2}v \right), \\ \frac{d\eta}{dt} = -\frac{1}{2} \frac{du}{dt} + \frac{1}{2} \frac{dv}{dt} = -\frac{1}{2} \hbar \left( -\frac{1}{2}u + \frac{1}{2}v \right), \end{cases}$$

that is

$$\begin{cases} \frac{du}{dt} + \frac{dv}{dt} = \frac{1}{2} \hbar u + \frac{1}{2} \hbar v, \\ -\frac{du}{dt} + \frac{dv}{dt} = \frac{1}{2} \hbar u - \frac{1}{2} \hbar v. \end{cases} \quad (20.1.9)$$

Subtracting the two equations in system (20.1.9) yields

$$2 \frac{du}{dt} = \hbar v,$$

i.e.,

$$\frac{du}{dt} = \frac{1}{2} \hbar v. \quad (20.1.10)$$

Adding the two equations in system (20.1.9) gives

$$2 \frac{dv}{dt} = \hbar u,$$

i.e.,

$$\frac{dv}{dt} = \frac{1}{2}\hbar u . \quad (20.1.11)$$

Combining equations (20.1.10) and (20.1.11) results in equation (20.1.1). Therefore, equation (20.1.5) or (20.1.6) indeed transforms equation (20.1.3) into equation (20.1.1). The determinant of the transformation matrix in equation (20.1.5) is  $|J|=1/2 > 0$ , so this transformation is orientation-preserving.

On the phase plane  $\xi\eta$ , a vector on the  $\xi$ -axis, say  $(1,0)$ , transforms under equation (20.1.7)

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

into  $(1,1)$ , while a vector on the  $\eta$ -axis,  $(0,1)$ , transforms under equation (20.1.7)

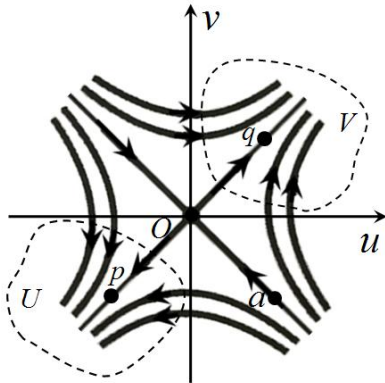
$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

into  $(-1,1)$ . Therefore, the lines corresponding to the  $\xi$ -axis and  $\eta$ -axis in the phase plane  $uv$  are given by:

$$\xi\text{-axis: } u - v = 0 ; \quad \eta\text{-axis: } u + v = 0 .$$

Based on this, we can sketch the shape of the singular point of  $\hat{S}_x$  in the phase plane  $uv$ , as shown in Figure 20.1.2. This computational result is consistent with previous calculations. Earlier calculations indicated that the circulation of  $\hat{S}_x$  around the unit circle is non-zero, implying that  $\hat{S}_x$  is rotational. Here, we find that the integral curves near the singular point of  $\hat{S}_x$  are also rotational.

Observing Figure 20.1.2 from the left and right sides, we notice that the directions of symmetric integral curves are opposite. Similarly, from the top and bottom views, symmetric integral curves also exhibit opposite directions. In particular, when viewed from any direction with the origin as the symmetry center, these integral curves show that symmetric pairs have opposite directions.

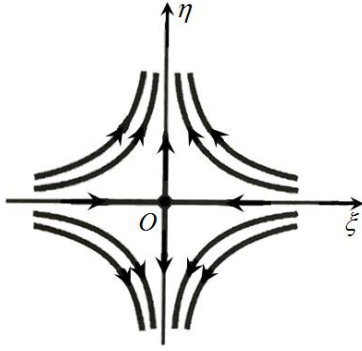


**Figure 20.1.3** Measured values of a local region  $V$  would yield  $+\frac{1}{2}\hbar$ .  
proton  $\hat{S}_x$

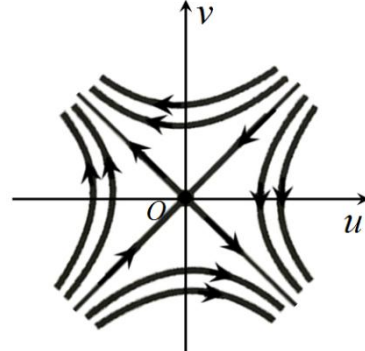
This is why the spin of a proton (or antiproton) simultaneously exists in two superimposed states. When measuring the  $z$ -component of the spin of a proton (or antiproton), one might obtain  $+\frac{1}{2}\hbar$  or  $-\frac{1}{2}\hbar$ .

For example, as shown in Figure 20.1.3, from point  $a$  to point  $O$  to point  $p$ , the direction is counterclockwise, but from point  $a$  to point  $O$  to point  $q$ , it is clockwise. Thus, the tangent vector directions at points  $p$  and  $q$  are opposite. If measuring the  $z$ -component of the proton spin at point  $p$  in a local region  $U$  yields  $-\frac{1}{2}\hbar$ , then measuring at point  $q$  in

The phase plane is a plane within the tangent space. Since the orientation of an antiproton is opposite to that of a proton, if the coordinates of the proton's tangent space are  $(t, \xi, \eta)$  or  $(t, u, v)$ , the coordinates of the antiproton's tangent space can be taken as  $(-t, \xi, \eta)$  or  $(-t, u, v)$ . Consequently, the direction of the integral curves of  $\hat{S}_x$  for the antiproton is exactly opposite to that of the proton, as shown in Figures 20.1.4 and 20.1.5.



**Figure 20.1.4** Integral curves of antiproton  $\hat{S}_x$  in the plane  $\xi\eta$



**Figure 20.1.5** Integral curves of antiproton  $\hat{S}_x$  in the plane  $uv$

At the singular point, the field components are continuously differentiable and non-singular. However, near the singular point, the direction of the tangent vectors of the field generally changes discontinuously. This is another reason for the randomness in the motion of point-like particles.

## 2. Integral Curves of the Spin Operator $\hat{S}_y$ for Neutrinos and Antineutrinos

We have concluded that the spin operator primarily responsible for the properties of neutrinos and antineutrinos is the  $SU(2)$  group operator  $\hat{S}_y$ . Now, we derive the integral curves of the spin operator  $\hat{S}_y$ . From the expression of  $\hat{S}_y$ ,

$$\hat{S}_y = \frac{i}{2}\hbar \left( v \frac{\partial}{\partial u} - u \frac{\partial}{\partial v} \right),$$

the system of equations satisfied by the integral curves is

$$\begin{cases} \frac{du}{dt} = \frac{i}{2}\hbar v, \\ \frac{dv}{dt} = -\frac{i}{2}\hbar u. \end{cases} \quad (20.1.12)$$

The system can be rewritten as

$$u \frac{du}{dt} = \frac{i}{2}\hbar uv, \quad v \frac{dv}{dt} = -\frac{i}{2}\hbar uv. \quad (20.1.13)$$

Combining these, we get

$$u du = -v dv. \quad (20.1.14)$$

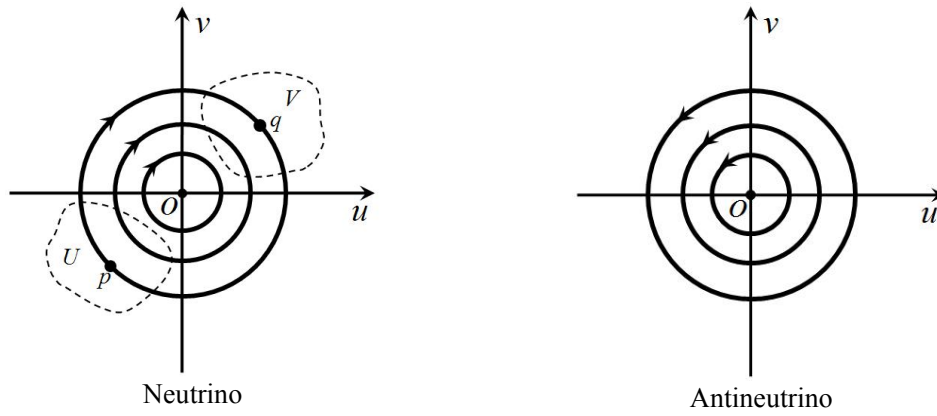
Integrating both sides yields

$$\frac{1}{2}u^2 = -\frac{1}{2}v^2 + \frac{1}{2}C.$$

Thus, the general solution is

$$u^2 + v^2 = C, \quad (20.1.15)$$

where  $C$  is an arbitrary non-negative constant. Therefore, the type of singular point is a center, as shown in the phase plane  $uv$  in Figure 20.1.6.



**Figure 20.1.6** Integral curves of  $\hat{S}_y$

Upon closer inspection of Figure 20.1.6, we observe that all integral curves have the same direction—only one direction. This is distinctly different from Figure 20.1.2. This is why neutrinos (or antineutrinos) do not have two superimposed spin states but only one state. When measuring the  $z$ -component of the spin of a neutrino (or antineutrino), only a unique value can be obtained. The integral curves in the left diagram of Figure 20.1.6 are left-handed, so they correspond to neutrinos. The tangent vector directions at points  $p$  and  $q$  are the same, both clockwise. When measuring the  $z$ -component of neutrino spin at point  $p$  in a local region  $U$ , the value obtained is  $-\frac{1}{2}\hbar$ . Similarly, measuring at point  $q$  in a local region  $V$  also yields  $-\frac{1}{2}\hbar$ . At any point, measuring the  $z$ -component of neutrino spin yields only this unique value because the integral curves of the neutrino spin operator have only one direction. Antineutrinos are right-handed, so their integral curves should be counterclockwise, as shown in the right diagram of Figure 20.1.6. When measuring the  $z$ -component of antineutrino spin, only the unique value  $+\frac{1}{2}\hbar$  can be obtained, as the integral curves of the antineutrino spin operator also have only one direction.

### 3. Integral Curves of the Spin Operator $\hat{S}_z$ for Electrons and Positrons

We have concluded that the spin operator primarily responsible for the properties of electrons and positrons is the  $SU(2)$  group operator  $\hat{S}_z$ . Now, we derive the integral curves of the spin operator  $\hat{S}_z$ . From the expression of  $\hat{S}_z$ ,

$$\hat{S}_z = \frac{1}{2}\hbar \left( u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v} \right),$$

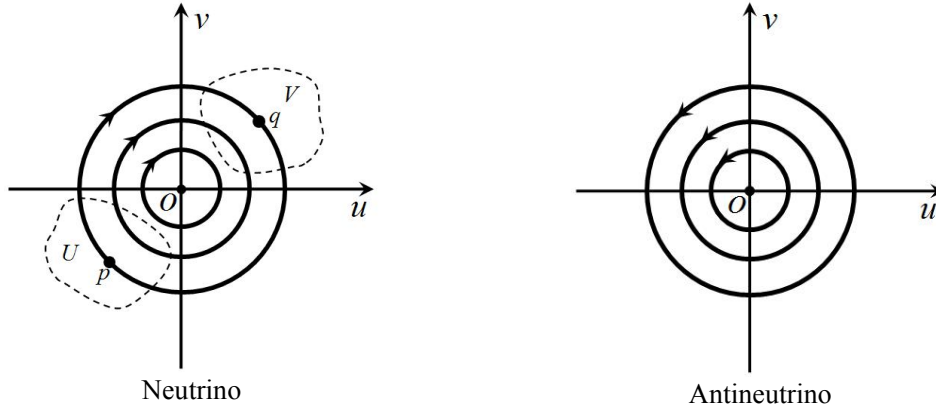
Integrating both sides yields

$$\frac{1}{2}u^2 = -\frac{1}{2}v^2 + \frac{1}{2}C.$$

Thus, the general solution is

$$u^2 + v^2 = C, \quad (20.1.15)$$

where  $C$  is an arbitrary non-negative constant. Therefore, the type of singular point is a center, as shown in the phase plane  $uv$  in Figure 20.1.6.



**Figure 20.1.6** Integral curves of  $\hat{S}_y$

Upon closer inspection of Figure 20.1.6, we observe that all integral curves have the same direction—only one direction. This is distinctly different from Figure 20.1.2. This is why neutrinos (or antineutrinos) do not have two superimposed spin states but only one state. When measuring the  $z$ -component of the spin of a neutrino (or antineutrino), only a unique value can be obtained. The integral curves in the left diagram of Figure 20.1.6 are left-handed, so they correspond to neutrinos. The tangent vector directions at points  $p$  and  $q$  are the same, both clockwise. When measuring the  $z$ -component of neutrino spin at point  $p$  in a local region  $U$ , the value obtained is  $-\frac{1}{2}\hbar$ . Similarly, measuring at point  $q$  in a local region  $V$  also yields  $-\frac{1}{2}\hbar$ . At any point, measuring the  $z$ -component of neutrino spin yields only this unique value because the integral curves of the neutrino spin operator have only one direction. Antineutrinos are right-handed, so their integral curves should be counterclockwise, as shown in the right diagram of Figure 20.1.6. When measuring the  $z$ -component of antineutrino spin, only the unique value  $+\frac{1}{2}\hbar$  can be obtained, as the integral curves of the antineutrino spin operator also have only one direction.

### 3. Integral Curves of the Spin Operator $\hat{S}_z$ for Electrons and Positrons

We have concluded that the spin operator primarily responsible for the properties of electrons and positrons is the  $SU(2)$  group operator  $\hat{S}_z$ . Now, we derive the integral curves of the spin operator  $\hat{S}_z$ . From the expression of  $\hat{S}_z$ ,

$$\hat{S}_z = \frac{1}{2}\hbar \left( u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v} \right),$$

the system of equations satisfied by the integral curves is

$$\begin{cases} \frac{du}{dt} = \frac{1}{2}\hbar u, \\ \frac{dv}{dt} = -\frac{1}{2}\hbar v. \end{cases} \quad (20.1.16)$$

The characteristic equation of system (20.1.16) is

$$\begin{vmatrix} \frac{1}{2}\hbar - \lambda & 0 \\ 0 & -\frac{1}{2}\hbar - \lambda \end{vmatrix} = 0,$$

i.e.,

$$\left( \frac{1}{2}\hbar - \lambda \right) \left( -\frac{1}{2}\hbar - \lambda \right) = 0. \quad (20.1.17)$$

Solving, we obtain

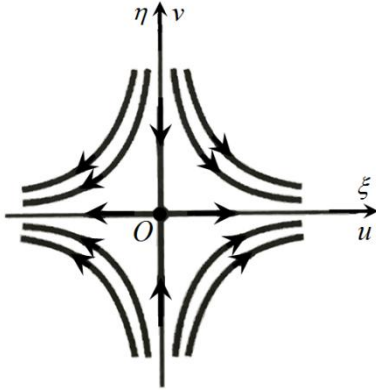


$$\lambda_1 = \frac{1}{2}\hbar, \quad \lambda_2 = -\frac{1}{2}\hbar.$$

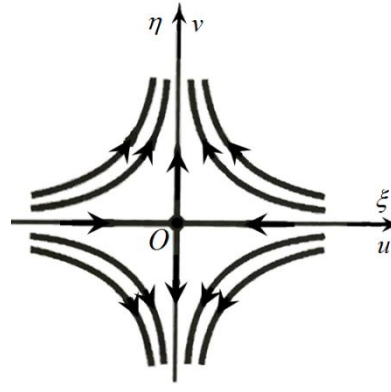
These are two real roots with opposite signs. Therefore, system (20.1.16) can be transformed into the following standard form:

$$\frac{d\xi}{dt} = \frac{1}{2}\hbar\xi, \quad \frac{d\eta}{dt} = -\frac{1}{2}\hbar\eta. \quad (20.1.18)$$

Since the characteristic roots are real with opposite signs, the type of singular point is a saddle point. However, because the form of the system is identical in both the phase plane  $\xi\eta$  and the phase plane  $uv$ , the shape of the singular point is also the same in both planes. The shape of the singular point is shown in Figure 20.1.7.



**Figure 20.1.7** Integral curves of electron  $\hat{S}_z$



**Figure 20.1.8** Integral curves of positron  $\hat{S}_z$

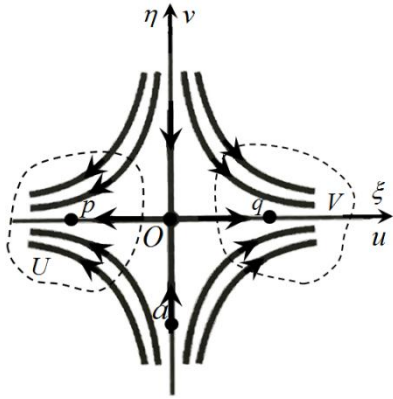
Observing Figure 20.1.7 from the left and right sides, we notice that the directions of symmetric integral curves are opposite. Similarly, from the top and bottom views, symmetric integral curves also exhibit opposite directions. In particular, when viewed from any direction with the origin as the symmetry center, symmetric integral curves have opposite directions. This is why

the spin of an electron (or positron) simultaneously exists in two superimposed states. When measuring the  $z$ -component of the spin of an electron (or positron), one might obtain  $+\frac{1}{2}\hbar$  or  $-\frac{1}{2}\hbar$ . For example, as

shown in Figure 20.1.9, from point  $a$  to point  $O$  to point  $p$ , the direction is counterclockwise, but from point  $a$  to point  $O$  to point  $q$ , it is clockwise. Thus, the tangent vector directions at points  $p$  and  $q$  are opposite.

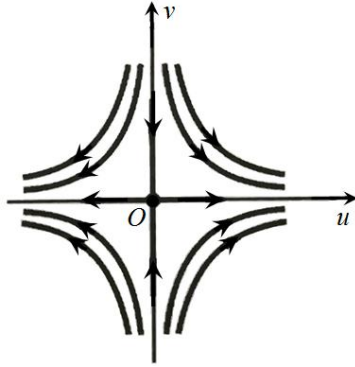
If measuring the  $z$ -component of electron spin at point  $p$  in a local region  $U$  yields  $-\frac{1}{2}\hbar$ , then measuring at

point  $q$  in a local region  $V$  would yield  $+\frac{1}{2}\hbar$ .

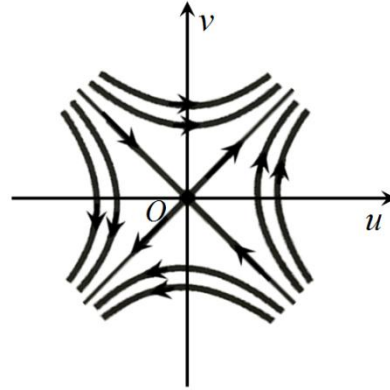


**Figure 20.1.9** Measured values of electron  $\hat{S}_z$

#### 4. The Relationship between $\hat{S}_x$ and $\hat{S}_z$



(Figure 20.1.7) Electron



(Figure 20.1.2) Proton

**Figure 20.1.10** The relationship between  $\hat{S}_x$  and  $\hat{S}_z$

The expressions for the spin operators  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  of the  $SU(2)$  group for electrons and protons discussed earlier are all presented in the same oriented phase plane  $uv$ . That is, both electrons and protons are assigned the same orientation. The same applies to positrons and antiprotons.

For ease of reading and comparison, we redraw Figures 20.1.7 and 20.1.2 in Figure 20.1.10. By comparing these two figures, we observe that rotating the integral curve diagram of the electron's spin operator  $\hat{S}_z$  counterclockwise by  $45^\circ$  yields the integral curve diagram of the proton's spin operator  $\hat{S}_x$ .

## §20.2 Integral Curves of Photon Spin Operators

The three spin operators for photons and antiphotons are

$$\hat{L}_x = -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right), \quad \hat{L}_y = -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right), \quad \hat{L}_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right).$$

Let us derive the integral curves of the spin operator  $\hat{L}_z$ . From its expression, the system of equations satisfied by the integral curves is

$$\begin{cases} \frac{dx}{dt} = i\hbar y, \\ \frac{dy}{dt} = -i\hbar x. \end{cases} \quad (20.2.1)$$

The system can be rewritten as

$$x \frac{dx}{dt} = i\hbar xy, \quad y \frac{dy}{dt} = -i\hbar xy. \quad (20.2.2)$$

Combining these, we obtain

$$x dx = -y dy. \quad (20.2.3)$$

Integrating both sides gives

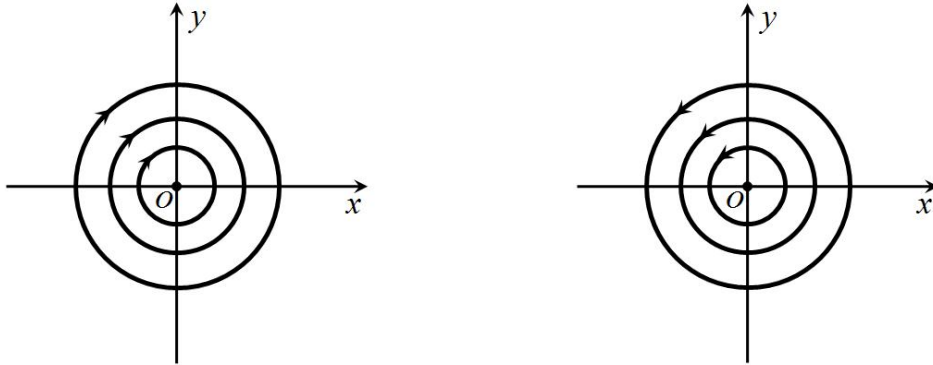
$$\frac{1}{2}x^2 = -\frac{1}{2}y^2 + \frac{1}{2}C.$$

Thus, the general solution is

$$x^2 + y^2 = C, \quad (20.2.4)$$

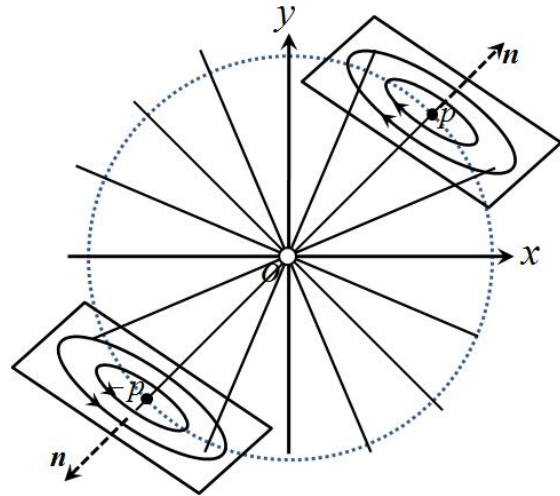
where  $C$  is an arbitrary non-negative constant. Therefore, the type of singular point is a center, as shown in the phase plane  $xy$  in Figure 20.2.1. The expressions for  $\hat{L}_x$  and  $\hat{L}_y$  are similar to  $\hat{L}_z$ ,

and their integral curves are also similar.



**Figure 20.2.1** Integral curves of photon spin operators

Carefully observing the left or right diagram in Figure 20.2.1, we find that all integral curves have the same direction—only one direction. This is similar to Figure 20.1.6. However, why does a neutrino (or antineutrino) have only one spin state, yielding a unique value when measuring its  $z$ -component, while a photon has two spin states, yielding two opposite values when measuring its  $z$ -component? This is because the topologies of photons and neutrinos/antineutrinos differ: a photon is a real projective space  $RP^3$ , whereas neutrinos and antineutrinos are three-dimensional unit spheres  $S^3$ . As shown in Figure 20.2.2, this figure is a model of  $RP^3$ , with a left-handed orientation. The dotted arrow  $\mathbf{n}$  in the figure indicates the normal vector. The integral curves of the spin operator at point  $p$  rotate clockwise, while the integral curves of the spin operator at point  $-p$  rotate counterclockwise. Points  $p$  and  $-p$  lie on the same line passing through the origin. Thus, these two points belong to the same point, denoted as  $A$ . When points  $p$  and  $-p$  are identified as the same point  $A$ , the neighborhood of  $A$  contains two tangent vector fields with exactly opposite integral curve directions. This shows that in the neighborhood of every point of a photon, there are two spin tangent vector fields (operators) with opposite directions. In other words, in the neighborhood of every point of a photon, there exist integral curves of two different directions, as shown in Figure 20.2.1. Therefore, at every point of a photon, there are two opposite spin  $z$ -components, leading to the possibility of measuring  $+\hbar$  or  $-\hbar$  for the photon's spin  $z$ -component. The same applies to the spin  $x$ - and  $y$ -components.



**Figure 20.2.2** Points  $p$  and  $-p$  are the same point

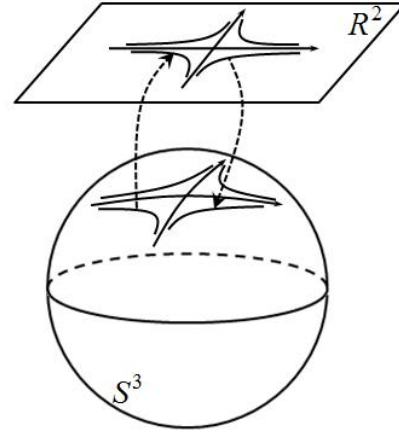
The differences among  $\hat{L}_y$ ,  $\hat{L}_z$ , and  $\hat{L}_x$  lie in their directions:  $\hat{L}_x$  is the spin operator along the  $x$ -direction, and  $\hat{L}_y$  is along the  $y$ -direction. However, their expressions are similar, so their integral curves are similar. Consequently, when measuring the spin component of a photon along any direction ( $x$ ,  $y$ , or  $z$ ), it is possible to obtain either  $+\hbar$  or  $-\hbar$ .

Finally, it should be noted that although we plot the integral curves of spin operators in the phase plane  $R^2$ , these integral curves actually exist on the Lie group, not on the phase plane. Nevertheless, we can regard the integral curves on the Lie group as those on the phase plane, or vice versa, for the following reasons:

We can view the tangent space  $T_e G$  at the identity element  $e$  of the Lie group  $G$  (i.e., the

Lie algebra  $\text{Lie}(G)$  at  $e$ ) as a three-dimensional Euclidean space  $R^3$ . The spin operators  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$  are the three components of the spin operator  $\hat{S}$ . We can consider  $\hat{S}_x$  as a tangent vector field on a plane in  $R^3$  and plot its integral curves, as shown in Figure 20.1.2. This plane is called the phase plane  $R^2$ , which naturally resides in  $R^3$  or the tangent space  $T_e G$ . The same understanding applies to the other two components  $\hat{S}_y$  and  $\hat{S}_z$ .

According to Theorem 12.5.7, the exponential map is a diffeomorphism  $\exp_e : B \rightarrow V$  from a neighborhood  $B \subset T_e G$  of the origin (zero vector) in the tangent space  $T_e G$  at the identity element  $e$  to a neighborhood  $V \subset G$  of the identity element  $e$  in the Lie group  $G$ . Therefore, there is a one-to-one correspondence between points in  $V$  and points in  $B$ . Correspondingly, each point in a two-dimensional neighborhood  $\tilde{V}$  of  $e$  in  $G$  corresponds one-to-one with a point in a neighborhood  $\tilde{B} \subset R^2$  of the origin in the phase plane  $R^2$ . Thus, the integral curves of the spin operator in  $\tilde{V}$  can be mapped one-to-one onto a neighborhood  $\tilde{B}$  in the phase plane, as shown in Figure 20.2.3. Conversely, the integral curves of the spin operator in a neighborhood  $\tilde{B}$  of the origin in the phase plane can be mapped one-to-one into a two-dimensional neighborhood  $\tilde{V}$  of  $e$ . Different integral curves of the spin operator in  $\tilde{V}$  correspond to different integral curves in the phase plane. The integral curves of the spin operator in  $\tilde{V}$  do not intersect, and their images in the phase plane also do not intersect.



**Figure 20.2.3** One-to-one correspondence of integral curves between two neighborhoods  $\tilde{V}$  and  $\tilde{B}$

In summary, there is a one-to-one correspondence between the integral curves of the spin operator in  $\tilde{V}$  and their images in a neighborhood  $\tilde{B}$  of the phase plane. Therefore, for intuition and convenience, we often use the integral curves of the spin operator in a neighborhood  $\tilde{B}$  of the origin in the phase plane  $R^2$  to represent the integral curves of the spin operator in a two-dimensional neighborhood  $\tilde{V}$  of the identity element  $e$  in the Lie group  $G$ . Additionally, since  $V$  and  $B$  are homeomorphic, and  $B$  is a neighborhood in the tangent space—which can be viewed as a Euclidean space—we can establish the coordinate system of neighborhood  $V$  through the coordinate system of neighborhood  $B$ , which is called the normal coordinate system.

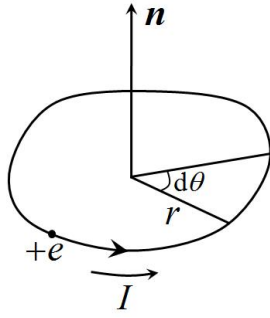
## §20.3 Orbital and Spin Magnetic Moments of Electrons

### 1. Magnetic Moment

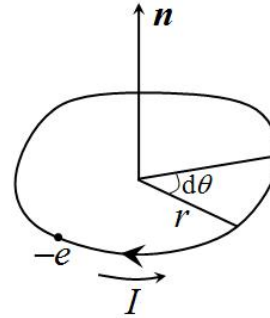
The magnetic moment is a concept in classical physics. When a current of intensity  $I$  flows through a closed loop, the magnetic moment produced is

$$\boldsymbol{\mu} = I\mathbf{A} = IAn, \quad (20.3.1)$$

where  $\mathbf{A}$  is the area vector, whose magnitude  $A$  is the area enclosed by the closed loop. The direction of  $\mathbf{A}$  is taken as the normal vector  $\mathbf{n}$  of the planar loop. The direction of the magnetic moment  $\boldsymbol{\mu}$  aligns with the normal vector  $\mathbf{n}$  of the planar loop (current-carrying circuit), as shown in Figures 20.3.1 and 20.3.2.



**Figure 20.3.1** Magnetic moment of a positively charged particle



**Figure 20.3.2** Magnetic moment of a negatively charged particle

## 2.Orbital Magnetic Moment of Electrons

As shown in Figure 20.3.1, consider a particle with a unit positive charge  $e$  ( $e > 0$ ) and rest mass  $m$  moving along a closed loop. The magnitude of the area  $A$  swept out by the particle in one revolution can be obtained from equation (9.6.3):

$$A = \iint_U r dr d\theta = \int_0^{2\pi} d\theta \int_0^r r dr = \int_0^{2\pi} \frac{1}{2} r^2 d\theta = \int_0^T \frac{1}{2} r^2 \frac{d\theta}{dt} dt,$$

where  $T$  is the period. Since

$$\frac{1}{2} r^2 \frac{d\theta}{dt} = \frac{L}{2m},$$

and the angular momentum  $L$  about the center of the orbit is conserved, we have

$$A = \int_0^T \frac{L}{2m} dt = \frac{LT}{2m}.$$

Written in vector form as

$$\mathbf{A} = \frac{T}{2m} \mathbf{L}. \quad (20.3.2)$$

The current intensity  $I$  is the product of the number of times  $\frac{1}{T}$  a positively charged particle passes any point on the orbit per second and the charge  $e$ , thus,

$$I = \frac{e}{T}. \quad (20.3.3)$$

Substituting equations (20.3.2) and (20.3.3) into equation (20.3.1), the orbital magnetic moment of a positively charged particle is obtained as

$$\hat{\mu}_L = \frac{e}{2m} \hat{L} = M_L \hat{L}, \quad (20.3.4)$$

where the magnetic moment and orbital angular momentum are written in operator form as  $\hat{\mu}_L$  and  $\hat{L}$ , respectively. Here,

$$M_L = \frac{e}{2m}.$$

The quantity  $M_L$  is called the orbital gyromagnetic ratio, which equals the ratio of the magnetic moment due to orbital motion to the orbital angular momentum:

$$M_L = \frac{\mu_L}{L}.$$

Since the electron carries a unit negative charge, the orbital magnetic moment of the electron is derived from equation (20.3.4) as

$$\hat{\mu}_L = -\frac{e}{2m_e} \hat{L} = M_L \hat{L}, \quad (20.3.5)$$

where

$$M_L = -\frac{e}{2m_e}, \quad (20.3.6)$$

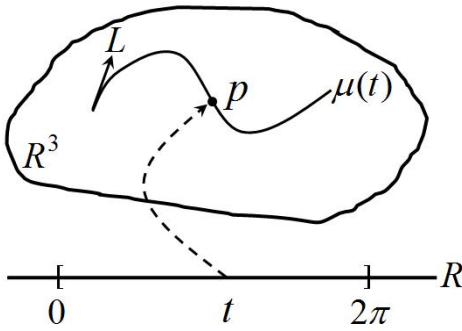
and  $m_e$  is the rest mass of the electron. See Figure 20.3.2.

### 3. Magnetic Moment as a Curve

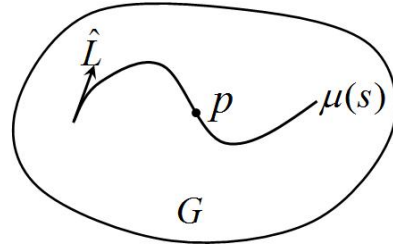
The scalar form  $\mu$  of the magnetic moment  $\boldsymbol{\mu}$  from equation (20.3.1) can be written as:

$$\mu(t) = IA = \frac{I}{2m} \int_0^t L d\alpha,$$

where  $t \in [0, 2\pi]$ . Based on this expression, we can interpret the magnetic moment  $\mu$  as a curve  $\mu(t)$  in space  $R^3$ , as shown in Figure 20.3.3. The angular momentum  $L$  is the tangent vector to this curve  $\mu(t)$ , and the integral of  $L$  along the entire curve yields the magnitude of the magnetic moment  $\mu$ .



**Figure 20.3.3** Magnetic moment  $\mu$  as a curve



**Figure 20.3.4** Magnetic moment on a Lie group

In the above, the particle is treated as a point mass, i.e., a point-like particle. Now, if we consider the particle as a Lie group (such as  $SU(2)$  or  $SO(3)$ ), the angular momentum  $L$  should be replaced by the particle's spin operator  $\hat{L}$ . The spin operator is a left- (or right-) invariant tangent vector field on the Lie group. The integral curves of a left- (or right-) invariant tangent vector field on a Lie group are geodesics on the Lie group. As shown in Figure 20.3.4,  $\hat{L}$  is a left- (or right-) invariant tangent vector field on the Lie group, or the particle's spin operator, and  $\mu(s)$  is its integral curve (geodesic).

Whenever a charged point-like particle undergoes directed motion, it generates a current, regardless of whether its trajectory is closed or open. Similarly, whenever the trajectory of a charged point-like particle is curved rather than straight, it produces a magnetic moment, regardless of whether the trajectory is closed or open.

According to equation (20.3.5), we can set the spin magnetic moment operator of the particle as  $\hat{\mu} = M\hat{L}$ , where  $M$  is a constant. Then

$$\mu = \iiint_G u^*(\mathbf{r}) M \hat{L} u(\mathbf{r}) d\tau = M \iiint_G u^*(\mathbf{r}) \hat{L} u(\mathbf{r}) d\tau.$$

This expression can also be written as a function of a parameter  $s$ , such as

$$\mu(s) = M \int_{z_0}^s \left\{ \int_{y_0}^{y_1} \left[ \int_{x_0}^{x_1} u^*(\mathbf{r}) \hat{L} u(\mathbf{r}) dx \right] dy \right\} dz,$$

where  $s \in [z_0, z_1]$ . Therefore, we can still interpret the magnetic moment  $\mu$  as a curve  $\mu(s)$  on the Lie group. The integral along this entire curve gives the magnitude of the magnetic moment  $\mu$ . When  $u(\mathbf{r})$  is an eigenfunction of the spin operator  $\hat{L}$ ,  $u^*(\mathbf{r}) \hat{L} u(\mathbf{r})$  is its length (eigenvalue), which is constant. In this case, the curve  $\mu(s)$  corresponds to an integral curve of

$\hat{L}$ .

#### 4. Spin Magnetic Moment of the Electron

The spin magnetic moment of the electron can be derived by taking the non-relativistic limit of the Dirac equation for an electron in an electromagnetic field.

For convenience in calculation, Gaussian units are adopted. The electromagnetic field can be expressed as  $(\mathbf{A}, \Phi)$ . Starting from the Dirac equation for a free electron (16.9.5), we make the following substitutions:

$$i\hbar \frac{\partial \psi}{\partial t} \rightarrow i\hbar \frac{\partial \psi}{\partial t} + e\Phi, \quad (20.3.7)$$

$$-i\hbar \nabla \rightarrow -i\hbar \nabla + \frac{e}{c} \mathbf{A} = \mathbf{P} + \frac{e}{c} \mathbf{A}, \quad \text{where } \mathbf{P} = -i\hbar \nabla, \quad (20.3.8)$$

we obtain the Dirac equation for an electron in the electromagnetic field  $(\mathbf{A}, \Phi)$  is

$$i\hbar \frac{\partial \psi}{\partial t} = \left[ c\boldsymbol{\alpha} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) - e\Phi + m_e c^2 \beta \right] \psi. \quad (20.3.9)$$

Let

$$\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix} \exp\left(-\frac{i}{\hbar} m_e c^2 t\right), \quad (20.3.10)$$

and substitute equation (20.3.10) into equation (20.3.9) to obtain

$$i\hbar \frac{\partial \varphi}{\partial t} = c\boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \chi - e\Phi \varphi, \quad (20.3.11)$$

$$i\hbar \frac{\partial \chi}{\partial t} = c\boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \varphi - e\Phi \chi - 2m_e c^2 \chi. \quad (20.3.12)$$

Transforming equation (20.3.12) yields

$$\chi = \frac{1}{2m_e c} \boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \varphi - \frac{1}{2m_e c^2} e\Phi \chi - \frac{1}{2m_e c^2} i\hbar \frac{\partial \chi}{\partial t},$$

which is an equation expanded in powers of  $1/c$ . By neglecting terms containing  $1/2m_e c^2$ , we obtain the first-order approximation (non-relativistic approximation) of the expansion:

$$\chi \approx \frac{1}{2m_e c} \boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \varphi, \quad (20.3.13)$$

thus  $\chi/\varphi \approx v/c$ . Substituting equation (20.3.13) into equation (20.3.11) gives

$$i\hbar \frac{\partial \varphi}{\partial t} = \frac{1}{2m_e} \left[ \boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \right]^2 \varphi - e\Phi \varphi. \quad (20.3.14)$$

Next, we simplify equation (20.3.14).

First, we need the following relation:

$$\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P} = -i\hbar (\nabla \times \mathbf{A}). \quad (20.3.15)$$

The derivation of equation (20.3.15) is as follows:

$$\begin{aligned} (\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P})_x \varphi &= -i\hbar (\nabla \times \mathbf{A} + \mathbf{A} \times \nabla)_x \varphi \\ &= -i\hbar \left[ \left( \frac{d}{dy} A_z \varphi - \frac{d}{dz} A_y \varphi \right) + \left( A_y \frac{d\varphi}{dz} - A_z \frac{d\varphi}{dy} \right) \right] \\ &= -i\hbar \left( \frac{dA_z}{dy} - \frac{dA_y}{dz} \right) \varphi = -i\hbar (\nabla \times \mathbf{A})_x \varphi. \end{aligned} \quad (20.3.16)$$

Similarly, we obtain

$$(\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P})_y \varphi = -i\hbar (\nabla \times \mathbf{A})_y \varphi, \quad (20.3.17)$$

$$(\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P})_z \varphi = -i\hbar (\nabla \times \mathbf{A})_z \varphi. \quad (20.3.18)$$

Combining equations (20.3.16), (20.3.17), and (20.3.18) gives

$$(\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P}) \varphi = -i\hbar (\nabla \times \mathbf{A}) \varphi,$$

that is



$$\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P} = -i\hbar(\nabla \times \mathbf{A}).$$

This is equation (20.3.15).

Second, we also need equation (16.11.9):

$$(\boldsymbol{\sigma} \cdot \mathbf{A})(\boldsymbol{\sigma} \cdot \mathbf{B}) = \mathbf{A} \cdot \mathbf{B} + i\boldsymbol{\sigma} \cdot (\mathbf{A} \times \mathbf{B}), \quad (20.3.19)$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are arbitrary vectors.

Using equations (20.3.15) and (20.3.19), we get

$$\begin{aligned} \left[ \boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \right]^2 &= \left[ \boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \right] \cdot \left[ \boldsymbol{\sigma} \cdot \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \right] \\ &= \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right)^2 + i\boldsymbol{\sigma} \cdot \left[ \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \times \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right) \right] \\ &= \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right)^2 + i\frac{e}{c} \boldsymbol{\sigma} \cdot [\mathbf{P} \times \mathbf{A} + \mathbf{A} \times \mathbf{P}] \\ &= \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right)^2 + \frac{e\hbar}{c} \boldsymbol{\sigma} \cdot (\nabla \times \mathbf{A}) \\ &= \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right)^2 + \frac{e\hbar}{c} \boldsymbol{\sigma} \cdot \mathbf{B}. \end{aligned}$$

Therefore, equation (20.3.14) simplifies to

$$i\hbar \frac{\partial}{\partial t} \varphi = \left[ \frac{1}{2m_e} \left( \mathbf{P} + \frac{e}{c} \mathbf{A} \right)^2 + \frac{e\hbar}{2m_e c} \boldsymbol{\sigma} \cdot \mathbf{B} - e\Phi \right] \varphi. \quad (20.3.20)$$

Let

$$\hat{\mu}_e = -\frac{e\hbar}{2m_e c} \boldsymbol{\sigma} = -\frac{e}{m_e c} \mathbf{S}, \quad \left( \mathbf{S} = \frac{\hbar}{2} \boldsymbol{\sigma} \right). \quad (20.3.21)$$

Here,  $\hat{\mu}_e$  is the spin magnetic moment of the electron. Its projection onto the positive  $z$ -axis is

$$\mu_e = -\frac{e}{m_e c} S_z, \quad (20.3.22)$$

where  $S_z = \frac{1}{2}\hbar$ . Conventionally, the projection value  $\mu_e$  of  $\hat{\mu}_e$  onto the positive  $z$ -axis is taken as the magnitude of  $\hat{\mu}_e$ . When referring to the spin magnetic moment of a particle, it generally denotes this magnitude.

In equation (20.3.21),  $\hat{\mu}_e$  is the spin magnetic moment of the electron in Gaussian units. Converting to SI units, the spin magnetic moment of the electron in SI units is

$$\hat{\mu}_e = -\frac{e\hbar}{2m_e} \boldsymbol{\sigma} = -\frac{e}{m_e} \mathbf{S},$$

with magnitude

$$\mu_e = -\frac{e}{m_e} S_z = -M_S S_z, \quad (20.3.23)$$

where

$$M_S = \frac{\hat{\mu}_e}{\mathbf{S}} = -\frac{e}{m_e}.$$

The quantity  $M_S$  is called the spin gyromagnetic ratio, which equals the ratio of the electron's spin magnetic moment to its spin angular momentum.

Since a point-like electron performing orbital motion resides in  $M^4$  spacetime, the angle rotated in one full revolution is  $2\pi$ . In contrast, a point-like electron performing spin motion resides in  $S^3$  space, where the angle required for one full revolution is  $4\pi$ . Therefore, the area swept by the current for a point-like electron in spin motion is twice that of its orbital motion, which is why  $M_S = 2M_L$ . However, experimental measurements of the electron's spin magnetic moment reveal that the experimental value of the spin gyromagnetic ratio  $M_{S,\text{exp}}$  is slightly less



than  $M_S = -e/m_e$  :

$$M_{S,\text{exp}} = -\frac{e}{m_e}(1 + 0.0011659202) ,$$

which can be attributed to photon perturbations on the electron.

The measured value of the electron's spin magnetic moment is

$$\mu_{e,\text{exp}} = -\frac{e}{m_e}(1 + 0.0011659202)S_z .$$

Given the Bohr magneton  $\mu_B = \frac{e\hbar}{2m_e}$ , we have

$$\frac{e}{m_e} = \frac{2\mu_B}{\hbar} .$$

Thus, the measured value of the electron's spin magnetic moment is

$$\mu_{e,\text{exp}} = -\frac{e}{m_e}(1 + 0.0011659202)S_z = -\frac{2\mu_B}{\hbar}(1 + 0.0011659202)\frac{1}{2}\hbar ,$$

i.e.,

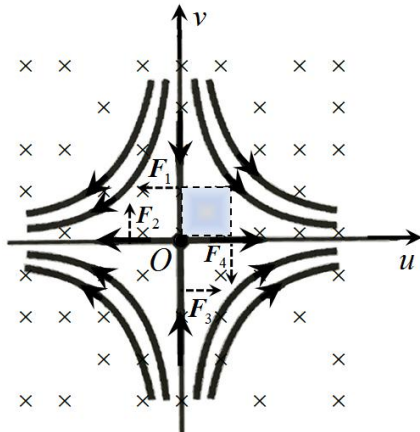
$$\mu_{e,\text{exp}} = -(1 + 0.0011659202)\mu_B .$$

The theoretical value of the electron's spin magnetic moment is

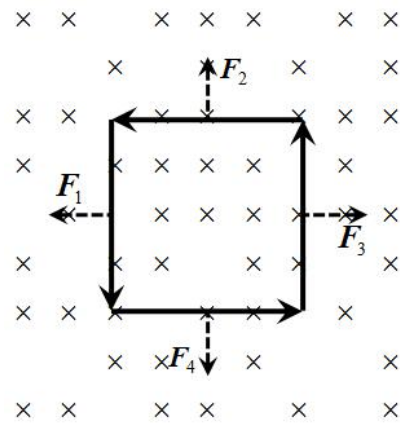
$$\mu_e = -\mu_B = -\frac{e\hbar}{2m_e} . \quad (20.3.24)$$

### 5.The Reason Why Electrons Possess a Spin Magnetic Moment

The Stern–Gerlach experiment revealed that when a beam of hydrogen atoms in the  $s$ -state passes through a non-uniform magnetic field, the beam splits, proving that electrons possess a spin magnetic moment. The magnetic moment is a concept from classical physics. We can analyze the reason why electrons have a spin magnetic moment using classical methods, based on the singularity diagram of the electron (Figure 20.1.7).



**Figure 20.3.5** Electron in a non-uniform magnetic field



**Figure 20.3.6** Rectangular coil within the electron

Based on Figure 20.1.7, we draw Figure 20.3.5. The curves in Figure 20.3.5 represent the integral curves of the spin operator  $\hat{S}_z$ . Since  $\hat{S}_z$  is a left (or right) invariant vector field, its integral curves are also geodesics (refer to Theorem 12.8.4), and thus these curves also represent the trajectory of the point-like electron. Since the electron carries a negative charge, the opposite direction of the arrows on the integral curves in the figure indicates the direction of the current generated by the point-like electron during motion. The current directions on the left and right halves of the  $u$ -axis, and on the upper and lower halves of the  $v$ -axis, are exactly opposite. When the electron is in a non-uniform magnetic field, these four semi-axes experience Ampère forces.

As shown in the figure, force  $F_1$  points left, force  $F_3$  points right, force  $F_2$  points upward, and force  $F_4$  points downward. Due to the non-uniformity of the magnetic field, these four forces are not equal, resulting in a net magnetic moment. We can reposition the left half of the  $u$ -axis and the lower half of the  $v$ -axis as shown in Figure 20.3.6. These four semi-axes then form a rectangular coil, with the current directions connecting head-to-tail to form a closed loop. When this coil is placed in a non-uniform magnetic field, this closed rectangular loop experiences the effect of a magnetic moment, which explains why the electron possesses a spin magnetic moment.

## §20.4 Transformation from Electron to Proton

### 1. General Transformation Formulas

Electrons and protons are different principal bundles with different structure groups. Their structure groups are generated by different spin operators  $\hat{S}_z$  and  $\hat{S}_x$ , respectively. The integral curves or trajectories of  $\hat{S}_z$  and  $\hat{S}_x$  are the solutions of equations (20.1.16) and (20.1.1), respectively. The transformation relationship between electrons and protons can be found through the solutions of these two sets of equations.

Solving equations (20.1.16) yields

$$u = B_1 e^{\frac{1}{2}ht}, \quad v = B_2 e^{-\frac{1}{2}ht},$$

where  $t \in (-\infty, +\infty) = R$ . Changing coordinate symbols by replacing  $u$  with  $\tilde{u}$  and  $v$  with  $\tilde{v}$ , we obtain

$$\tilde{u} = B_1 e^{\frac{1}{2}ht}, \quad \tilde{v} = B_2 e^{-\frac{1}{2}ht}. \quad (20.4.1)$$

Solving equations (20.1.1) yields

$$u = C_1 e^{\frac{1}{2}ht} + C_2 e^{-\frac{1}{2}ht}, \quad v = C_1 e^{\frac{1}{2}ht} - C_2 e^{-\frac{1}{2}ht}. \quad (20.4.2)$$

The solutions of equations (20.1.1) can also be expressed in the following form:

$$u = C_1 e^{\frac{1}{2}ht} - C_2 e^{-\frac{1}{2}ht}, \quad v = C_1 e^{\frac{1}{2}ht} + C_2 e^{-\frac{1}{2}ht}. \quad (20.4.3)$$

Substituting solutions (20.4.1) into solutions (20.4.3), we obtain the coordinate transformation formulas

$$\begin{cases} u = k_1 \tilde{u} - k_2 \tilde{v}, \\ v = k_1 \tilde{u} + k_2 \tilde{v}, \end{cases} \quad (20.4.4)$$

where

$$k_1 = C_1/B_1, \quad k_2 = C_2/B_2. \quad (20.4.5)$$

Previously, we expressed the spin operators  $\hat{S}_z$  and  $\hat{S}_x$  on the same  $uv$ -plane or on two identical planes. Therefore, equations (20.1.16) and (20.1.1) are also defined on the same  $uv$ -plane or on two identical planes, and their solutions are likewise on the same  $uv$ -plane or on two identical planes. Thus, the  $uv$ -plane and the  $\tilde{u}\tilde{v}$ -plane are either the same plane or two identical planes, so the coordinate transformation formulas (20.4.4) satisfy

$$k_1 = k_2 = 1.$$

This is because if  $k_1$  and  $k_2$  were not equal to 1, the coordinates between the  $uv$ -plane and the  $\tilde{u}\tilde{v}$ -plane would involve scaling, meaning they are not the same plane or two identical planes. Therefore, formula (20.4.4) becomes

$$\begin{cases} u = \tilde{u} - \tilde{v}, \\ v = \tilde{u} + \tilde{v}, \end{cases} \quad (20.4.6)$$

and equation (20.4.5) becomes

$$B_1 = C_1, \quad B_2 = C_2. \quad (20.4.7)$$

Denote the transformation (20.4.6) as  $F$ . Equation (20.4.6) is the reason for the difference between

the spin magnetic moments of the electron and the proton. The Jacobian matrix of transformation (20.4.6) is

$$J = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \quad (20.4.8)$$

Since  $|J| = 2 > 0$ , this transformation is orientation-preserving; thus, the proton and the electron have the same orientation.

The area element of the electron is  $d\tilde{u}d\tilde{v}$ . Using equation (20.4.6), the area element of the proton can be obtained as

$$dudv = |J|d\tilde{u}d\tilde{v} = 2d\tilde{u}d\tilde{v}. \quad (20.4.9)$$

Therefore, the area element of the proton  $dudv$  is 2 times that of the electron  $d\tilde{u}d\tilde{v}$ .

From equation (20.4.6), we have

$$\begin{cases} \frac{\partial}{\partial u} = \frac{\partial}{\partial \tilde{u}} - \frac{\partial}{\partial \tilde{v}}, \\ \frac{\partial}{\partial v} = \frac{\partial}{\partial \tilde{u}} + \frac{\partial}{\partial \tilde{v}}, \end{cases}$$

that is,

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \tilde{u}} \\ \frac{\partial}{\partial \tilde{v}} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix}.$$

Let a vector in the electron be  $(\rho, 0)$ . Using the above relation, the corresponding vector in the proton associated with  $(\rho, 0)$  is

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \rho \\ 0 \end{pmatrix} = \begin{pmatrix} \rho \\ \rho \end{pmatrix}, \quad (20.4.10)$$

which is  $(\rho, \rho)$ .

## 2.Spherical Right Triangle Formula

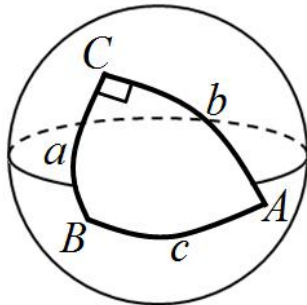


Figure 20.4.1 A spherical right triangle

Consider a right spherical triangle  $ABC$  on a unit sphere, as shown in Figure 20.4.1. Let its three angles be denoted by  $A, B, C$ , with angle  $C$  being the right angle,  $C = \frac{\pi}{2}$ . Let the sides opposite angles  $A, B$ , and  $C$  be  $a, b$ , and  $c$ , respectively. Then the following relation holds:

$$\cos c = \cos a \cos b. \quad (20.4.11)$$

## §20.5 Spin Magnetic Moment of the Proton

We now calculate the spin magnetic moment of the proton. We compare the electron and the proton, deriving the proton's spin magnetic moment from that of the electron.

From equation (20.3.24), the spin magnetic moment of the electron is obtained as

$$\mu_e = I_e A_e = -\frac{e\hbar}{2m_e} = -\frac{e}{m_e} \frac{\hbar}{2}, \quad (20.5.1)$$

where  $I_e$  is the current generated by the point-like electron moving inside the electron itself, and  $A_e$  is the area enclosed by the current loop formed by the motion of the point-like electron.

We can also apply the Dirac equation (20.3.9) to the proton. Replacing the electron's rest

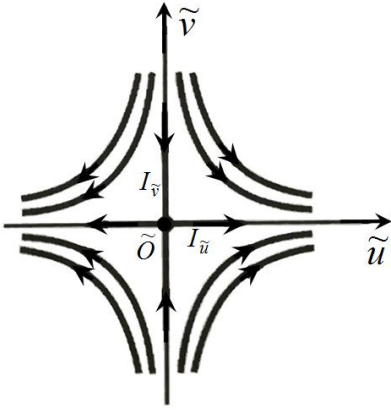
mass  $m_e$  in equation (20.3.9) with the proton's rest mass  $m_p$ , we can obtain a calculation formula for the proton's spin magnetic moment  $\mu_p$  similar to equation (20.3.23), namely

$$\mu_p = \frac{e\hbar}{2m_p} = \frac{e}{m_p} \frac{\hbar}{2}. \quad (20.5.2)$$

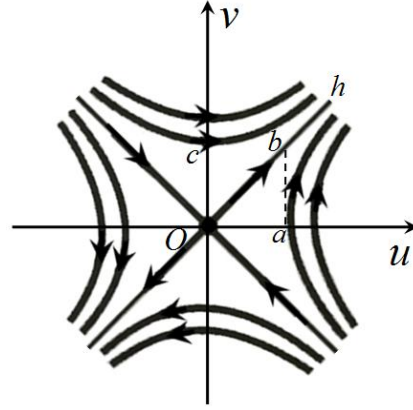
Therefore, for the proton, replacing the electron's rest mass  $m_e$  in equation (20.5.1) with the proton's rest mass  $m_p$ , we get

$$\mu_p = I_p A_p = \frac{e\hbar}{2m_p} = \frac{e}{m_p} \frac{\hbar}{2} = \frac{e}{m_p} S_z. \quad (20.5.3)$$

This is called the **normal spin magnetic moment of the proton**, because the actually measured proton spin magnetic moment is larger than this value. The reason for the difference between the spin magnetic moments of the electron and the proton is that the electron is the principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , while the proton is the principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ . Nevertheless, we can calculate the **anomalous spin magnetic moment of the proton** based on equation (20.5.3), by replacing the two numerical values of current intensity and area in equation (20.5.3).



**Figure 20.5.1** Trajectory for the proton's normal spin magnetic moment



**Figure 20.5.2** Trajectory for the proton's anomalous spin magnetic moment

Based on the similarity between equations (20.5.3) and (20.5.1), it can be considered that when the trajectory diagram of the proton's spin operator is the same as that of the electron's spin operator, the expression (20.5.3) for the proton's normal spin magnetic moment is similar to the expression (20.5.1) for the electron's spin magnetic moment. Therefore, it can be considered that the trajectory diagram of the electron's spin operator (Figure 20.1.7) is also the trajectory diagram when the proton's spin magnetic moment is the normal moment. Thus, we take Figure 20.5.1 as the trajectory diagram when the proton's spin magnetic moment is the normal moment, while Figure 20.5.2 (which is the same as Figure 20.1.2) is the trajectory diagram when the proton's spin magnetic moment is the anomalous moment. Since the proton carries a positive charge, the direction of the trajectory in the diagram is both the direction of motion of the point-like proton and the direction of the current intensity.

Figures 20.5.1 and 20.5.2 are located on two different smooth manifolds, respectively. The transformation from Figure 20.5.1 to Figure 20.5.2 is given by equation (20.4.6). We have already denoted this transformation as  $F$ , where  $F$  is a diffeomorphism.

Since the  $SU(2)$  group is a compact connected Lie group, according to Theorem 6.6.1, the smooth tangent vector field  $\hat{S}_z$  on the  $SU(2)$  group determines a one-parameter differentiable transformation group  $\varphi(t)$ , whose integral curves are shown in Figure 20.1.7 or Figure 20.1.8. For the electron, when  $\tilde{u} \neq 0, \tilde{v} = 0$  or  $\tilde{u} = 0, \tilde{v} \neq 0$  in equation (20.4.1), equation (20.4.1) represents an integral curve passing through the origin.

Setting  $B_1 = 1, B_2 = 1, t = 0$  in equation (20.4.1) yields

$$\tilde{u} = \tilde{v} = 1,$$

which represents the unit element  $\varphi(0)$  of the transformation group  $\varphi(t)$ . In Figure 20.5.1, let the magnitudes of the current intensity flowing through the  $\tilde{u}$ -axis and  $\tilde{v}$ -axis be  $I_{\tilde{u}}$  and  $I_{\tilde{v}}$ , respectively. We can set

$$I_{\tilde{u}} = \tilde{u} = 1, \quad I_{\tilde{v}} = \tilde{v} = 1.$$

Let

$$I_p = I_{\tilde{u}} = I_{\tilde{v}} = 1. \quad (20.5.4)$$

Since the parameter  $t=0$ ,  $I_p$  is precisely the magnitude of the current intensity flowing out from the origin when the proton's spin magnetic moment is the normal moment.

One cannot treat  $I_p=1$  as a tangent vector  $(1, 0)$  and then obtain  $(1, 1)$  via the transformation (20.4.10)

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and simply regard  $(1, 1)$  as the proton's current intensity. This is because the line segment where  $I_p$  lies is on the integral curve within the electron manifold, while tangent vectors reside in the tangent space. Therefore,  $I_p$  should not be viewed as a vector, but rather as the length of a segment of the integral curve (equivalent to the  $\tilde{u}$ -axis or  $\tilde{v}$ -axis in the phase plane).

Since we have already set  $B_1 = B_2 = 1$  in equation (20.4.1), according to equation (20.4.7), we should set  $C_1 = C_2 = 1$  in equation (20.4.2). Further setting  $t=0$  yields

$$u = 2, \quad v = 0,$$

From equation (20.4.3), we get

$$u = 0, \quad v = 2.$$

As shown in Figure 20.5.2, the line segment  $oa$  from the origin  $o$  to point  $a(2,0)$  lies on the  $u$ -axis of Figure 20.5.2, and the line segment  $oc$  from the origin  $o$  to point  $c(0,2)$  lies on the  $v$ -axis of Figure 20.5.2. However, the  $u$ -axis and  $v$ -axis are not integral curves of the proton's spin operator  $\hat{S}_x$ . But the line segments  $oa$ ,  $oc$ , and  $ob$  form a right triangle  $oab$  ( $oc = ab$ ), and the line segment  $ob$  lies on the integral curve of  $\hat{S}_x$ . The length  $d(o,b)$  of the line segment  $ob$  can be taken as the magnitude of the current intensity  $I_a$  corresponding to  $I_{\tilde{u}}$ . That is, when transforming from Figure 20.5.1 to Figure 20.5.2,  $I_{\tilde{u}}$  transforms to  $I_a$ ,

$$I_a = d(o,b).$$

Similarly, since the parameter  $t=0$ ,  $I_a$  is precisely the magnitude of the current intensity flowing out from the origin when the proton's spin magnetic moment is the anomalous spin magnetic moment.

Since  $\triangle oab$  is on the smooth manifold of the proton, calculating the length  $I_a$  requires using the unit spherical right triangle formula (20.4.11), yielding

$$I_a = \arccos[\cos(oa)\cos(oc)] = \arccos[\cos(2)\cos(2)]. \quad (20.5.5)$$

In equation (20.5.3),  $A_p$  is the area enclosed by the current loop for the proton's normal spin magnetic moment. Assume  $A_p$  equals the unit area, i.e., let

$$A_p = \iint d\tilde{u}d\tilde{v} = \int_0^1 d\tilde{u} \int_0^1 d\tilde{v}. \quad (20.5.6)$$

Assume the area  $A_a$  enclosed by the current loop for the proton's anomalous spin magnetic moment, as depicted in Figure 20.5.2, is also the unit area in Figure 20.5.2, i.e., let

$$A_a = \iint du dv = \int_0^1 du \int_0^1 dv. \quad (20.5.7)$$

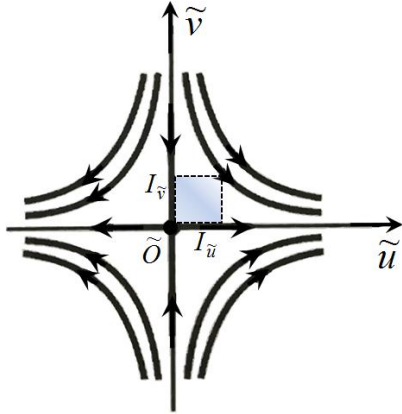
Integrating both sides of equation (20.4.9), we get

$$\iint du dv = \iint |J| d\tilde{u} d\tilde{v} = \iint 2 d\tilde{u} d\tilde{v} . \quad (20.5.8)$$

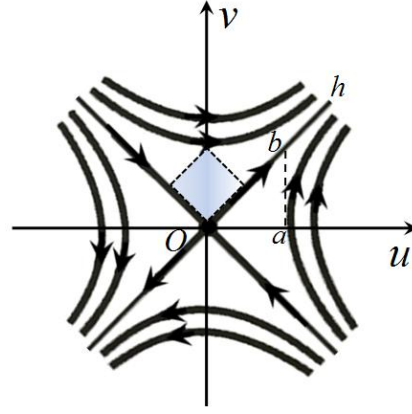
Integrating both sides of equation (20.5.8) simultaneously from 0 to 1 for the variables yields the relationship between  $A_p$  and  $A_a$ :

$$A_a = \int_0^1 du \int_0^1 dv = 2 \int_0^1 d\tilde{u} \int_0^1 d\tilde{v} = 2A_p . \quad (20.5.9)$$

Therefore, the area  $A_a$  enclosed by the current loop for the proton's anomalous spin magnetic moment is twice the area  $A_p$  enclosed by the current loop for the proton's normal spin magnetic moment.



**Figure 20.5.3** Rectangular loop for the proton's normal spin magnetic moment



**Figure 20.5.4** Rectangular loop for the proton's anomalous spin magnetic moment

As shown in Figures 20.5.3 and 20.5.4, the rectangular loops in both figures are squares of unit area in their respective coordinate systems. However, the area of the rectangular loop in Figure 20.5.4 is twice the area of the rectangular loop in Figure 20.5.3. Simultaneously, the current intensity flowing through the rectangular loop in Figure 20.5.4 is  $\arccos[\cos(2)\cos(2)]$  times that flowing through the rectangular loop in Figure 20.5.3. Therefore, from equations (20.5.4), (20.5.5), and (20.5.9), the anomalous spin magnetic moment of the proton is obtained as

$$\mu_a = I_a A_a = (\arccos[\cos(2)\cos(2)]) I_p \cdot (2A_p) = 2 \arccos[\cos(2)\cos(2)] I_p A_p . \quad (20.5.10)$$

Then, from equation (20.5.3), we have

$$\mu_a = I_a A_a = 2 \arccos[\cos(2)\cos(2)] \frac{e\hbar}{2m_p} = 2.79348123991637 \frac{e\hbar}{2m_p} . \quad (20.5.11)$$

Since

$$\mu_N = \frac{e\hbar}{2m_p} , \quad (20.5.12)$$

where  $\mu_N$  is called the nuclear magneton, we obtain

$$\mu_a = 2.79348123991637 \mu_N . \quad (20.5.13)$$

The experimentally measured value of the proton's anomalous spin magnetic moment is  $\mu_{p,\text{exp}} = 2.79284734462 \mu_N$ . Therefore, the difference between the calculated value and the experimental value is

$$\mu_a - \mu_{p,\text{exp}} = 0.0006338952 \ 9636515 \ \mu_N ,$$

$$\left( \frac{\mu_a}{\mu_{p,\text{exp}}} - 1 \right) \times 100\% = 0.0226970\% .$$

The theoretical calculated value is very close to the experimental value.

Deriving from the Dirac equation, the spin gyromagnetic ratio of the electron is twice its orbital gyromagnetic ratio. Therefore, as long as we can experimentally measure the electron's



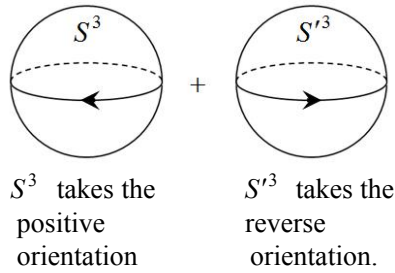
orbital gyromagnetic ratio, we can calculate the electron's spin gyromagnetic ratio, and then calculate the electron's rest mass. In this section, by deriving the proton's anomalous spin magnetic moment from the electron's spin magnetic moment, knowing the proton's anomalous spin magnetic moment allows us to calculate the proton's rest mass. Thus, from this perspective, we are theoretically capable of calculating the rest masses of both the electron and the proton. In other words, the three physical quantities—spin, spin magnetic moment, and rest mass—of the electron and positron (or proton and antiproton) are closely linked. Knowing the value of the spin magnetic moment also reveals the rest mass.

## §20.6 Spin Magnetic Moment of the Neutron

According to our first assumption, the neutron  $n$  is composed of its decay products which are the proton  $p$ , the electron  $e^-$ , and the antineutrino  $\bar{\nu}_e$ . Although the neutron is electrically neutral, experiments show that it possesses a spin magnetic moment, which is negative in value. In this section, we calculate the spin magnetic moment of the neutron.

### 1. The Proton's Spin Magnetic Moment within the Neutron is Halved

Both the proton and the antineutrino are  $SU(2)$  groups. To distinguish them, the proton's  $SU(2)$  group is still denoted as the  $SU(2)$  group, while the antineutrino's  $SU(2)$  group is denoted as the  $SU(2)'$  group. Within the neutron, the proton and the antineutrino form a direct inner product group  $SU(2) \otimes SU(2)'$ . We have already considered the proton as a left-handed oriented particle and the antineutrino as a right-handed oriented particle. Due to their opposite orientations,



**Figure 20.6.1** Two objects  $S^3$  with opposite orientations combine to form one object  $RP^3$

according to the analysis in §16.13, the three-dimensional sphere  $S^3_\otimes$ , which is homeomorphic to the direct inner product group  $SU(2) \otimes SU(2)'$  formed by the two oppositely oriented  $SU(2)$  and  $SU(2)'$  groups, has its antipodal points identified. Therefore,  $S^3_\otimes$  can be regarded as a projective space  $RP^3$ . That is, the topological structure of this inner product group  $SU(2) \otimes SU(2)'$  is  $RP^3$ , which is isomorphic to the  $SO(3)$  group:

$$SU(2) \otimes SU(2)' \cong SO(3).$$

Thus, because the proton and antineutrino combine to form an  $RP^3$  space, within this  $RP^3$  space, the point-like proton only needs to rotate through an angle of  $2\pi$  to complete one full turn, which is half the rotation required in the free state (where it must rotate  $4\pi$  in  $S^3$ ). This results in the current loop area for the spin magnetic moment of the proton within the neutron being only half of that in the free state. Therefore, the spin magnetic moment of the proton within the neutron should be half of its free-state value, i.e.,

$$\mu_p = 2.7934812399 \ 1637 \times \frac{1}{2} \mu_N. \quad (20.6.1)$$

The electron is also a component of the neutron. Within the neutron, the electron also forms a direct inner product group together with the proton and antineutrino. Let the group representing the electron be the  $SU(2)''$  group.  $SU(2)''$  also forms a direct inner product group with  $SU(2) \otimes SU(2)'$ :

$$SU(2) \otimes SU(2)' \otimes SU(2)'' \cong SU(2).$$

Therefore, the neutron is also an  $SU(2)$  group and also a smooth manifold of  $S^3$ . Let  $g \in SU(2)$ ,  $g' \in SU(2)'$ ,  $g'' \in SU(2)''$ , then

$$(g, g', g'') \in SU(2) \otimes SU(2)' \otimes SU(2)'' . \quad (20.6.2)$$

$(g, g', g'')$  becomes a point within the neutron; that is, the neutron is a smooth manifold of  $S^3$ , where each point in  $S^3$  is composed of three group elements  $(g, g', g'')$ .

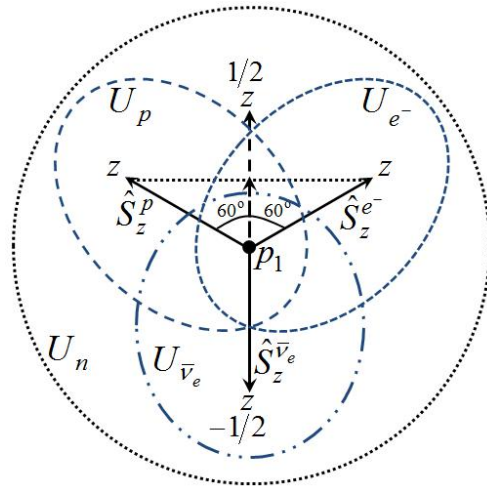
## 2. Angles Between the z-Axis Components of Spin for the Proton, Electron, and Antineutrino within the Neutron

The neutron is an inner direct product group  $n = p \otimes e^- \otimes \bar{\nu}_e$  composed of a proton, an electron, and an antineutrino, each with spin  $-\hbar/2$ . The resulting neutron also has spin  $-\hbar/2$ . For a neutron composed of a proton, electron, and antineutrino, each with spin  $-\hbar/2$ , to also have spin  $-\hbar/2$ , the angles between the z-axis components of the spin vectors of these three particles must all be  $120^\circ$ , as shown in Figure 20.6.2.

Figure 20.6.2 represents the neutron smooth manifold  $n$ . In Figure 20.6.2, the point  $p_1$  lies on both the neutron manifold  $n$  and the manifolds of the proton  $p$ , electron  $e^-$ , and antineutrino  $\bar{\nu}_e$ , respectively. In the tangent space at each point of this neutron smooth manifold  $n$ , there exist the z-axis component vectors of the spin operators for the proton  $p$ , electron  $e^-$ , and antineutrino  $\bar{\nu}_e$ , which intersect each other at angles of  $120^\circ$ .

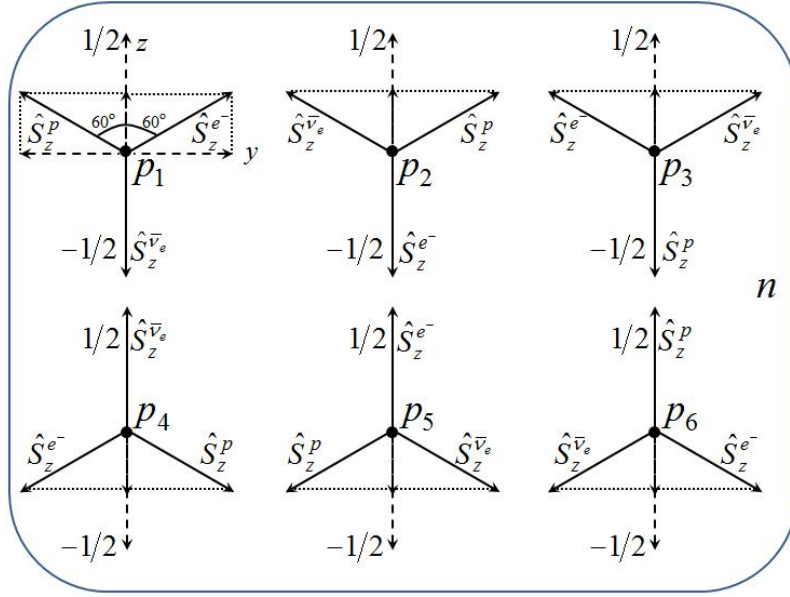
For example, the local coordinate system of the proton  $p$  at point  $p_1$  on its own manifold is  $(U_p; x, y, z)$ ; In this local coordinate system, the z-axis component vector of the proton's spin is  $\hat{S}_z^p$ . The local coordinate system of the electron  $e^-$  at point  $p_1$  on its own manifold is  $(U_{e^-}; x, y, z)$ ; In this local coordinate system, the z-axis component vector of the electron's spin is  $\hat{S}_z^{e^-}$ . The local coordinate system of the antineutrino  $\bar{\nu}_e$  at point  $p_1$  on its own manifold is  $(U_{\bar{\nu}_e}; x, y, z)$ ; In this local coordinate system, the z-axis component vector of the antineutrino's spin is  $\hat{S}_z^{\bar{\nu}_e}$ . However,  $\hat{S}_z^p$ ,  $\hat{S}_z^{e^-}$ , and  $\hat{S}_z^{\bar{\nu}_e}$  are all also located in the tangent space  $T_{p_1}n$  at the same point of the neutron manifold. Let the local coordinate system of the neutron  $n$  at point  $p_1$  on its own manifold be  $(U_n; x, y, z)$ .

At point  $p_1$ ,  $\hat{S}_z^p$  and  $\hat{S}_z^{e^-}$  intersect at an angle of  $120^\circ$  in the local coordinate system  $(U_n; x, y, z)$ . Their projections onto the y-axis (transverse direction) are equal in magnitude but opposite in direction, summing to zero; the sum of their projections onto the z-axis equals  $\hbar/2$ , which is exactly equal in magnitude but opposite in direction to the z-axis component vector  $\hat{S}_z^{\bar{\nu}_e}$  of the antineutrino  $\bar{\nu}_e$  in the local coordinate system  $(U_n; x, y, z)$ . Therefore, if the z-axis component of the neutron's spin is measured near point  $p_1$ , it could yield either  $+\hbar/2$  or  $-\hbar/2$ . The value  $+\hbar/2$  is obtained when the measurement captures the sum of the projections of  $\hat{S}_z^p$  and  $\hat{S}_z^{e^-}$  onto the positive z-direction at point  $p_1$  in the figure; the value  $-\hbar/2$  is obtained when the measurement captures the value of the vector  $\hat{S}_z^{\bar{\nu}_e}$  below point  $p_1$  in the figure.



**Figure 20.6.2** Relationship between the z-axis components of spin for the three particles





**Figure 20.6.3** The six permutations of the arrangement of spin operator components within the neutron

There are a total of six possible arrangement distributions for the  $z$ -axis component vectors of the spin operators of the proton  $p$ , electron  $e^-$ , and antineutrino  $\bar{\nu}_e$  in the tangent space at each point of the neutron manifold  $n$ , as shown in Figure 20.6.3.

Point  $p_1$  in Figure 20.6.3 is the same as in Figure 20.6.2. In the local coordinate system  $(U_n; x, y, z)$  at point  $p_2$ , the  $z$ -axis component vector  $\hat{S}_z^{\bar{\nu}_e}$  of the antineutrino's spin intersects the  $z$ -axis component vector  $\hat{S}_z^p$  of the proton's spin at an angle of  $120^\circ$ . The sum of the projections of  $\hat{S}_z^{\bar{\nu}_e}$  and  $\hat{S}_z^p$  onto the positive  $z$ -direction equals  $\hbar/2$ , which is exactly equal in magnitude but opposite in direction to the  $z$ -axis component vector  $\hat{S}_z^{e^-}$  of the electron  $e^-$  in the negative  $z$ -direction. Therefore, if the  $z$ -axis component of the neutron's spin is measured near point  $p_2$ , it could yield either  $+\hbar/2$  or  $-\hbar/2$ . The value  $+\hbar/2$  is obtained when the measurement captures the sum of the projections of  $\hat{S}_z^{\bar{\nu}_e}$  and  $\hat{S}_z^p$  onto the positive  $z$ -direction at point  $p_2$  in the figure; the value  $-\hbar/2$  is obtained when the measurement captures the value of the vector  $\hat{S}_z^{e^-}$  below point  $p_2$  in the figure.

Similarly, in the local coordinate system  $(U_n; x, y, z)$  at point  $p_3$  in Figure 20.6.3, the  $z$ -axis component vector  $\hat{S}_z^{e^-}$  of the electron's spin intersects the  $z$ -axis component vector  $\hat{S}_z^{\bar{\nu}_e}$  of the antineutrino's spin at an angle of  $120^\circ$ . The sum of the projections of  $\hat{S}_z^{e^-}$  and  $\hat{S}_z^{\bar{\nu}_e}$  onto the positive  $z$ -direction equals  $+\hbar/2$ , which is exactly equal in magnitude but opposite in direction to the  $z$ -axis component vector  $\hat{S}_z^p$  of the proton  $p$  in the negative  $z$ -direction. Therefore, if the  $z$ -axis component of the neutron's spin is measured near point  $p_3$ , it could yield either  $+\hbar/2$  or  $-\hbar/2$ . The value  $+\hbar/2$  is obtained when the measurement captures the sum of the projections of  $\hat{S}_z^{e^-}$  and  $\hat{S}_z^{\bar{\nu}_e}$  onto the positive  $z$ -direction at point  $p_3$  in the figure; the value  $-\hbar/2$  is obtained when the measurement captures the value of the vector  $\hat{S}_z^p$  below point  $p_3$  in the figure.

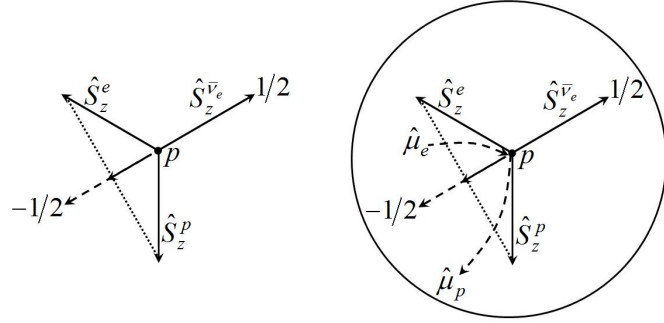
Therefore, assuming that at each point of the neutron, the  $z$ -axis component vectors of the spin operators for the proton  $p$ , electron  $e^-$ , and antineutrino  $\bar{\nu}_e$  in the tangent space intersect each other at angles of  $120^\circ$ , then the angular momentum of these three particles is conserved in the  $z$ -direction. Simultaneously, measuring the  $z$ -axis component of the neutron's spin yields either

$+\hbar/2$  or  $-\hbar/2$ , consistent with experiments. Hence, this assumption will be used in the following calculation of the neutron's spin magnetic moment.

### 3.The Spin Magnetic Moment of the Neutron is Negative

The antineutrino  $\bar{\nu}_e$  is a component of the neutron. Since the  $z$ -axis component vector of the antineutrino's spin has only one direction, in its free state this direction always maintains a right-handed helical relationship with its direction of motion, with an eigenvalue of  $S_z^{\bar{\nu}_e} = \hbar/2$ . Therefore, within the neutron, the  $z$ -axis component vector of the antineutrino's spin can be taken as the positive direction.

In the free state, the proton's spin magnetic moment  $\hat{\mu}_p$  is in the same direction as its  $z$ -axis spin component vector and is a positive value. However, when the proton becomes a component of the neutron, it is no longer a free particle, and its spin magnetic moment is influenced by the other particles within the neutron. Since the  $z$ -axis component vectors of the proton and the antineutrino within the neutron intersect at an



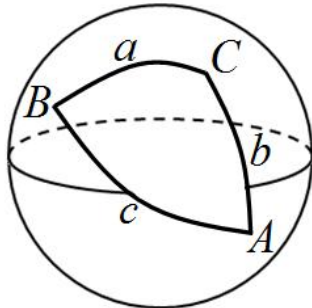
**Figure 20.6.4** Magnetic moments of the proton and electron within the neutron

angle of  $120^\circ$ , and we take the antineutrino's  $z$ -axis spin component as the positive direction within the neutron, if the proton's spin magnetic moment is projected onto the antineutrino's  $z$ -axis, it can only project onto the negative direction of the antineutrino's spin  $z$ -axis, resulting in a negative projection value, as shown in Figure 20.6.4. Consequently, within the neutron, the proton's spin magnetic moment must be negative.

When the electron is in its free state, its spin magnetic moment  $\hat{\mu}_e$  is negative. Since the directions of the electron's and proton's magnetic moments are opposite, within the neutron, when the proton's spin magnetic moment projects onto the negative direction of the antineutrino's spin  $z$ -axis, the projection of the electron's spin magnetic moment  $\hat{\mu}_e$  onto the antineutrino's  $z$ -axis lies in the positive direction of that axis. Therefore, within the neutron, the electron's spin magnetic moment is positive, as shown in Figure 20.6.4.

Because within the neutron the proton's spin magnetic moment is negative, the electron's spin magnetic moment is positive, and the absolute value of the proton's spin magnetic moment is greater than that of the electron's, the combined magnetic moment of the proton and electron is negative. Thus, the spin magnetic moment of the neutron is negative.

### 4.Spherical Law of Cosines



**Figure 20.6.5** Triangle on a sphere

**Spherical Triangle Side Cosine Theorem** As shown in Figure 20.6.5, consider an arbitrary triangle  $ABC$  on a unit sphere. Let its three interior angles be denoted by  $A$ ,  $B$ , and  $C$ , and let the sides opposite angles  $A$ ,  $B$ , and  $C$  be  $a$ ,  $b$ , and  $c$ , respectively. The following relationships hold:

$$\begin{aligned}\cos a &= \cos b \cos c + \sin b \sin c \cos A, \\ \cos b &= \cos c \cos a + \sin c \sin a \cos B, \\ \cos c &= \cos a \cos b + \sin a \sin b \cos C.\end{aligned}\tag{20.6.3}$$

### 5.The Neutron as a Curved Manifold

According to equation (20.3.1), the magnetic moment is the product of the current intensity

and the area enclosed by a closed current loop. Based on the previous discussion, we interpret the magnetic moment as a curve on a smooth manifold. The spin magnetic moments of the proton and the electron are two intersecting curves on the neutron manifold. Since the neutron is a Lie group, according to Theorem 12.8.4, the sectional curvature at each point of the neutron is greater than zero—that is, the neutron is curved at every point. Therefore, we must use the cosine theorem for curved space, namely formula (20.6.3), to calculate the combined magnetic moment formed by the spin magnetic moments of the proton and the electron, i.e., the spin magnetic moment of the neutron  $\mu_n$ .

## 6.Units for the Spin Magnetic Moments of the Proton and Electron within the Neutron

To calculate the combined magnetic moment formed by the spin magnetic moments of the proton and the electron using the cosine theorem, it is first necessary to know the spin magnetic moments of the proton and the electron. Since we assume that within the neutron, the proton and the electron are tightly bound together, if the Dirac equation (20.3.9) is applied to the electron within the neutron, the electron's rest mass should be replaced by the neutron's rest mass. That is,  $m_e$  in equation (20.3.9) should be replaced by the neutron's rest mass  $m_n$ . Therefore, the calculated spin magnetic moment of the electron should be

$$\mu'_e = -\frac{e\hbar}{2m_n}. \quad (20.6.4)$$

Similarly, if the Dirac equation (20.3.9) is applied to the proton within the neutron, the proton's rest mass should also be replaced by the neutron's rest mass. That is,  $m_e$  in equation (20.3.9) should be replaced by the neutron's rest mass  $m_n$ . Therefore, the calculated spin magnetic moment of the proton should be

$$\mu'_p = \frac{e\hbar}{2m_n}.$$

Then, using the method from §20.5, the proton's spin magnetic moment is calculated as

$$\mu'_p = 2.7934812399 \ 1637 \frac{e\hbar}{2m_n}. \quad (20.6.5)$$

In summary, for both the electron and the proton within the neutron, the unit for their spin magnetic moments should be changed to  $\frac{e\hbar}{2m_n}$ .

## 7.Two States of the Neutron's Spin Magnetic Moment

To calculate the combined spin magnetic moment formed by the proton and the electron using the cosine theorem, it is also necessary to know the angle between the spin magnetic moments of the proton and the electron. Since the tangent space at each point of the neutron is a Euclidean flat space, the  $120^\circ$  angle between the  $z$ -axis components of the proton's and electron's spins is an angle measured in the Euclidean plane. However, to compute the combined spin magnetic moment of the proton and the electron, one must know the angle between their spin magnetic moments **on the neutron manifold**, because the point-like proton and point-like electron move on the neutron manifold, not in the tangent space. The angle between the two magnetic moments is the angle between the two magnetic-moment curves on the manifold.

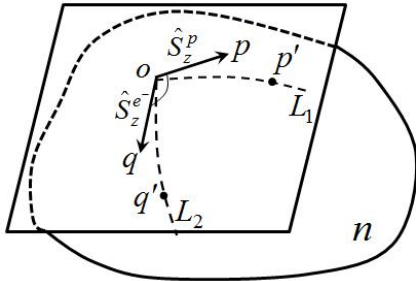
### (1)Hinge State

The direction of the proton's spin magnetic moment along its  $z$ -axis is the same as the direction of its  $z$ -axis spin component  $\hat{S}_z^p$ , but the direction of the electron's spin magnetic moment along its  $z$ -axis is opposite to that of its  $z$ -axis spin component  $\hat{S}_z^e$ . As shown in Figure 20.6.6, suppose there are two geodesics  $L_1$  and  $L_2$  on the neutron smooth manifold  $n$ , intersecting at point  $o$  with an angle equal to  $120^\circ$ . The direction of segment  $L_2$  is from point  $q'$  to point  $o$ , and the direction of segment  $L_1$  is from point  $o$  to point  $p'$ . The starting point  $o$  of segment  $L_1$  is the endpoint of segment  $L_2$ , so segments  $op'$  and  $oq'$  form a hinge

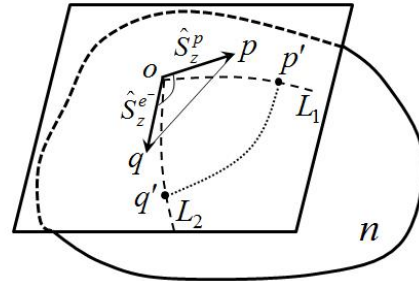
$(o, op', oq')$ . The length of segment  $op'$  on geodesic  $L_1$  is  $L_{op'}$ ,  $L_{op'} = \frac{1}{2}$ , and the length of segment  $oq'$  on geodesic  $L_2$  is  $L_{oq'}$ ,  $L_{oq'} = \frac{1}{2}$ .

Since the sectional curvature of the neutron manifold is greater than zero, according to the Toponogov comparison theorem, there exists a comparison hinge  $(o, op, oq)$  in the tangent space at point  $o$  such that the length of tangent vector  $\hat{S}_z^p$  equals  $|\hat{S}_z^p| = L_{op'} = \frac{1}{2}$ , the length of tangent vector  $\hat{S}_z^{e-}$  equals  $|\hat{S}_z^{e-}| = L_{oq'} = \frac{1}{2}$ , the angle between tangent vectors  $\hat{S}_z^p$  and  $\hat{S}_z^{e-}$  in this Euclidean tangent space equals  $120^\circ$ , the angle between geodesics  $L_1$  and  $L_2$  also equals  $120^\circ$ , and the distance between points  $p'$  and  $q'$  is less than the distance between points  $p$  and  $q$ , i.e.,  $d(p', q') < d(p, q)$ . The distance between points  $p$  and  $q$  is

$$L_{pq} = d(p, q) = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 - 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\cos\left(\frac{2\pi}{3}\right)} = \frac{\sqrt{3}}{2}.$$



**Figure 20.6.6** Hinge state



**Figure 20.6.7** Triangle state

It can be considered that the hinge  $(o, op', oq')$  on the neutron manifold at point  $o$  is the image of the hinge  $(o, op, oq)$  in the tangent space at  $o$  under the exponential map. That is, one may consider that through the exponential map, the tangent vector  $\hat{S}_z^p$  is mapped to the geodesic segment  $op'$  on the neutron manifold, and the tangent vector  $\hat{S}_z^{e-}$  is mapped to the geodesic segment  $oq'$  on the neutron manifold. Since the exponential map is locally a diffeomorphism (see Theorem 12.5.7), it is also invertible. Therefore, conversely, one may consider that through the inverse of the exponential map, the geodesic segment  $op'$  on the neutron manifold is mapped to the tangent vector  $\hat{S}_z^p$  in the tangent space, and the geodesic segment  $oq'$  is mapped to the tangent vector  $\hat{S}_z^{e-}$  in the tangent space.

Comparing with the  $z$ -axis spin components of the proton and electron, it can be considered that the tangent vector  $\hat{S}_z^p$  corresponds to the proton's  $z$ -axis spin component and  $\hat{S}_z^{e-}$  corresponds to the electron's  $z$ -axis spin component, while the geodesic segments  $op'$  and  $oq'$  represent the spin magnetic moments of the proton and electron, respectively. The angle between the  $z$ -axis spin components of the proton and electron equals  $120^\circ$ , so the angle between the  $z$ -axis components of the proton's and electron's spin magnetic moments on the neutron manifold also equals  $120^\circ$ . Knowing this angle allows the calculation of the combined magnetic moment of the proton's and electron's spin magnetic moments along the  $z$ -axis.

The proton, electron, and neutron are all homeomorphic to the three-dimensional unit sphere, so they are all constant-curvature manifolds. Therefore, we can use the cosine theorem mentioned earlier to compute the neutron's magnetic moment. From equation (20.6.3) we obtain

$$\cos a = \cos \mu_e \cos \mu_p + \sin \mu_e \sin \mu_p \cos\left(\frac{2\pi}{3}\right),$$

hence

$$\mu_n = \arccos(\cos a) = \arccos\left(\cos \mu_e \cos \mu_p + \sin \mu_e \sin \mu_p \cos\left(\frac{2\pi}{3}\right)\right). \quad (20.6.6)$$

Substituting the following numerical values

$$\mu_e = 1, \quad \mu_p = 2.79348123991637 \times \frac{1}{2}, \quad \pi = 3.14159265358979 \quad (20.6.7)$$

into equation (20.6.6), the magnitude of the neutron's spin magnetic moment is calculated as

$$\mu'_{n,\text{cal}} = 1.89738070129692 \frac{e\hbar}{2m_n}. \quad (20.6.8)$$

It should be noted that equation (20.6.3) is the cosine theorem for the **unit sphere**, so the data substituted into equation (20.6.6) must be data on the three-dimensional unit sphere (refer to the discussion in Chapter 11 on the proof of the cosine theorem for spherical-triangle sides); i.e., we must set  $\frac{e\hbar}{2m_n} = 1$ .

## (2) Triangular State

As shown in Figure 20.6.7, there is a tangent plane at point  $o$  on the neutron manifold. The tangent vector  $\hat{S}_z^p$  represents the proton's spin operator with length  $|\hat{S}_z^p| = \frac{1}{2}$ , and the tangent vector  $\hat{S}_z^e$  represents the electron's spin operator with length  $|\hat{S}_z^e| = \frac{1}{2}$ . Let the length between endpoints  $p$  and  $q$  be  $L_{pq}$ . Points  $o$ ,  $p$ , and  $q$  form a triangle  $\Delta opq$ .

Since the sectional curvature of the neutron manifold is greater than zero, according to the Toponogov comparison theorem, in the neighborhood of point  $o$  on the neutron manifold, there exists a triangle  $\Delta op'q'$  composed of three geodesic segments  $op'$ ,  $oq'$ , and  $p'q'$ . The length of  $op'$  is  $L_{op'} = \frac{1}{2}$ , the length of  $oq'$  is  $L_{oq'} = \frac{1}{2}$ , and the length of  $p'q'$  is  $L_{p'q'} = L_{pq} = \frac{\sqrt{3}}{2}$ , meaning that the corresponding sides of triangle  $\Delta op'q'$  and triangle  $\Delta opq$  are equal; that is,

$$L_{op'} = |\hat{S}_z^p| = \frac{1}{2}, \quad L_{oq'} = |\hat{S}_z^e| = \frac{1}{2}, \quad L_{p'q'} = L_{pq} = \frac{\sqrt{3}}{2}.$$

But each interior angle of  $\Delta op'q'$  is **larger** than the corresponding angle of  $\Delta opq$ ; for example,  $\angle p'oq' > \angle poq$ .

The size of  $\angle p'oq' = \varphi$  on the neutron manifold can be calculated using equation (20.6.3). From (20.6.3) we have

$$\cos(\sqrt{3}/2) = \cos(1/2)\cos(1/2) + \sin(1/2)\sin(1/2)\cos\varphi$$

hence

$$\begin{aligned} \varphi &= \arccos\left(\frac{\cos(\sqrt{3}/2) - \cos(1/2)\cos(1/2)}{\sin(1/2)\sin(1/2)}\right) \\ &= 2.131819928951840 \\ &= 122.14428461082^\circ. \end{aligned} \quad (20.6.9)$$

This shows that the angle  $\varphi$  is indeed larger on the curved manifold than in the tangent space. Again, note that equation (20.6.3) is the cosine theorem for the **unit sphere**, so the data substituted into (20.6.9) must be data on the three-dimensional unit sphere; i.e., we must set  $\hbar = 1$ .

The geodesic segments  $op'$  and  $oq'$  are, respectively, the spin magnetic moments of the proton and electron. Using the present value of  $\varphi$ , we now recalculate the neutron's spin magnetic moment via equation (20.6.3). Substituting the numerical values from (20.6.7) and the angle  $\varphi$  from (20.6.9) into

$$\mu_n = \arccos(\cos \alpha) = \arccos(\cos \mu_e \cos \mu_p + \sin \mu_e \sin \mu_p \cos \varphi), \quad (20.6.10)$$

we obtain the magnitude of the neutron's spin magnetic moment as

$$\mu_{n,\text{cal}}'' = 1.92556569152266 \frac{e\hbar}{2m_n}. \quad (20.6.11)$$

Once more, it must be emphasized that equation (20.6.3) is the cosine theorem for the unit sphere; therefore, the data used in (20.6.10) must correspond to the three-dimensional unit sphere; i.e., we

must set  $\frac{e\hbar}{2m_n} = 1$ .

### 8. Average Value of the Neutron's Spin Magnetic Moment

Since in the tangent spaces of Figure 20.6.6 and Figure 20.6.7, the corresponding sides of the two triangles are equal, these two triangles are congruent. Consequently, their corresponding angles  $\angle poq$  are equal, both being  $120^\circ$ . Therefore, it can be considered that the same triangle  $\Delta opq$  in the tangent space at point  $o$  corresponds to two states on the neutron manifold: the hinge state  $(o, op', oq')$  and the triangle state  $\Delta op'q'$ . The experimentally measured value of the neutron's spin magnetic moment is then the average value  $\bar{\mu}_{n,\text{cal}}$  of the neutron's spin magnetic moment in these two states. Below, three methods are used to calculate this average.

#### (1) Arithmetic Mean Method

We first use the simplest arithmetic mean method to find the average  $\bar{\mu}_{n,\text{cal}}$ . Adding the values from equation (20.6.8) and equation (20.6.11) and dividing by 2 gives the arithmetic mean of the neutron's spin magnetic moment:

$$\bar{\mu}_{n,\text{cal}} = (\mu'_{n,\text{cal}} + \mu''_{n,\text{cal}}) \times \frac{1}{2} = 1.91147319640979 \frac{e\hbar}{2m_n}. \quad (20.6.12)$$

The  $\bar{\mu}_{n,\text{cal}}$  calculated above are in units of  $\frac{e\hbar}{2m_n}$ . To compare with experimental values, we need to convert them to units of the nuclear magneton  $\frac{e\hbar}{2m_p}$ .

The neutron rest mass  $m_n = 939.56563 \text{ MeV}/c^2$  and proton rest mass  $m_p = 938.27231 \text{ MeV}/c^2$  are known. The ratio of the neutron to proton rest mass is

$$\frac{m_n}{m_p} = 1.00137840580631. \quad (20.6.13)$$

Multiplying the calculated  $\bar{\mu}_{n,\text{cal}}$  by  $\frac{m_n}{m_p}$  yields the average spin magnetic moment in units of  $\frac{e\hbar}{2m_p}$ :

$$\begin{aligned} \bar{\mu}_{n,\text{cal}} &= 1.91147319640979 \frac{e\hbar}{2m_n} \times \frac{m_n}{m_p} \\ &= 1.91147319640979 \times 1.00137840580631 \frac{e\hbar}{2m_p} = 1.91410798216232 \frac{e\hbar}{2m_p}. \end{aligned}$$

The experimentally measured absolute value of the neutron's spin magnetic moment is

$$\mu_{n,\text{exp}} = 1.91304275 \frac{e\hbar}{2m_p},$$

The difference between the calculated and experimental values is

$$\bar{\mu}_{n,\text{cal}} - \mu_{n,\text{exp}} = 1.91410798216232 - 1.91304275 = 0.00106523216231702,$$



and the relative error is

$$\left( \frac{\bar{\mu}_{n,\text{cal}}}{\mu_{n,\text{exp}}} - 1 \right) \times 100\% = 0.0556826115\%.$$

### (2) Harmonic Mean Method

We now use the harmonic mean method to calculate the average value  $\bar{\mu}_{n,\text{cal}}$ . Substituting the values from equations (20.6.8) and (20.6.11) into the following formula, the average spin magnetic moment of the neutron can be calculated as

$$\bar{\mu}_{n,\text{cal}} = \frac{2}{1/\mu'_{n,\text{cal}} + 1/\mu''_{n,\text{cal}}} = 1.91136929831753 \frac{e\hbar}{2m_n}. \quad (20.6.14)$$

Multiplying  $\bar{\mu}_{n,\text{cal}}$  by  $\frac{m_n}{m_p}$  yields the average spin magnetic moment in units of  $\frac{e\hbar}{2m_p}$ :

$$\begin{aligned} \bar{\mu}_{n,\text{cal}} &= 1.91136929831753 \frac{e\hbar}{2m_n} \times \frac{m_n}{m_p} \\ &= 1.91136929831753 \times 1.00137840580631 \frac{e\hbar}{2m_p} = 1.91400394085633 \frac{e\hbar}{2m_p}. \end{aligned}$$

The difference between the calculated and experimental values is

$$\bar{\mu}_{n,\text{cal}} - \mu_{n,\text{exp}} = 1.91400394085633 - 1.91304275 = 0.000961190856327399,$$

and the error is

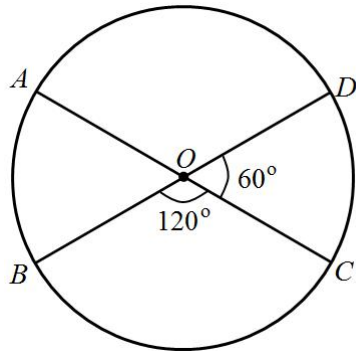
$$\left( \frac{\bar{\mu}_{n,\text{cal}}}{\mu_{n,\text{exp}}} - 1 \right) \times 100\% = 0.050244\%.$$

The arithmetic mean method treats the probabilities of occurrence for the hinge state  $(o, op', oq')$  and the triangle state  $\Delta op'q'$  as equal, each being 0.5. The harmonic mean method yields an average that leans toward the smaller value, implying that the smaller value has a higher probability of occurrence than the larger one. From the results of the two averaging methods above, it can be seen that the harmonic mean gives a  $\bar{\mu}_{n,\text{cal}}$  closer to the experimental value. This indicates that the probabilities of occurrence for the two states are not equal—the hinge state  $(o, op', oq')$  has a higher probability of occurrence than the triangle state  $\Delta op'q'$ .

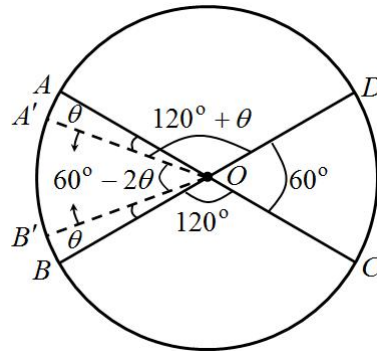
### (3) Classical Probability Method

**Problem 1** Figure 20.6.8 represents the smooth manifold of the neutron. At any point  $O$  on the neutron's smooth manifold, four line segments  $OA$ ,  $OB$ ,  $OC$ , and  $OD$  are connected, with  $OA$  and  $OD$  intersecting at an angle of  $120^\circ$ , and  $OB$  and  $OC$  also intersecting at an angle of  $120^\circ$ . The question is: What is the probability that at this point  $O$ , the pairs  $OA$  and  $OD$ , and  $OB$  and  $OC$ , transform into intersecting angles of  $120^\circ + \theta$ ? From equation (20.6.9),

$$\theta = 122.14428461082^\circ - 120^\circ = 2.14428461082^\circ. \quad (20.6.15)$$



**Figure 20.6.8** Angles intersecting at  $120^\circ$



**Figure 20.6.9** Transformed into angles of  $120^\circ + \theta$

**Solution** Let the event “ $OA$  and  $OD$ ,  $OB$  and  $OC$  intersect at angles of  $120^\circ$ ” be denoted as  $H$ . Suppose the probability of event  $H$  occurring is  $P(H)$ . Let the event “ $OA$  and  $OD$ ,  $OB$  and  $OC$  transform into intersecting angles of  $120^\circ + \theta$ ” be denoted as  $T$ . Suppose the probability of event  $T$  occurring is  $P(T)$ . Problem 1 asks for the conditional probability of event  $T$  occurring given that event  $H$  has occurred first, i.e., to find  $P(T|H)$ .

The following three methods can generate an angle of  $120^\circ + \theta$ :

The first method, as shown in Figure 20.6.9, involves rotating  $OA$  counterclockwise by an angle of  $\theta + \omega$  and rotating  $OD$  counterclockwise by an angle of  $\omega$ . If  $\omega = 0^\circ$ , then  $OD$  remains stationary, but  $OA$  is rotated counterclockwise by an angle of  $\theta$  to obtain the segment  $OA'$ , and  $OA'$  intersects  $OD$  at an angle of  $120^\circ + \theta$ . Simultaneously, if  $OB$  is rotated clockwise by an angle of  $\theta$  to obtain the segment  $OB'$ , then  $OB'$  intersects  $OC$  at an angle of  $120^\circ + \theta$ .

The second method involves rotating  $OD$  clockwise by an angle of  $\theta + \omega$  and rotating  $OA$  clockwise by an angle of  $\omega$ , causing  $OD$  and  $OA$  to intersect at an angle of  $120^\circ + \theta$ . At the same time, rotating  $OC$  counterclockwise by an angle of  $\theta$  causes  $OB$  and  $OC$  to also intersect at an angle of  $120^\circ + \theta$ .

The first method and the second method are similar.

The third method involves rotating  $OA$  counterclockwise and  $OD$  clockwise simultaneously by an angle of  $\theta/2$ , while also rotating  $OB$  clockwise and  $OC$  counterclockwise by an angle of  $\theta/2$ , which also results in two pairs of segments intersecting at an angle of  $120^\circ + \theta$ .

Therefore, we can adjust the angles of these four line segments within  $\angle AOB$  or within  $\angle COD$  to form two  $120^\circ + \theta$  angles. This experiment of adjusting the segment angles is the random experiment  $E$ . Since  $\angle AOB + \angle COD = 120^\circ$ , the total number of sample points in the sample space of random experiment  $E$  is  $120^\circ$ .

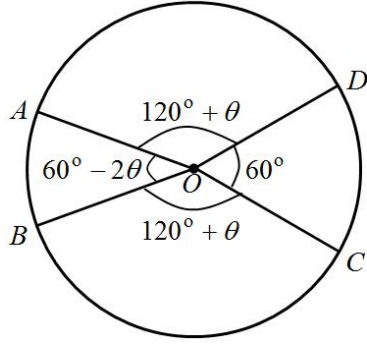
In the first method, when  $\omega = 0^\circ$ ,  $OA'$  and  $OD$  form an angle of  $120^\circ + \theta$ . However, as  $\omega$  varies within this range,  $0^\circ \leq \omega \leq 60^\circ - 2\theta$ ,  $OA'$  and  $OD$  still maintain an angle of  $120^\circ + \theta$ . Therefore, the number of sample points for event  $T$  is  $60^\circ - 2\theta$ , and thus the conditional probability of event  $T$  occurring is

$$P(T|H) = \frac{60^\circ - 2\theta}{120^\circ}. \quad (20.6.16)$$

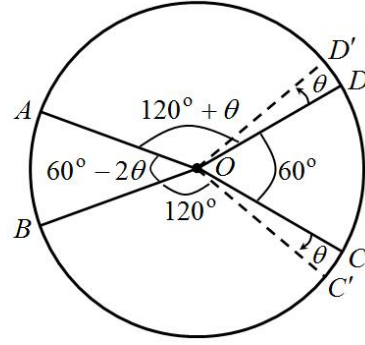
**Problem 2** Figure 20.6.10 represents the smooth manifold of the neutron. At any point  $O$  on the neutron's smooth manifold, four line segments  $OA$ ,  $OB$ ,  $OC$ , and  $OD$  are connected, with  $OA$  and  $OD$  intersecting at an angle of  $120^\circ + \theta$ , and  $OB$  and  $OC$  also intersecting at an angle of  $120^\circ + \theta$ . The question is: What is the probability that at this point  $O$ , the pairs  $OA$  and  $OD$ , and  $OB$  and  $OC$ , transform into intersecting angles of  $120^\circ$ ? The magnitude of angle  $\theta$  is given by equation (20.6.15).

**Solution** Problem 2 is the inverse of Problem 1. Let the event “ $OA$  and  $OD$ ,  $OB$  and  $OC$  intersect at angles of  $120^\circ + \theta$ ” be denoted as  $T$ . Suppose the probability of event  $T$  occurring is  $P(T)$ . Let the event “ $OA$  and  $OD$ ,  $OB$  and  $OC$  transform into intersecting angles of  $120^\circ$ ” be denoted as  $H$ . Suppose the probability of event  $H$  occurring is  $P(H)$ . Problem 2 asks for the conditional probability of event  $H$  occurring given that event  $T$  has occurred first, i.e., to find  $P(H|T)$ .





**Figure 20.6.10** Angles intersecting at  $120^\circ + \theta$



**Figure 20.6.11** Transformed into angles of  $120^\circ$

The following three methods can generate an angle of  $120^\circ$ :

The first method, as shown in Figure 20.6.11, involves rotating  $OA$  clockwise by an angle of  $\omega$  and rotating  $OD$  counterclockwise by an angle of  $\theta - \omega$ , so that

$$120^\circ + \theta - \omega - (\theta - \omega) = 120^\circ.$$

If  $\omega = 0^\circ$ , then  $OA$  remains stationary, while  $OD$  is rotated counterclockwise by an angle of  $\theta$  to obtain segment  $OD'$ . Then  $OA$  and  $OD'$  intersect at an angle of  $120^\circ$ . Simultaneously, if  $OC$  is rotated clockwise by an angle of  $\theta$  to obtain segment  $OC'$ , then  $OB$  and  $OC'$  intersect at an angle of  $120^\circ$ .

In Figure 20.6.11, if the clockwise direction is defined as positive, then when  $\omega \geq 0$ , segment  $OD$  moves within  $\angle C'OD'$ , and the range of  $\omega$  is  $0^\circ \leq \omega \leq 60^\circ + 2\theta$ . When  $\omega < 0$ , segment  $OA$  moves within  $\angle AOB$ .

The second method involves rotating  $OA$  clockwise and rotating  $OB$  counterclockwise by an angle of  $\theta$  each, which also yields two pairs of segments intersecting at an angle of  $120^\circ$ .

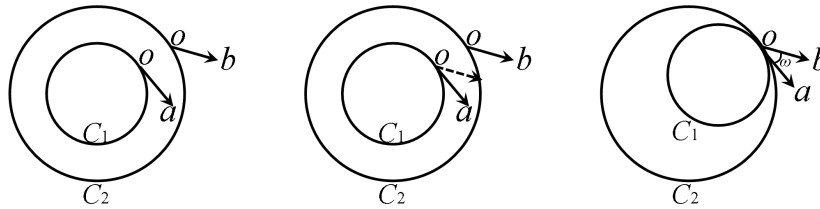
The third method involves rotating  $OD$  counterclockwise and rotating  $OA$  clockwise by an angle of  $\theta/2$ , while simultaneously rotating  $OC$  clockwise and rotating  $OB$  counterclockwise by an angle of  $\theta/2$ , which also yields two pairs of segments intersecting at an angle of  $120^\circ$ .

These experiments of adjusting the segment angles constitute the random experiment  $E$ . These three methods indicate that the total number of sample points in the sample space of random experiment  $E$  is  $120^\circ$ .

Now,  $\angle AOD'$  and  $\angle BOC'$  are already angles of  $120^\circ$ . However, in the first method, the maximum possible range for  $\omega$  is  $0^\circ \leq \omega \leq 60^\circ + 2\theta$ . Therefore, the number of sample points for event  $H$  is  $60^\circ + 2\theta$ , and thus the conditional probability of event  $H$  occurring is

$$P(H|T) = \frac{60^\circ + 2\theta}{120^\circ}. \quad (20.6.17)$$

Both problems above have multiple solution methods, but we should adopt the first method for the following reasons:



**Figure 20.6.12** Method for finding the angle between vectors on two circles

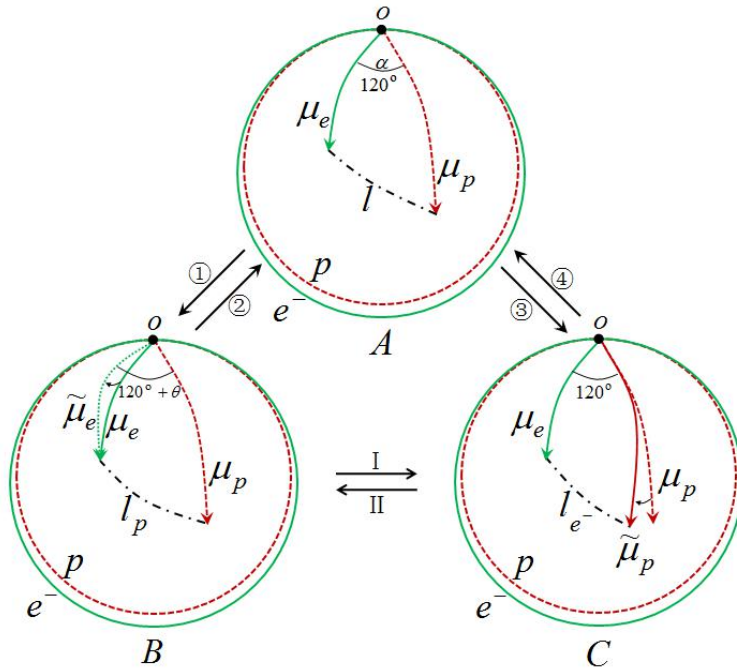
As shown in the left diagram of Figure 20.6.12, to find the angle between vector  $a$  on circle

$C_1$  and vector  $b$  on circle  $C_2$ , there are two approaches: First, as shown in the middle diagram, translate vector  $b$  from circle  $C_2$  to point  $o$  on circle  $C_1$ ; second, as shown in the right diagram, parallel-translate circle  $C_1$  onto circle  $C_2$  so that the points  $o$  on both circles coincide. Both methods yield the angle between the two vectors as  $\omega$ . This angle  $\omega$  is the angle between vectors  $a$  and  $b$  in the plane.

Similarly, as shown in diagram  $A$  of Figure 20.6.13, the solid circle represents the electron smooth manifold  $e^-$ , and the dashed circle represents the proton smooth manifold  $p$ . The spin magnetic moment  $\mu_e$  lies on the electron smooth manifold  $e^-$ , and the spin magnetic moment  $\mu_p$  lies on the proton smooth manifold  $p$ .  $\mu_e$  and  $\mu_p$  reside on different smooth manifolds, and their angle is the angle  $\alpha$  between  $\mu_e$  and  $\mu_p$  in a flat space such as  $M^4$ . Assume initially  $\alpha = 120^\circ$ .

① In the flat space  $M^4$ , the angle between  $\mu_e$  and  $\mu_p$  is  $120^\circ$ . As shown in diagram  $B$  of Figure 20.6.13, if  $\mu_e$  is moved or projected onto the proton smooth manifold  $p$  to become  $\tilde{\mu}_e$ , and if  $\mu_e$  and  $\tilde{\mu}_e$  have equal length, denoted  $|\mu_e| = |\tilde{\mu}_e|$ , and  $l = l_p$ , where  $l = d(\mu_e, \mu_p)$ , and  $l_p = d(\tilde{\mu}_e, \mu_p)$ , then because the sectional curvature of the proton smooth manifold  $p$  is greater than zero, according to Theorem 11.12.1 and the Toponogov comparison theorem, the angle  $\alpha_p$  between  $\tilde{\mu}_e$  and  $\mu_p$  on the proton manifold is greater than the angle  $\alpha = 120^\circ$  in flat  $M^4$ , i.e.,  $\alpha_p > \alpha$ . By suitably adjusting the position of  $\tilde{\mu}_e$ , we can achieve  $\alpha_p = 120^\circ + \theta$ .

② Conversely, when the angle between  $\tilde{\mu}_e$  and  $\mu_p$  on the proton manifold is  $\alpha_p = 120^\circ + \theta$ , according to Theorem 11.12.1, one can find a comparison triangle in  $M^4$  (i.e., in diagram  $A$ ) such that their corresponding sides are equal:  $|\mu_e| = |\tilde{\mu}_e|$ ,  $|\mu_p| = |\mu_p|$ ,  $l = l_p$ , and such that the angle between  $\mu_e$  and  $\mu_p$  in flat  $M^4$  is  $\alpha = 120^\circ$ .



**Figure 20.6.13** Reason for choosing the first solution method

③ Returning to diagram  $A$ , the angle between  $\mu_e$  and  $\mu_p$  in flat  $M^4$  is  $\alpha = 120^\circ$ . As shown in diagram  $C$  of Figure 20.6.13, suppose  $\mu_p$  is moved or projected onto the electron

smooth manifold  $e^-$  to become  $\tilde{\mu}_p$ , and  $\mu_p$  and  $\tilde{\mu}_p$  have equal length, denoted  $|\mu_p| = |\tilde{\mu}_p|$ . According to Theorem 11.12.1, in the electron smooth manifold  $e^-$  of diagram  $C$ , one can find a comparison hinge such that  $|\mu_e| = |\mu_e|$ ,  $|\mu_p| = |\tilde{\mu}_p|$ ,  $l > l_{e^-}$ ,  $l_{e^-} = d(\mu_e, \tilde{\mu}_p)$ . By the Toponogov comparison theorem, suitably adjusting the position of  $\tilde{\mu}_p$ —that is, suitably adjusting the length  $l_{e^-}$ —allows the angle  $\alpha_{e^-}$  between  $\mu_e$  and  $\tilde{\mu}_p$  on the electron manifold to equal the angle  $\alpha$  in flat  $M^4$ , i.e.,  $\alpha_{e^-} = \alpha = 120^\circ$ .

④ Conversely, when the angle between  $\mu_e$  and  $\tilde{\mu}_p$  on the electron manifold is  $\alpha_{e^-} = 120^\circ$ , according to Theorem 11.12.1, one can find a comparison hinge in  $M^4$  (i.e., in diagram  $A$ ) such that their corresponding sides are equal:  $|\mu_e| = |\mu_e|$ ,  $|\mu_p| = |\tilde{\mu}_p|$ , and such that the angle between  $\mu_e$  and  $\mu_p$  in flat  $M^4$  is  $\alpha = 120^\circ$ .

We can regard segment  $OA$  in Figure 20.6.9 as the electron spin magnetic moment  $\mu_e$  and segment  $OD$  as the proton spin magnetic moment  $\mu_p$ . Rotating  $OA$  counterclockwise by  $\theta$  corresponds to moving or projecting  $\mu_e$  onto the proton smooth manifold to become  $\tilde{\mu}_e$  in diagram  $B$  of Figure 20.6.13. The change in the angle between the spin magnetic moments  $\mu_e$  and  $\mu_p$  from  $120^\circ$  to  $120^\circ + \theta$  corresponds to transitioning from the state depicted in diagram  $A$  of Figure 20.6.13 ( $\alpha = 120^\circ$ ) to the state depicted in diagram  $B$  ( $\alpha = 120^\circ + \theta$ ).

Similarly, we can regard segment  $OA$  in Figure 20.6.11 as the electron spin magnetic moment  $\mu_e$  and segment  $OD$  as the proton spin magnetic moment  $\mu_p$ . Rotating  $OD$  counterclockwise by  $\theta$  corresponds to moving or projecting  $\mu_p$  onto the electron smooth manifold to become  $\tilde{\mu}_p$  in diagram  $C$  of Figure 20.6.13. The change in the angle between the spin magnetic moments  $\mu_e$  and  $\mu_p$  from  $120^\circ + \theta$  to  $120^\circ$  corresponds to transitioning from the state in diagram  $B$  to the state in diagram  $A$ , and then to the state in diagram  $C$ :  $\textcircled{2} \rightarrow \textcircled{3} = \text{I}$ . Thus, these two transformation processes ( $\textcircled{2} \rightarrow \textcircled{3} = \text{I}$  and  $\textcircled{4} \rightarrow \textcircled{1} = \text{II}$ ) align with the first solution method for Problems 1 and 2; therefore, we should adopt the first solution method for both problems.

Since the neutron is a smooth manifold composed of infinitely many points, and each point can potentially serve as the identity element, at each point either the hinge state or the triangle state can occur. That is, at each point, events  $H$  and  $T$  can occur, and the probabilities of these two states appearing at different points are equal. Therefore, events  $H$  and  $T$  are independent, so  $P(HT) = P(H)P(T)$ . From the conditional probability formula

$$P(T | H) = \frac{P(TH)}{P(H)} = \frac{P(T)P(H)}{P(H)} = P(T),$$

and using equation (20.6.16), we obtain

$$P(T) = P(T | H) = \frac{60^\circ - 2\theta}{120^\circ}. \quad (20.6.18)$$

Similarly, from the conditional probability formula

$$P(H | T) = \frac{P(TH)}{P(T)} = \frac{P(T)P(H)}{P(T)} = P(H),$$

and using equation (20.6.17), we obtain

$$P(H) = P(H | T) = \frac{60^\circ + 2\theta}{120^\circ}. \quad (20.6.19)$$

Now, using the probabilities  $P(T)$  and  $P(H)$ , we compute the average neutron spin magnetic moment

$$\bar{\mu}_{n,\text{cal}} = \mu'_{n,\text{cal}} \times P(H) + \mu''_{n,\text{cal}} \times P(T),$$

that is

$$\bar{\mu}_{n,\text{cal}} = \mu'_{n,\text{cal}} \times \frac{60^\circ + 2\theta}{120^\circ} + \mu''_{n,\text{cal}} \times \frac{60^\circ - 2\theta}{120^\circ}. \quad (20.6.20)$$

Substituting the numerical values from equations (20.6.8), (20.6.15), and (20.6.11) into equation (20.6.20), we calculate

$$\bar{\mu}_{n,\text{cal}} = 1.91046591906317 \frac{e\hbar}{2m_n}.$$

Multiplying  $\bar{\mu}_{n,\text{cal}}$  by  $\frac{m_n}{m_p}$  gives the average spin magnetic moment in units of the nuclear magneton  $\frac{e\hbar}{2m_p}$ ,

$$\begin{aligned} \bar{\mu}_{n,\text{cal}} &= 1.91046591906317 \frac{e\hbar}{2m_n} \times \frac{m_n}{m_p} \\ &= 1.91046591906317 \times 1.00137840580631 \frac{e\hbar}{2m_p} = 1.91309931637875 \frac{e\hbar}{2m_p}. \end{aligned}$$

The difference between the calculated and experimental values is

$$\bar{\mu}_{n,\text{cal}} - \mu_{n,\text{exp}} = 1.91309931637875 - 1.91304275 = 0.0000565663787543969,$$

and the relative error is

$$\left( \frac{\bar{\mu}_{n,\text{cal}}}{\mu_{n,\text{exp}}} - 1 \right) \times 100\% = 0.002956880\%.$$

If we substitute the experimental values of the electron spin magnetic moment

$$M_{S,\text{exp}} = (1 + 0.0011659202) \frac{e}{m_e}$$

and the proton spin magnetic moment

$$\mu_{p,\text{exp}} = 2.79284734462 \mu_N$$

into the above formulas for calculation—that is, change the values of  $\mu_e$  and  $\mu_p$  in equation (20.6.7) to

$$\mu_e = 1 + 0.0011659202, \quad \mu_p = 2.7928473446 \times \frac{1}{2},$$

then the calculated neutron spin magnetic moment differs too much from the experimental value. A possible reason is that during the measurement of the electron and proton spin magnetic moments, photons exist in the surrounding environment. Because photons constitute a rotation group, they may influence the spin magnetic moments of the measured electron and proton, causing the experimentally obtained values for the electron and proton spin magnetic moments not to be their true magnetic moments. In other words, the measured values have significant errors compared to the true magnetic moments of the electron and proton.

## §20.7 Weak and Strong Interactions

In the previous calculation of the neutron's spin magnetic moment, we considered the neutron to be composed of a proton  $p$ , an electron  $e^-$ , and an antineutrino  $\bar{\nu}_e$ . The proton  $p$  and the antineutrino  $\bar{\nu}_e$  together form a projective space. One may naturally ask: since the antineutrino  $\bar{\nu}_e$  carries no electric charge, can it really bind with the proton  $p$ ? What force holds them together? This section addresses these questions and explains the mechanisms underlying weak and strong interactions.

### 1. Attractive Forces Near point-like Neutrinos, Antineutrinos, and Photons

The spin operator that determines the structure group of the principal bundle for neutrinos

and antineutrinos is

$$\hat{S}_y = \frac{i}{2} \hbar \left( v \frac{\partial}{\partial u} - u \frac{\partial}{\partial v} \right). \quad (20.7.1)$$

We have already solved the integral–curve equation of this operator as  $u^2 + v^2 = C$ ; see equation (20.1.15). The operator  $\hat{S}_y$  is a complex vector. Multiplying both sides of (20.7.1) by the imaginary unit  $i$  yields a real vector

$$\hat{S}'_y = i\hat{S}_y = \frac{1}{2} \hbar \left( -v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \quad (20.7.2)$$

From operator (20.7.2) one can also obtain the integral–curve equations  $u^2 + v^2 = C$ . Therefore, to avoid complications introduced by the imaginary unit  $i$ , when studying the internal forces of neutrinos and antineutrinos we replace the operator  $\hat{S}_y$  with  $\hat{S}'_y$ .

From operator (20.7.2) we derive the system of equations:

$$\frac{du}{dt} = -\frac{1}{2} \hbar v, \quad (20.7.3)$$

$$\frac{dv}{dt} = \frac{1}{2} \hbar u. \quad (20.7.4)$$

Differentiating equation (20.7.3) gives

$$\frac{d^2u}{dt^2} = -\frac{1}{2} \hbar \frac{dv}{dt} = -\frac{1}{2} \hbar \left( \frac{1}{2} \hbar u \right) = -\frac{1}{4} \hbar^2 u. \quad (20.7.5)$$

Comparing equation (20.7.5) with Newton's second law, we can set

$$F(u) = -\frac{1}{4} \hbar^2 u. \quad (20.7.6)$$

We call  $F(u)$  the force that exists in the vicinity of a point-like neutrino or antineutrino. Here  $u$  is taken as the coordinate of the point-like neutrino/antineutrino,  $\frac{du}{dt} = -\frac{1}{2} \hbar v$  as its velocity, and

$$T = \frac{1}{2} \left( \frac{du}{dt} \right)^2 = \frac{1}{2} \left( -\frac{1}{2} \hbar v \right)^2 = \frac{1}{8} \hbar^2 v^2 \quad (20.7.7)$$

as its kinetic energy.

Assuming the interior of a neutrino/antineutrino is a conservative system with a potential-energy function  $U(u)$ , we have

$$F(u) = -\frac{\partial U}{\partial u}. \quad (20.7.8)$$

Integrating gives

$$U = -\int_{u_0}^u F(\lambda) d\lambda. \quad (20.7.9)$$

Substituting (20.7.6) into (20.7.9) yields

$$U = \int_{u_0}^u \frac{1}{4} \hbar^2 \lambda d\lambda = \frac{1}{4} \hbar^2 \left[ \frac{1}{2} \lambda^2 \right]_{u_0}^u = \frac{1}{8} \hbar^2 u^2 - \frac{1}{8} \hbar^2 u_0^2 = \frac{1}{8} \hbar^2 u^2. \quad (20.7.10)$$

In equation (20.7.10) we set the potential energy to zero at the point  $u_0$ , i.e., take  $\frac{1}{8} \hbar^2 u_0^2 = 0$ .

The total mechanical energy of this conservative system is

$$E = T + U = \frac{1}{8} \hbar^2 v^2 + \frac{1}{8} \hbar^2 u^2.$$

Since

$$\frac{dE}{dt} = \frac{d}{dt} \left( \frac{1}{8} \hbar^2 v^2 + \frac{1}{8} \hbar^2 u^2 \right) = \frac{1}{8} \hbar^2 \left( 2v \frac{dv}{dt} + 2u \frac{du}{dt} \right) = \frac{1}{4} \hbar^2 \left( v \left( \frac{1}{2} \hbar u \right) + u \left( -\frac{1}{2} \hbar v \right) \right) = 0,$$

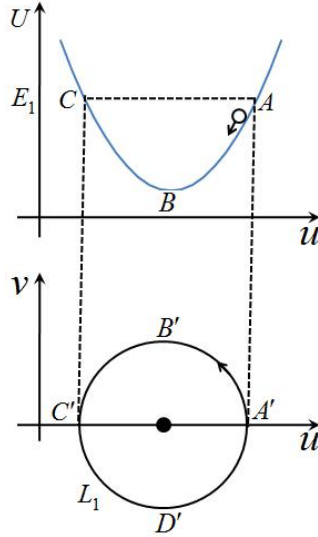
the total mechanical energy  $E$  is conserved.

From equations (20.7.3) and (20.7.4) we see that  $(u, v) = (0, 0)$  is a singularity of the smooth

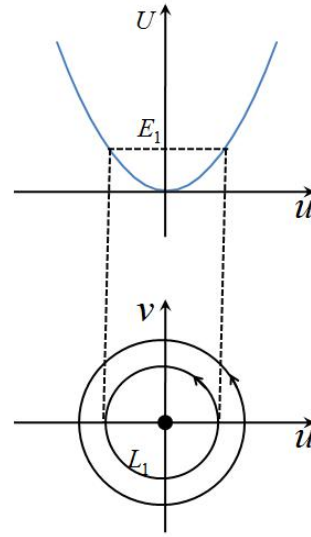
tangent-vector field  $\hat{S}'_y$ . Equation (20.7.6) shows that at this singularity  $F(u)=0$ , and from (20.7.8) we obtain

$$F(0) = -\left.\frac{\partial U}{\partial u}\right|_0 = 0,$$

so the potential energy attains a critical (minimum) value at this singularity.



**Figure 20.7.1** A point-like particle in a potential well and the corresponding integral curve



**Figure 20.7.2** Attractive force near a singularity whose integral curves are circles

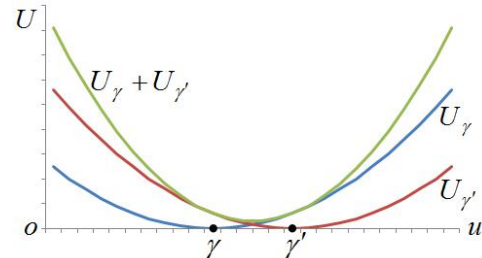
Equation (20.7.10) tells us that  $U \geq 0$ . This indicates that **near the singularity of the spin operator (smooth tangent-vector field)  $\hat{S}'_y$** —i.e., near the identity element or near a point-like neutrino/antineutrino—**an attractive force exists**.

As illustrated in Figure 20.7.1, assume the total mechanical energy is  $E$ . Because the potential energy  $U$  cannot exceed the total energy  $E$ , the bead reaches its highest point at  $A$ . The bead falling from  $A$  to the lowest point  $B$  corresponds to the point-like particle moving from point  $A'$  to point  $B'$ . The bead moving from  $B$  to the highest point  $C$  corresponds to the point-like particle moving from point  $B'$  to point  $C'$ .

Since the integral curves of the three spin operators of the photon are also circles, a similar analysis shows that **an attractive force likewise exists around a point-like photon**. In summary, **near any singularity of a spin operator whose integral curves are circles, an attractive force is present**, as shown in Figure 20.7.2.

Experiments reveal that **Bose–Einstein condensation of light** occurs near absolute zero. The reason such condensation can happen is precisely because **an attractive force exists around point-like photons**. The potential energy of a system formed by two or more aggregated point-like photons is also positive, as shown in Figure 20.7.3 (we set  $\hbar = 1$  and plot the potential-energy function  $U = \frac{1}{8}u^2$ ; the

same plotting method is used for subsequent figures in this section). Consequently, aggregated point-like photons exert an attractive force on other nearby point-like photons, drawing more of them together and thereby producing Bose–Einstein condensation. **Any particle whose spin operator has integral curves identical to those of the photon (i.e., circles) will, like the photon, exhibit condensation phenomena.**



**Figure 20.7.3** Superposition of the potential-energy curves for two photons

In practical calculations, the potential-energy function (20.7.10) should be multiplied by a proportionality constant; for example, let the potential-energy function be  $U = \frac{1}{2}ku^2$ , where

$k = \omega^2$  is the frequency of light. Assuming this is the potential-energy function near a point-like photon, and because the photon is a smooth manifold, the point-like photon can appear at any point of this smooth manifold, we ought to employ a quantum-mechanical approach and **operator-ize** (i.e., tensor-field-ize) the physical quantities. Operator-izing the kinetic-energy expression (20.7.7) for the photon gives:

$$T = \frac{1}{2} \left( \frac{du}{dt} \right)^2 \rightarrow -\frac{\hbar^2}{2} \frac{d^2}{du^2},$$

leading to the Schrödinger equation satisfied by the photon:

$$-\frac{\hbar^2}{2} \frac{d^2\psi}{du^2} + \frac{1}{2}ku^2\psi = E\psi.$$

This is the familiar **linear harmonic-oscillator equation**. Solving it yields

$$E = \left( n + \frac{1}{2} \right) \hbar\omega, \quad n = 0, 1, 2, \dots$$

Therefore, **the energy of a point-like photon is quantized**—it comes in discrete quanta—precisely because its potential-energy function is  $U = \frac{1}{2}ku^2$ . That the energy of point-like photons is quantized has been fully confirmed by experiment, which also validates the conclusions we have drawn from the above analysis.

## 2.Repulsive Forces Near Point-like Protons and Antiprotons

The spin operator that determines the structure group of the principal bundle for protons and antiprotons is

$$\hat{S}_x = \frac{1}{2} \hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \quad (20.7.11)$$

From operator (20.7.11) we obtain the system of equations:

$$\frac{du}{dt} = \frac{1}{2} \hbar v, \quad (20.7.12)$$

$$\frac{dv}{dt} = \frac{1}{2} \hbar u. \quad (20.7.13)$$

Differentiating equation (20.7.12) gives

$$\frac{d^2u}{dt^2} = \frac{1}{2} \hbar \frac{dv}{dt} = \frac{1}{2} \hbar \left( \frac{1}{2} \hbar u \right) = \frac{1}{4} \hbar^2 u. \quad (20.7.14)$$

Comparing equation (20.7.14) with Newton's second law, we can set

$$F(u) = \frac{1}{4} \hbar^2 u. \quad (20.7.15)$$

We call  $F(u)$  the force that exists in the vicinity of a point-like proton or antiproton. Here  $u$  is taken as the coordinate of the point-like proton/antiproton,  $\frac{du}{dt} = \frac{1}{2} \hbar v$  as its velocity, and

$$T = \frac{1}{2} \left( \frac{du}{dt} \right)^2 = \frac{1}{2} \left( \frac{1}{2} \hbar v \right)^2 = \frac{1}{8} \hbar^2 v^2 \quad (20.7.16)$$

as its kinetic energy.

Assuming the interior of a proton/antiproton is a conservative system with a potential-energy function  $U(u)$ , we have

$$F(u) = -\frac{\partial U}{\partial u}. \quad (20.7.17)$$

Integrating gives



$$U = -\int_{u_0}^u F(\lambda) d\lambda. \quad (20.7.18)$$

Substituting (20.7.15) into (20.7.18) yields

$$U = -\int_{u_0}^u \frac{1}{4} \hbar^2 \lambda d\lambda = -\frac{1}{4} \hbar^2 \left[ \frac{1}{2} \lambda^2 \right]_{u_0}^u = -\frac{1}{8} \hbar^2 u^2 + \frac{1}{8} \hbar^2 u_0^2 = -\frac{1}{8} \hbar^2 u^2. \quad (20.7.19)$$

In equation (20.7.19) we set the potential energy to zero at the point  $u_0$ , i.e., take  $\frac{1}{8} \hbar^2 u_0^2 = 0$ .

The total mechanical energy of this conservative system is

$$E = T + U = \frac{1}{8} \hbar^2 v^2 - \frac{1}{8} \hbar^2 u^2.$$

Since

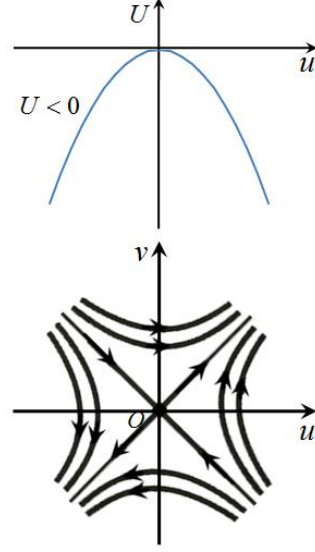
$$\begin{aligned} \frac{dE}{dt} &= \frac{d}{dt} \left( \frac{1}{8} \hbar^2 v^2 - \frac{1}{8} \hbar^2 u^2 \right) \\ &= \frac{1}{8} \hbar^2 \left( 2v \frac{dv}{dt} - 2u \frac{du}{dt} \right) \\ &= \frac{1}{4} \hbar^2 \left( v \left( \frac{1}{2} \hbar u \right) - u \left( \frac{1}{2} \hbar v \right) \right) = 0, \end{aligned}$$

the total mechanical energy  $E$  is conserved.

From equations (20.7.12) and (20.7.13) we see that  $(u, v) = (0, 0)$  is a singularity of the smooth tangent-vector field  $\hat{S}_x$ . Equation (20.7.15) shows that at this singularity  $F(u) = 0$ , and from (20.7.17) we obtain

$$F(0) = -\left. \frac{\partial U}{\partial u} \right|_0 = 0,$$

so the potential energy attains a critical (maximum) value at this singularity.



**Figure 20.7.4** A repulsive force exists near the singularity of the proton spin operator  $\hat{S}_x$

Equation (20.7.19) tells us that  $U \leq 0$ . This indicates that **near the singularity of the spin operator (smooth tangent-vector field)  $\hat{S}_x$** —i.e., near the identity element or near a point-like proton/antiproton—a **repulsive force exists**, as illustrated in Figure 20.7.4.

### 3.Repulsive Forces Near Point-like Electrons and Positrons

The spin operator that determines the structure group of the principal bundle for electrons and positrons is

$$\hat{S}_z = \frac{1}{2} \hbar \left( u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v} \right). \quad (20.7.20)$$

From operator (20.7.20) we obtain the system of equations:

$$\frac{du}{dt} = \frac{1}{2} \hbar u, \quad (20.7.21)$$

$$\frac{dv}{dt} = -\frac{1}{2} \hbar v. \quad (20.7.22)$$

Differentiating equation (20.7.21) gives

$$\frac{d^2 u}{dt^2} = \frac{1}{2} \hbar \frac{du}{dt} = \frac{1}{2} \hbar \left( \frac{1}{2} \hbar u \right) = \frac{1}{4} \hbar^2 u. \quad (20.7.23)$$

Comparing equation (20.7.23) with Newton's second law, we can set

$$F(u) = \frac{1}{4} \hbar^2 u. \quad (20.7.24)$$



We call  $F(u)$  the force that exists in the vicinity of a point-like electron or positron. Here  $u$  is taken as the coordinate of the point-like electron/positron,  $\frac{du}{dt} = \frac{1}{2}\hbar u$  as its velocity, and

$$T = \frac{1}{2} \left( \frac{du}{dt} \right)^2 = \frac{1}{2} \left( \frac{1}{2}\hbar u \right)^2 = \frac{1}{8} \hbar^2 u^2 \quad (20.7.25)$$

as its kinetic energy.

Assuming the interior of an electron/positron is a conservative system with a potential-energy function  $U(u)$ , we have

$$F(u) = -\frac{\partial U}{\partial u}. \quad (20.7.26)$$

Integrating gives

$$U = -\int_{u_0}^u F(\lambda) d\lambda. \quad (20.7.27)$$

Substituting (20.7.24) into (20.7.27) yields

$$U = -\int_{u_0}^u \frac{1}{4} \hbar^2 \lambda d\lambda = -\frac{1}{4} \hbar^2 \left[ \frac{1}{2} \lambda^2 \right]_{u_0}^u = -\frac{1}{8} \hbar^2 u^2 + \frac{1}{8} \hbar^2 u_0^2 = -\frac{1}{8} \hbar^2 u^2. \quad (20.7.28)$$

In equation (20.7.28) we set the potential energy to zero at the point  $u_0$ , i.e., take  $\frac{1}{8} \hbar^2 u_0^2 = 0$ .

The total mechanical energy of this conservative system is

$$E = T + U = \frac{1}{8} \hbar^2 u^2 - \frac{1}{8} \hbar^2 u^2 = 0,$$

which shows that for a point-like electron/positron the kinetic and potential energies are opposite in sign at any instant.

From equations (20.7.21) and (20.7.22) we see that  $(u, v) = (0, 0)$  is a singularity of the smooth tangent-vector field  $\hat{S}_z$ . Equation (20.7.24) shows that at this singularity  $F(u) = 0$ , and from (20.7.26) we obtain

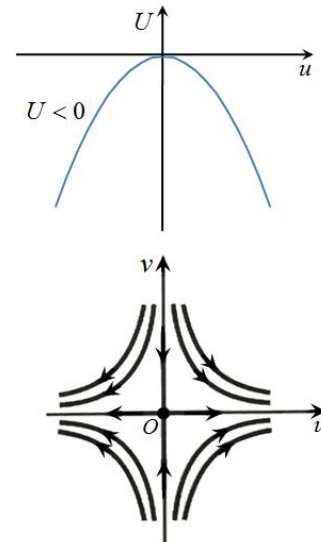
$$F(0) = -\left. \frac{\partial U}{\partial u} \right|_0 = 0,$$

so the potential energy attains a critical (maximum) value at this singularity.

Equation (20.7.28) tells us that  $U \leq 0$ . This indicates that **near the singularity of the spin operator (smooth tangent-vector field)  $\hat{S}_z$** —i.e., near the identity element or near a point-like electron/positron—**a repulsive force exists**, as illustrated in Figure 20.7.5.

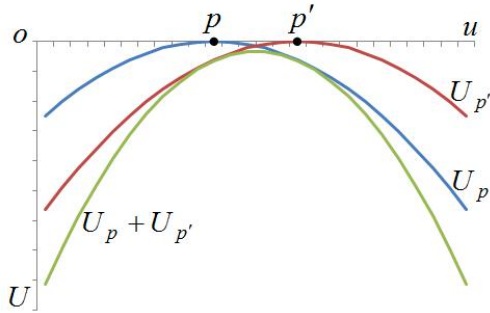
Because the potential energies near a point-like electron and a point-like proton are both negative (giving rise to repulsive forces), when two point-like electrons, two point-like protons, or a point-like electron and a point-like proton approach each other, the combined potential energy of the system is also negative (see Figure 20.7.6), and hence they repel one another. Therefore, two or more identical point-like electrons or point-like protons cannot completely merge; they cannot occupy the same state. In contrast, the potential near a point-like photon is positive (producing attraction), so two or more point-like photons can perfectly overlap, occupy the same state, and even undergo condensation.

If only an antineutrino  $\bar{\nu}_e$  and a proton  $p$  are present, the potential near the point-like antineutrino is positive, while that near the point-like proton is negative. When they come sufficiently close, the total potential-energy curve  $U_{\bar{\nu}_e} + U_p$  in their vicinity slopes upward to the right, as shown in Figure 20.7.7. Because a conservative force does work that reduces the total

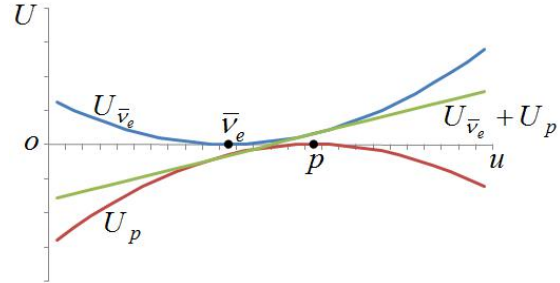


**Figure 20.7.5** A repulsive force exists near the singularity of the electron spin operator  $\hat{S}_z$

potential energy of the system, the point-like proton is drawn toward the point-like antineutrino. This means the point-like antineutrino exerts an **attractive force** on the point-like proton, allowing the two point-like particles to bind together. Similarly, a point-like antineutrino and a point-like electron can also bind.

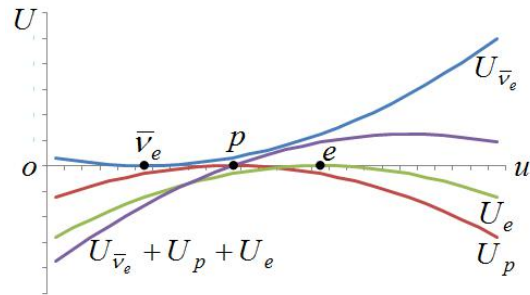


**Figure 20.7.6** A system of two point-like protons produces repulsion



**Figure 20.7.7** A system of a point-like antineutrino and a point-like proton produces attraction

If an antineutrino  $\bar{\nu}_e$ , a proton  $p$ , and an electron  $e^-$  combine to form a neutron, the neutron's potential-energy curve  $U_{\bar{\nu}_e} + U_p + U_e$  appears as in Figure 20.7.8. The point-like electron and point-like proton, already subject to Coulomb attraction, move closer together. Near the point-like antineutrino, the neutron's potential-energy curve slopes upward to the right, and because the conservative force in this system does work that lowers the total potential energy, the point-like electron and point-like proton are pulled toward the point-like antineutrino. Eventually, the three point-like particles come very close and bind tightly, forming a neutron.



**Figure 20.7.8** Formation of a neutron

From Figure 20.7.8 we can see that the potential-energy curve near a point-like neutron slopes upward to the right, indicating an **attractive force** near the point-like neutron. Consequently, many point-like neutrons can attract one another and aggregate, thereby forming a neutron star.

The attractive and repulsive forces discussed above all occur in the immediate vicinity of point-like particles; they are short-range forces. These are precisely the forces observed experimentally that give rise to **weak and strong interactions**.

## §20.8 Definition of Rest Mass and Its Applications

### 1.A New Definition of Rest Mass

In classical mechanics, the rest mass  $m_0$  of an object is defined as

$$m_0 = \rho V, \quad (20.8.1)$$

where  $\rho$  is the mass density and  $V$  is the volume of the object. Rest mass  $m_0$  is also defined as

$$m_0 = \frac{F}{a}. \quad (20.8.2)$$

where  $F$  is the force applied to the object and  $a$  is the acceleration produced.  $m_0$  is also referred to as **inertial mass**. This section presents an alternative definition of rest mass  $m_0$ .

Based on the content of Chapter 19, assume the gravitational field is generated by a quasi-stationary distribution of dust. The energy-momentum tensor for the quasi-stationary dust distribution is

$$T^{\mu\nu} = \rho u^\mu u^\nu, \quad (20.8.3)$$

where  $\rho$  is the mass density measured by a local inertial observer at rest instantaneously relative to the dust. If a coordinate system at rest relative to the dust is adopted, the trace of the energy-momentum tensor is

$$T = g_{\mu\nu} T^{\mu\nu} = g_{00} T^{00} = -\rho c^2. \quad (20.8.4)$$

Also, from equation (19.3.3),  $R = -\kappa T$ , we obtain

$$R = -\kappa T = \kappa \rho c^2. \quad (20.8.5)$$

Substituting equation (20.8.1) into (20.8.5) gives

$$R = \kappa c^2 \frac{m_0}{V}. \quad (20.8.6)$$

Equation (20.8.6) is similar to equation

$$R(p) = \frac{1}{V(B^m)} \int_{S^{m-1}} \langle X(v), v \rangle dV_{S^{m-1}}$$

in Theorem 11.6.4. Inspired by this, we define the rest mass  $m_0$  as

$$m_0 = \sigma \int_{S^2} \langle X(v), v \rangle dV_{S^2}, \quad (20.8.7)$$

where  $\sigma$  is a proportionality constant, and  $X(v)$  is a tangent vector field in the tangent space  $T_e M$  at the identity element  $e$  of the Lie group.  $m = 3$ , since the particle is three-dimensional. According to equation (11.6.9), equation (20.8.7) can also be written in the following vector form:

$$m_0 = \sigma \oint_{S^2} X \cdot dS. \quad (20.8.8)$$

Equation (20.8.8) indicates that the rest mass  $m_0$  of the particle is the flux of the tangent vector field  $X$  through the unit sphere  $S^2$  at the particle's identity element  $e$ .

Since

$$\sigma \oint_{S^2} X \cdot dS = m_0 = \int_V \rho dV,$$

equation (20.8.8) can be written in the following divergence form:

$$\nabla \cdot X = \frac{\rho}{\sigma}. \quad (20.8.9)$$

Therefore, it can be concluded that: if  $\nabla \cdot X \neq 0$ , then  $\rho \neq 0$ ; if  $\nabla \cdot X = 0$ , then  $\rho = 0$ .

## 2.Reasons for Different Rest Masses

Based on equation (20.8.8), we can understand why the rest mass of a proton is larger than that of an electron, and why the rest masses of photons and neutrinos are zero.

We know that every elementary particle with spin-1/2 is an  $SU(2)$  group, whereas the photon is an  $SO(3)$  group. The  $X$  in equation (20.8.8) is the spin operator that determines the structure group of the particle's principal bundle.

According to the previous analysis, the direction of the spin operator  $\hat{S}_x$ , which governs the main properties of the proton, is along the radial direction of the unit sphere  $S^2$  (see Figure. 20.1.2). Similarly, the direction of the spin operator  $\hat{S}_z$ , which governs the main properties of the electron, is also along the radial direction of the unit sphere  $S^2$  (see Figure. 20.1.7). Therefore, if we replace  $X$  in equation (20.8.8) with  $\hat{S}_x$  or  $\hat{S}_z$ , the integral in (20.8.8) becomes non-zero, so the rest masses of the proton and electron are not zero. The spin operator  $\hat{S}_y$ , which governs the main properties of the neutrino, is tangent to the unit sphere  $S^2$  (see Figure. 20.1.6). Thus, replacing  $X$  in equation (20.8.8) with  $\hat{S}_y$  results in an integral equal to zero, giving the neutrino a

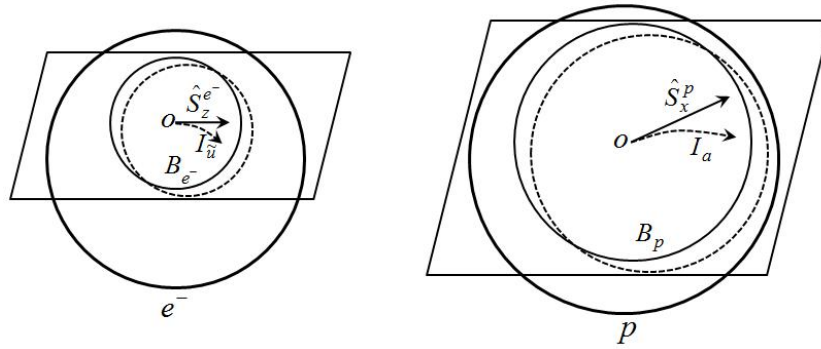
rest mass  $m_0 = 0$ . Similarly, for the photon, its spin operators  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$  are also tangent to the unit sphere  $S^2$  (see Figure. 20.2.1). Therefore, replacing  $\mathbf{X}$  in equation (20.8.8) with any one of the photon's spin operators  $\hat{L}_x$ ,  $\hat{L}_y$ , or  $\hat{L}_z$  also yields an integral equal to zero, so the photon's rest mass  $m_0 = 0$ .

Let us denote the three spin operators of the electron as  $\hat{S}_x^{e^-}$ ,  $\hat{S}_y^{e^-}$ , and  $\hat{S}_z^{e^-}$ . For the electron,  $\mathbf{X} = \hat{S}_z^{e^-}$ , and equation (20.8.8) should be revised as

$$m_{e^-} = \sigma \oint_{S_{e^-}^2} \hat{S}_z^{e^-} \cdot d\mathbf{S}. \quad (20.8.10)$$

Similarly, denote the three spin operators of the proton as  $\hat{S}_x^p$ ,  $\hat{S}_y^p$ , and  $\hat{S}_z^p$ . For the proton,  $\mathbf{X} = \hat{S}_x^p$ , and equation (20.8.8) should be revised as

$$m_p = \sigma \oint_{S_p^2} \hat{S}_x^p \cdot d\mathbf{S}. \quad (20.8.11)$$



**Figure 20.8.1** Exponential mapping for the electron and proton

According to Definition 11.1.6 or Theorem 12.5.7, the exponential map is a mapping from a neighborhood  $B \subset T_e G$  of the origin in the tangent space  $T_e G$  at the identity element  $e$  of a Lie group  $G$  to the smooth manifold  $G$ , i.e.,  $\exp_e : B \rightarrow G$ . As shown in the left panel of Figure 20.8.1, the neighborhood  $B_{e^-}$  contains the tangent vectors at the identity element  $e$  that lie in the tangent space at  $e$ . Therefore, the spin operator  $\hat{S}_z^{e^-}$  is in the neighborhood  $B_{e^-}$ , and the boundary of  $B_{e^-}$  is a two-dimensional unit sphere  $S_{e^-}^2$ .

### 3.Reasons Why the Proton-to-Electron Rest Mass Ratio is Approximately 1836

Experimentally, the proton's rest mass is measured to be  $1836.152701(37)^5$  times that of the electron. Let us analyze the reasons for this large ratio.

#### (1)Preliminary Analysis

Assuming that the proton's rest mass is an integer multiple, 1836 times the electron's rest mass, then from equations (20.8.10) and (20.8.11) we obtain

$$m_p = \sigma \oint_{S_p^2} \hat{S}_x^p \cdot d\mathbf{S} = 1836 \sigma \oint_{S_{e^-}^2} \hat{S}_z^{e^-} \cdot d\mathbf{S},$$

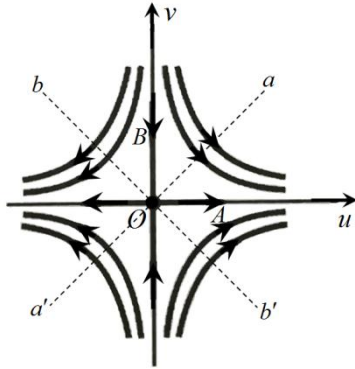
that is,

$$\oint_{S_p^2} \hat{S}_x^p \cdot d\mathbf{S} = 1836 \oint_{S_{e^-}^2} \hat{S}_z^{e^-} \cdot d\mathbf{S}. \quad (20.8.12)$$

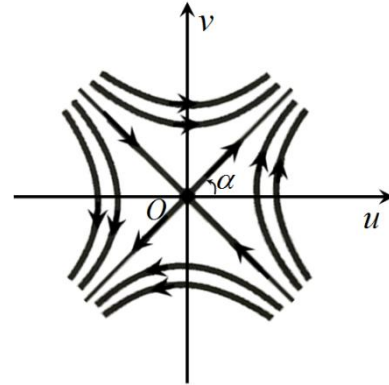
<sup>5</sup>Zeng Jinyan. *Quantum Mechanics (Volume I)* [M]. Beijing: Science Press, 2014: 544.

This implies that in the tangent space at the point-like proton  $e_p$ , the flux of  $\hat{S}_x^p$  enclosed by the two-dimensional integration surface  $S_p^2$  is 1836 times the flux of  $\hat{S}_z^e$  enclosed by the two-dimensional integration surface  $S_e^2$  in the tangent space at the point-like electron  $e_e$ .

For ease of reading, we redraw the integral curve diagrams of the spin operators for the electron and proton, as shown in Figures 20.8.2 and 20.8.3. Observing these two figures, we find that if Figure 20.8.2 is rotated counterclockwise by an angle of  $\pi/4$ , it precisely becomes Figure 20.8.3. However, rotating it by an angle of  $2n\pi + \pi/4$  (where  $n$  is an integer) would also make Figure 20.8.2 exactly become Figure 20.8.3.

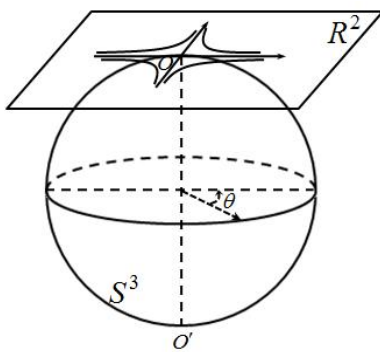


**Figure 20.8.2** Integral curves of the electron's spin operator

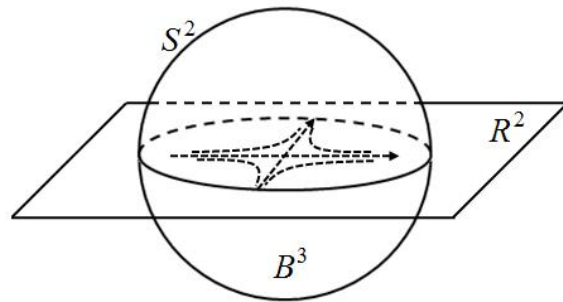


**Figure 20.8.3** Integral curves of the proton's spin operator

The phase plane  $R^2$  of the electron can be viewed as a plane in the three-dimensional tangent space at any point on the three-dimensional unit sphere  $S^3$  representing the electron, as shown in Figure 20.8.4. Particularly at the identity element, the phase plane  $R^2$  can be regarded as a cross-section of the three-dimensional unit solid open ball  $B^3$  centered at the zero vector within the three-dimensional tangent space, as illustrated in Figure 20.8.5. The boundary of this solid open ball  $B^3$  is precisely the integration surface  $S^2$  in equation (20.8.8). If this phase plane  $R^2$  rotates, it implies that this three-dimensional solid open ball  $B^3$  is also rotating.



**Figure 20.8.4** The phase plane  $R^2$  is a tangent plane to the sphere  $S^3$

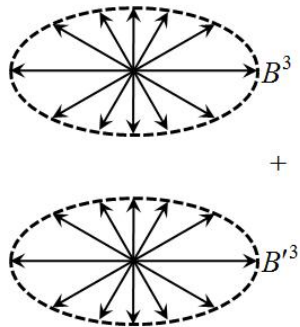


**Figure 20.8.5** The solid open ball  $B^3$ , its boundary  $S^2$ , and its cross-section  $R^2$

The rotation of the phase plane  $R^2$  drives the rotation of the electron's unit sphere  $S^3$ . This rotation can be viewed as a forward, orientation-preserving mapping  $f$  from the electron  $e^-$  to the proton  $p$ :

$$f: e^- \rightarrow p.$$

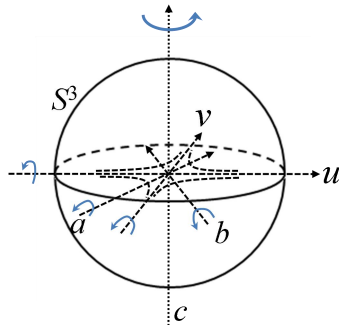
This mapping is similar to the one shown in Figure 9.7.3. It causes the electron's three-dimensional unit sphere  $S^3$  to rotate approximately 1836 times to become the proton  $p$ . The number 1836 is the degree of this mapping  $f$ . Based on the geometric meaning of the mapping degree, consider  $f$  as a covering map. The electron itself being stationary can be viewed as the electron completing its first rotation. At this moment, the proton begins to form, and its rest mass equals that of the electron. When the electron completes its second rotation, it covers itself once. This causes the solid open ball  $B^3$  within the tangent space at point  $o$  (the identity element) of the electron, as shown in Figure 20.8.1 and bounded by the two-dimensional sphere  $S^2$ , to also be covered once (two solid open balls coincide, equivalent to a single solid open ball rotating twice). Consequently, the tangent vectors passing through the boundary sphere  $S^2$  of this solid



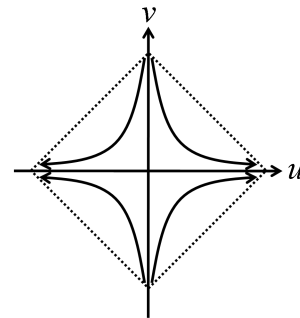
**Figure 20.8.6** Two solid open balls overlapping, doubling the flux.

open ball  $B^3$  double, as shown in Figure 20.8.6. From equation (20.8.10), the flux contained within the sphere  $S^2$  doubles, and thus the electron's rest mass would double. Therefore, we can infer: if the electron itself is stationary (equivalent to one rotation), the proton's rest mass equals the electron's rest mass; if the electron rotates twice, the proton's rest mass is twice the electron's rest mass; if it rotates 1836 times, the proton's rest mass is 1836 times the electron's rest mass, which is nearly the proton's rest mass, and the electron is nearly completely transformed into a proton. This is our preliminary estimate. But why does it require over 1836 rotations?

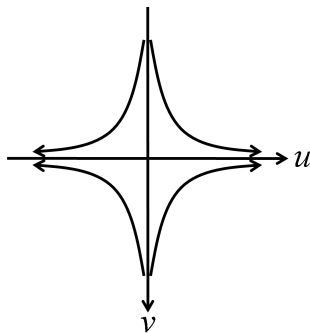
## (2) Synchronous (Coaxial) Rotation



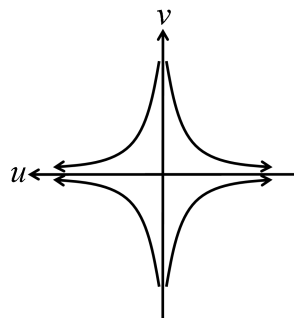
**Figure 20.8.7** Symmetry transformations



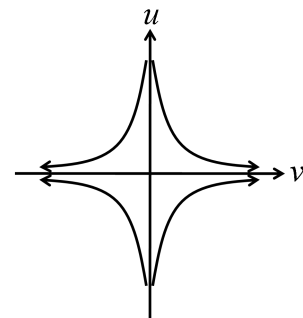
**Figure 20.8.8** The integral curves of the electron spin operator



**Figure 20.8.9** Rotation  $g_1$

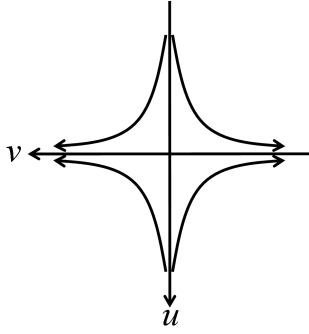


**Figure 20.8.10** Rotation  $g_2$

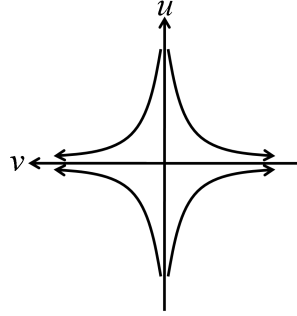


**Figure 20.8.11** Rotation  $g_3$

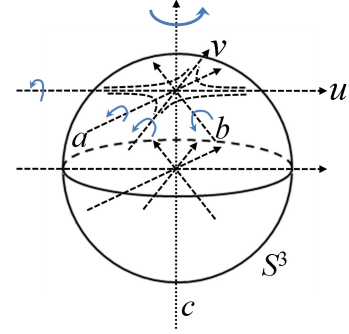




**Figure 20.8.12** Rotation  $g_4$



**Figure 20.8.13** Rotation  $g_5$



**Figure 20.8.14** Symmetry transformations

The integral curve diagrams of the electron spin operator in Figure 20.8.7 and Figure 20.8.8 are identical. Observing Figure 20.8.8, we find that the symmetry of the integral curve diagram of the electron spin operator resembles that of a square, and thus it possesses the same symmetry transformations as a square. If the integral curve diagram of the electron spin operator in Figure 20.8.7 is rotated counterclockwise by  $180^\circ$  about the  $u$ -axis, we obtain Figure 20.8.9, and denote this rotation as  $g_1$ . Similarly, if the integral curve diagram of the electron spin operator in Figure 20.8.7 is rotated counterclockwise by  $180^\circ$  about the  $v$ -axis, we obtain Figure 20.8.10, and denote this rotation as  $g_2$ . If the integral curve diagram of the electron spin operator in Figure 20.8.7 is rotated counterclockwise by  $180^\circ$  about the  $a$ -axis, we obtain Figure 20.8.11, and denote this rotation as  $g_3$ . If the integral curve diagram of the electron spin operator in Figure 20.8.7 is rotated counterclockwise by  $180^\circ$  about the  $b$ -axis, we obtain Figure 20.8.12, and denote this rotation as  $g_4$ . If the integral curve diagram of the electron spin operator in Figure 20.8.7 is rotated counterclockwise by  $90^\circ$  about the  $c$ -axis, we obtain Figure 20.8.13, and denote this rotation as  $g_5$ ; similarly, denote the counterclockwise rotation by  $180^\circ$  as  $g_6$ , the counterclockwise rotation by  $270^\circ$  as  $g_7$ , and the counterclockwise rotation by  $360^\circ$  as  $e$ .

Figures 20.8.9 to 20.8.13 have the same shape as Figure 20.8.8 (due to symmetric transformations), with the only difference being the orientation of the coordinate axes, which results from rotations. However, in the sphere  $S^3$ , owing to its spherical symmetry, the change in the orientation of the coordinate axes does not lead to any variation in the properties of the particle.

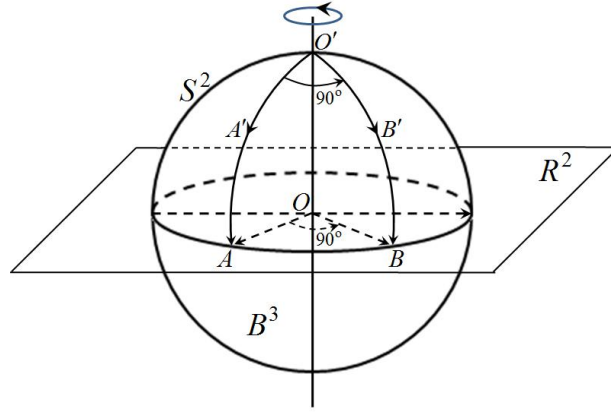
The symmetry axes in Figure 20.8.7 all intersect at the center of the sphere  $S^3$ , and the center remains stationary during the aforementioned rotations. If the phase plane  $R^2$  does not pass through the center of the sphere, as shown in Figure 20.8.14, then the rotation axis of the phase plane  $R^2$  is parallel to the rotation axis that passes through the center of the sphere.

The aforementioned rotations can form a point group  $D_4$ , denoted as

$$D_4 = \{e, g_1, g_2, g_3, g_4, g_5, g_6, g_7\}.$$

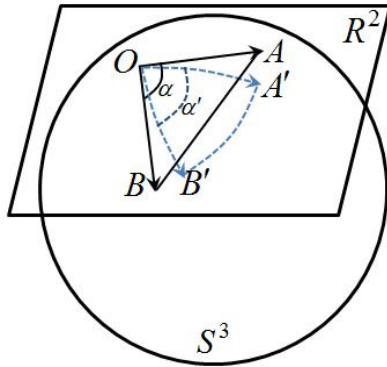
Here,  $e$  is the identity element, indicating a rotation of the sphere  $S^3$  by  $360^\circ$ , which is equivalent to the sphere  $S^3$  remaining stationary.

The mapping  $f$  contains the point group  $D_4$ . The mapping generated by the point group  $D_4$  can be denoted as  $f: D_4 \times e^- \rightarrow p$ .

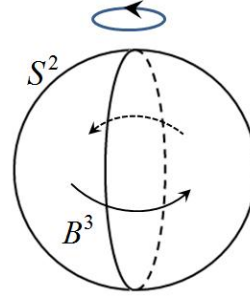


**Figure 20.8.15** Synchronous rotation

As shown in Figure 20.8.2,  $\angle AOB = \pi/2$ . Let the lengths of the tangent vectors  $\overrightarrow{OA}$  and  $\overrightarrow{OB}$  in the phase plane both be equal to 1, as illustrated in Figure 20.8.16. Furthermore, assume that under the exponential map, the tangent vectors  $\overrightarrow{OA}$  and  $\overrightarrow{OB}$  map to the geodesic segments  $OA'$  and  $OB'$  on  $S^3$ , respectively, with the arc lengths of  $OA'$  and  $OB'$  both equal to 1, and  $\angle A'OB' = \pi/2$ . According to the Toponogov comparison theorem, the length  $d(A', B')$  between points  $A'$  and  $B'$  is less than the length  $d(A, B)$  between points  $A$  and  $B$ . This is the hinge version of the Toponogov comparison theorem.



**Figure 20.8.16** Vector bundle



**Figure 20.8.17** Mutual covering of the left and right hemispheres of the open ball  $B^3$

During the rotation represented by each element of the point group  $D_4$ , the phase plane  $R^2$  rotates by an angle  $\theta$ , and the sphere  $S^3$  also rotates by an angle  $\theta$ . In the process of rotation,  $\angle AOB = \angle A'OB'$ , because if these two angles were not equal, the integral curve diagram of the electron spin operator would change, meaning that each element in the point group  $D_4$  would no longer be a symmetry transformation. We refer to these rotations as synchronous rotations. Since the point group  $D_4$  has a total of eight elements, there are eight distinct types of synchronous rotations, corresponding to the eight elements.

As shown in Figure 20.8.5, the phase plane  $R^2$  is the cross-section of the three-dimensional solid open ball  $B^3$  centered at the zero vector in the three-dimensional tangent space at the identity element of the electron's unit sphere  $S^3$ . Since both the phase plane  $R^2$  and the sphere  $S^3$  have rotated, the open ball  $B^3$  naturally rotates along with them, and it also rotates by an angle  $\theta$ .

A half-turn (rotation by  $180^\circ$ ) of the unit sphere  $S^3$  (or the open ball  $B^3$ ) is equivalent to covering the left hemisphere (left half-ball) onto the right hemisphere (right half-ball) once, and the right hemisphere (right half-ball) onto the left hemisphere (left half-ball) once, as illustrated in



Figure 20.8.17. Therefore, the mapping degree of this transformation is +1 (the rotation mapping preserves orientation). Similarly, rotations of the unit sphere  $S^3$  (or the open ball  $B^3$ ) by  $90^\circ$  or  $270^\circ$  also have a mapping degree of +1. A full rotation (by  $360^\circ$ ) of the unit sphere  $S^3$  (or the open ball  $B^3$ ) is equivalent to no motion at all, i.e., the identity mapping, and its mapping degree is also +1. Since the point group  $D_4$  has a total of eight rotation operations, the sum of the mapping degrees for all synchronous rotations is eight, i.e.,

$$\deg(f) = 8. \quad (20.8.13)$$

### (3) Asynchronous (Off-axis) Rotation

Next, we consider the triangle version of the Toponogov comparison theorem.

If, during the rotation process,  $\angle AOB \neq \angle A'O'B'$  in Figure 20.8.15 or  $\angle AOB \neq \angle A'OB'$  in Figure 20.8.16, then the phase plane  $R^2$  is not a symmetric transformation under the aforementioned rotation transformation. Consequently, this rotation is not one of the rotations represented by the point group  $D_4$ , i.e., it is not a synchronous rotation. Thus, the rotation axes of the sphere  $S^3$  and the phase plane  $R^2$  are different (neither coincident nor parallel). In the case of synchronous rotation (hinge version), a full rotation of the sphere  $S^3$  corresponds to a rotation by  $360^\circ$ . However, in the triangle version, a full rotation of the sphere  $S^3$  does not correspond to a rotation by  $360^\circ$ .

The rotation of the phase plane  $R^2$  implies that the tangent vectors on the phase plane  $R^2$  are also rotating. We can describe the rotation of any tangent vector on the phase plane  $R^2$  by describing the rotation of the unit tangent vector on the phase plane  $R^2$  (the electron's spin operator at the identity element can be regarded as a unit tangent vector). The rotation of the unit tangent vector on the phase plane  $R^2$  can be described by the unitary group

$$U(1) = \{z \mid |z| = 1; z \in \mathbb{C}\} = S^1.$$

Since we are studying the arbitrary rotational motion of vectors, if a vector's rotation process is completed in two steps—first rotating by  $30^\circ$ , then by  $50^\circ$ —it must be equivalent to completing the process through a single rotation by  $80^\circ$ . This requires that all vectors participating in the rotation form a group. The unit vectors emanating from the coordinate origin satisfy this requirement. Therefore, we set the tangent vectors  $\overrightarrow{OA}$  and  $\overrightarrow{OB}$  on the phase plane to have length 1 (i.e., unit tangent vectors) rather than any other length.

$S^3$  and  $R^2$  can be combined to form a vector bundle  $(P, S^3, R^2, \pi, U(1))$ , as shown in Figure 20.8.16, where the total space is  $P$ , the base manifold is  $S^3$ , the fiber type is  $R^2$ , and the structure group is  $U(1)$ . An element of the unitary group  $U(1)$  can be expressed as

$$U(1) = e^{in\theta}, \quad (20.8.14)$$

where  $i$  is the imaginary unit,  $n$  is an integer, and  $\theta$  is the rotation angle.

Let the smooth structure of the base manifold  $S^3$  be  $\{(U_\alpha, \varphi_\alpha); \alpha \in I\}$ , then the transition function family of the vector bundle is  $\{g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow U(1)\}$ , i.e.,

$$g_{\alpha\beta} = U(1) = e^{in\theta}.$$

Since  $g_{\alpha\beta}$  is the Jacobian matrix of the coordinate transformation

$$\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(U_\alpha \cap U_\beta) \rightarrow \varphi_\alpha(U_\alpha \cap U_\beta)$$

we can derive the functional expression for the coordinate transformation  $\varphi_\alpha \circ \varphi_\beta^{-1}$ . Since

$$\frac{d}{d\theta} \left( \frac{1}{in} e^{in\theta} + C \right) = e^{in\theta},$$

where  $C$  is a constant, the functional expression for the coordinate transformation  $\varphi_\alpha \circ \varphi_\beta^{-1}$  is

$$f(\theta) = -\frac{1}{n} i e^{in\theta} + C.$$

By choosing appropriate coordinates, we can make  $C = 0$ , and thus we obtain

$$f(\theta) = -\frac{1}{n}ie^{in\theta}. \quad (20.8.15)$$

Since

$$-\frac{1}{n}ie^{in\theta} = \frac{1}{n}(-ie^{in\theta}) = \frac{1}{n}\left[\left(\cos(-\frac{\pi}{2}) + i\sin(-\frac{\pi}{2})\right)e^{in\theta}\right] = \frac{1}{n}e^{i(n\theta - \frac{\pi}{2})},$$

we can also obtain

$$f(\theta) = \frac{1}{n}e^{i(n\theta - \frac{\pi}{2})}. \quad (20.8.16)$$

As shown in equations (20.8.14) and (20.8.16), when  $n = 1$ , if the phase plane  $R^2$  rotates from angle 0 to  $\pi/2$ , then the unit sphere  $S^3$  rotates from angle  $-\pi/2$  to angle 0. This indicates that when the phase plane  $R^2$  rotates, the electron sphere  $S^3$  also rotates accordingly, but the sphere  $S^3$  lags behind the phase plane  $R^2$  by an angle of  $\pi/2$ .

The three-dimensional unit sphere  $S^3$  of the electron is an  $SU(2)$  group. According to Theorem 12.5.7, the exponential map from a neighborhood  $B^3$  of the zero vector in the Lie algebra of the  $SU(2)$  group to a neighborhood  $O^3$  of the identity element  $e$  of the  $SU(2)$  group is a diffeomorphism. That is, the exponential map from a neighborhood  $B^3$  of the zero vector in the tangent space of the unit sphere  $S^3$  to a neighborhood  $O^3$  of the origin of the sphere  $S^3$  (taking the identity element  $e$  of the  $SU(2)$  group as the origin of the unit sphere  $S^3$ ) is a diffeomorphism (within a sufficiently small range, for example, the radii of neighborhoods  $B^3$  and  $O^3$  are on the order of  $10^{-12}$  meters). Since neighborhoods  $B^3$  and  $O^3$  are diffeomorphic, they are identical as smooth manifolds. Therefore, all smooth tangent vector fields on neighborhood  $O^3$  can be entirely transferred to neighborhood  $B^3$ .

Neighborhood  $O^3$  is part of the unit sphere  $S^3$ , so if the unit sphere  $S^3$  rotates, neighborhood  $O^3$  naturally rotates as well. Since the unit sphere  $S^3$  and the phase plane  $R^2$  form a vector bundle  $(P, S^3, R^2, \pi, U(1))$ , the rotation of the phase plane  $R^2$  drives the rotation of the unit sphere  $S^3$  and consequently the rotation of neighborhood  $O^3$  and neighborhood  $B^3$ .

Observing equation (20.8.14), we see that a unit vector in the phase plane requires a rotation of  $2\pi$  to complete one full revolution. If we replace  $\theta$  in equation (20.8.14) with  $\frac{\theta}{2}$ , or set in equation (20.8.14)

$$\theta = \frac{\omega}{2}, \quad (20.8.17)$$

then equation (20.8.14) becomes

$$U(1) = e^{ik(\frac{\omega}{2})}. \quad (20.8.18)$$

If the range of  $\theta$  is  $0 \leq \theta \leq 2\pi$ , then the range of  $\omega$  is  $0 \leq \omega \leq 4\pi$ . Therefore, the unit vector in the phase plane described by equation (20.8.14) requires a rotation of  $2\pi$  to complete one revolution, meaning the phase plane requires a rotation of  $2\pi$  per revolution. In contrast, the unit vector described by equation (20.8.18) requires a rotation of  $4\pi$  to complete one revolution, meaning the phase plane requires a rotation of  $4\pi$  per revolution. Consequently, we have  $k = 2n$ .

Correspondingly, equations (20.8.15) and (20.8.16) should be modified to

$$f(\omega) = -\frac{2}{k}ie^{ik(\frac{\omega}{2})}, \quad (20.8.19)$$

$$f(\omega) = \frac{2}{k}e^{i(k\frac{\omega}{2} - \frac{\pi}{2})}. \quad (20.8.20)$$

In the hinge version case,  $S^3$  rotating one revolution requires an angle of  $2\pi$ , and the corresponding phase plane  $R^2$  rotating one revolution also requires  $2\pi$ . However, after modifying the parameters to the triangle version, both require an angle of  $4\pi$ .

Suppose the rotation angle of  $S^3$  from start to end is  $\Delta\omega$ , then equations (20.8.18), (20.8.19), and (20.8.20) become

$$U(1) = e^{ik(\frac{\Delta\omega}{2})}, \quad (20.8.21)$$

$$f(\omega) = -\frac{2}{k} i e^{ik(\frac{\Delta\omega}{2})}, \quad (20.8.22)$$

$$f(\omega) = \frac{2}{k} e^{i(k\frac{\Delta\omega}{2} - \frac{\pi}{2})}. \quad (20.8.23)$$

As shown in Figure 20.8.16, on the phase plane,  $\angle AOB = \pi/2$ , the lengths of the tangent vectors  $\overrightarrow{OA}$  and  $\overrightarrow{OB}$  are both equal to 1, and the distance between points  $A$  and  $B$  is  $d(A, B) = \sqrt{2}$ . The tangent vectors  $\overrightarrow{OA}$  and  $\overrightarrow{OB}$  correspond to the geodesic segments  $OA'$  and  $OB'$  on  $S^3$ , respectively, and the arc lengths of  $OA'$  and  $OB'$  are both equal to 1. According to Toponogov's comparison theorem, we can always find a triangle  $\Delta A'OB'$  on the sphere  $S^3$  such that its corresponding sides are equal to those of the triangle  $\Delta AOB$  on the phase plane. Thus, we can let  $d(A', B') = d(A, B) = \sqrt{2}$ . Then, according to Toponogov's comparison theorem,  $\angle A'OB' = \alpha' > \angle AOB = \pi/2$ . By the cosine theorem for spherical triangle sides, given in equation (11.11.1), we have

$$\cos \sqrt{2} = \cos 1 \cos 1 + \sin 1 \sin 1 \cos \alpha'.$$

Thus, the magnitude of  $\angle A'OB'$  can be found as

$$\alpha' = \arccos \left( \frac{\cos \sqrt{2} - \cos 1 \cos 1}{\sin 1 \sin 1} \right) = 1.76404317203753 = 101.072228636621^\circ. \quad (20.8.24)$$

$\pi = 3.14159265358979$  was used in the calculation.

When the phase plane  $R^2$  rotates by  $90^\circ$ , the rotation angle of sphere  $S^3$  is greater than  $90^\circ$ , so their rotations are asynchronous. When the phase plane  $R^2$  rotates by  $90^\circ$ , the extra rotation angle for  $S^3$  compared to  $90^\circ$  is

$$\alpha' - 90^\circ = 101.072228636621^\circ - 90^\circ = 11.072228636621^\circ. \quad (20.8.25)$$

Since the phase plane  $R^2$  requires an angle of  $4\pi$  (i.e., 8 sectors of  $\frac{\pi}{2}$ ) to complete one revolution, the corresponding sphere  $S^3$  needs to rotate by an angle of  $8\alpha'$  greater than  $8 \times 90^\circ$ . The extra angle is  $8 \times (\alpha' - 90^\circ)$  per revolution of  $S^3$ . If sphere  $S^3$  rotates  $k = 2n$  revolutions, the total extra rotation is

$$\begin{aligned} \Delta\omega_1 &= 2n \times 8 \times (\alpha' - 90^\circ) \\ &= 2n \times 8 \times (101.072228636621^\circ - 90^\circ) = 2n \times 88.577829092968^\circ, \end{aligned}$$

that is

$$\Delta\omega_1 = 2n \times 88.577829092968^\circ. \quad (20.8.26)$$

If equation (20.8.14) is used to describe the rotation of the phase plane in Figure 20.8.2, then the phase plane requires  $2\pi$  per revolution. Comparing the phase planes in Figure 20.8.2 and Figure 20.8.3, we find that the phase plane in Figure 20.8.3 has rotated an extra  $45^\circ$  relative to that in Figure 20.8.2, denote  $\Delta\theta_2 = \alpha = 45^\circ$ . This extra  $45^\circ$  may appear after the phase plane in Figure 20.8.2 rotates just 1 revolution, or it may appear only after  $n$  revolutions.

If equation (20.8.18) is used to describe the rotation of the phase plane in Figure 20.8.2, then

the phase plane requires  $4\pi$  per revolution. The phase plane in Figure 20.8.2 rotating 2 revolutions under equation (20.8.14) is equivalent to it rotating 1 revolution under equation (20.8.18). Suppose the extra rotation angle of the phase plane in Figure 20.8.3 compared to that in Figure 20.8.2 is  $\Delta\omega_2$ . Since  $\theta = \frac{\omega}{2}$ , we have  $\Delta\theta_2 = \frac{1}{2}\Delta\omega_2 = 45^\circ$ , giving  $\Delta\omega_2 = 90^\circ$ . This extra  $\Delta\omega_2 = 90^\circ$  appears when the phase plane in Figure 20.8.2 rotates  $k = 2n$  revolutions to become the phase plane in Figure 20.8.3.

Purely numerically,  $\Delta\omega_1$  and  $\Delta\omega_2$  should be equal, i.e.,  $\Delta\omega_1 = \Delta\omega_2$ , because equations (20.8.21) and (20.8.22) require the parameter  $\Delta\omega$  to be the same, i.e., they must belong to the same parameter  $\Delta\omega$ . From a purely mathematical perspective, it is sufficient that the numerical values of parameter  $\Delta\omega$  in these two equations are equal. Sphere  $S^3$  rotates faster, while phase plane  $R^2$  rotates slower. If sphere  $S^3$  rotates  $k = 2n$  revolutions, it rotates an extra  $\Delta\omega_1 = 2n \times 88.577829092968^\circ$ , which is greater than the extra rotation angle  $\Delta\omega_2 = 90^\circ$  of phase plane  $R^2$ , especially when  $n$  is large,  $\Delta\omega_1$  is far greater than  $\Delta\omega_2$ . Since equations (20.8.21) and (20.8.22) require the parameter  $\Delta\omega$  to be equal,  $90^\circ$  must be changed to a large number. We can convert  $90^\circ$  to  $90 \times 60$  minutes, expanding it by a factor of 60. But when  $n$  is a large number, e.g.,  $n = 100$ ,  $2 \times 100 \times 88.577829092968 \gg 90 \times 60$ ,  $90 \times 60$  is still relatively small compared to the needed total. Therefore, we ultimately must convert  $90^\circ$  to

$$\Delta\omega_2 = 90 \times 60 \times 60$$

seconds ("), the largest number. This approach is feasible because, purely mathematically, the parameter  $\Delta\omega$  in equations (20.8.21) and (20.8.22) must be numerically equal, but there is no requirement that the units of  $\Delta\omega$  in these two equations be the same. In mathematics, to make the conclusion universal, it is unnecessary to consider specific physical units; for example,  $\Delta\omega$  can be treated as a dimensionless quantity. From a physical perspective, we need to assign units to physical quantities. We can consider that for sphere  $S^3$ , the unit of parameter  $\Delta\omega$  should be degrees ( $^\circ$ ), and for phase plane  $R^2$ , the unit of parameter  $\Delta\omega$  should be seconds ("). This is acceptable.

Previously, we required  $\angle AOB \neq \angle A'OB'$  in Figure 20.8.16, which is equivalent to requiring  $\Delta\omega_1 \neq \Delta\omega_2$ . However, we now also require  $\Delta\omega_1 = \Delta\omega_2$ . This creates a contradiction. The only way to resolve this contradiction is to let  $\Delta\omega_1$  and  $\Delta\omega_2$  be numerically equal, but with  $\Delta\omega_1$  in units of degrees and  $\Delta\omega_2$  in units of seconds. Although it is also possible to let  $\Delta\omega_2$  be measured in minutes, doing so would cause subsequent calculations to deviate from empirical facts. On the other hand, the sphere  $S^3$  is a spherical space, while the phase plane  $R^2$  is a tangent plane; they are fundamentally different from each other. Thus, adopting different units for each does not lead to interference. This also illustrates that the sphere  $S^3$  rotates quickly, whereas the phase plane  $R^2$  rotates slowly. This is analogous to the rotation of the second hand and minute hand on a watch. Although both hands rotate through the same 30 ticks, we say the second hand has moved 30 seconds, while the minute hand has moved 30 minutes, using different units to measure their respective speeds of rotation.

Therefore,  $\Delta\omega_1 = \Delta\omega_2$ , i.e.,

$$2n \times 88.577829092968 = 90 \times 60 \times 60, \quad (20.8.27)$$

solving gives

$$n = 1828.90009451429.$$

Since  $n$  must be an integer, we finally obtain  $n = 1828$ . This indicates that if the phase plane  $R^2$  rotates

$$A = 2 \times 1828 \times 88.577829092968$$

seconds ("), then the sphere  $S^3$  rotates

$$B = 2 \times 1828 \times 88.577829092968$$

degrees ( $^\circ$ ). This also shows that if equation (20.8.13) is used to describe the rotation of phase

plane  $R^2$ , then  $R^2$  rotates 1828 revolutions. That is, the mapping degree in the asynchronous rotation case is 1828 (the rotation mapping is orientation-preserving).

Dividing both sides of equation (20.8.27) by 2, we get

$$n \times 88.577829092968 = 45 \times 60 \times 60. \quad (20.8.28)$$

If we choose parameters for phase plane  $R^2$  such that one revolution requires only  $2\pi$ , and parameters for sphere  $S^3$  such that one revolution requires  $4\pi$ , we obtain equation (20.8.28). This is similar to the relationship between equations (16.8.1) and (16.8.2) earlier.

Due to the relativity of motion, a unit vector on phase plane  $R^2$  completing one revolution is equivalent to the unit vector being stationary and the phase plane  $R^2$  completing one revolution.

When  $n = 1$ , phase plane  $R^2$  rotates  $k = 2n = 2 \times 1 = 2$  revolutions. Sphere  $S^3$  correspondingly rotates  $k = 2n = 2 \times 1 = 2$  revolutions but needs to rotate an additional:

$$\Delta\omega_1 = 2 \times 1 \times 88.577829092968$$

degrees ( $^\circ$ ). Due to the constraint  $\Delta\omega_1 = \Delta\omega_2$ , phase plane  $R^2$  must also rotate an additional:

$$\Delta\omega_2 = 2 \times 1 \times 88.577829092968$$

seconds ( $''$ ).

When  $n = 2$ , phase plane  $R^2$  rotates  $k = 2n = 2 \times 2 = 4$  revolutions. Sphere  $S^3$  correspondingly rotates  $k = 2n = 2 \times 2 = 4$  revolutions but needs to rotate an additional:

$$\Delta\omega_1 = 2 \times 2 \times 88.577829092968$$

degrees ( $^\circ$ ). Due to the constraint  $\Delta\omega_1 = \Delta\omega_2$ , phase plane  $R^2$  must also rotate an additional:

$$\Delta\omega_2 = 2 \times 2 \times 88.577829092968$$

seconds ( $''$ ).

.....

When  $n = 1828$ , phase plane  $R^2$  rotates  $k = 2n = 2 \times 1828$  revolutions. Sphere  $S^3$  correspondingly rotates  $k = 2n = 2 \times 1828$  revolutions but needs to rotate an additional:

$$\Delta\omega_1 = 2 \times 1828 \times 88.577829092968$$

degrees ( $^\circ$ ). Due to the constraint  $\Delta\omega_1 = \Delta\omega_2$ , phase plane  $R^2$  must also rotate an additional:

$$\Delta\omega_2 = 2 \times 1828 \times 88.577829092968$$

seconds ( $''$ ). At this point, phase plane  $R^2$  has rotated an extra total of:

$$A = 2 \times 1828 \times 88.577829092968$$

seconds, i.e., an extra:

$$A = 2 \times 1828 \times 88.577829092968'' / 60 / 60 = 89.9557064344142^\circ,$$

and to reach the full extra rotation  $90^\circ$ , it still needs to rotate an additional:

$$\begin{aligned} C &= 90^\circ - 89.9557064344142^\circ = 0.0442935655858^\circ \\ &= 0.0442935655858 \times 60 \times 60'' = 159.45683610888'' \end{aligned}$$

In the case of synchronous rotation,  $g_5$  rotating  $90^\circ$  and  $g_6$  rotating  $180^\circ$  are different modes of rotation. Similarly, the 1828 rotation processes mentioned above are also different modes of rotation.

Let us analyze why the integer  $n$  cannot be 1829. If  $n$  were taken as 1829, both the phase plane  $R^2$  and the sphere  $S^3$  would have to rotate an additional  $k = 2 \times 1$  revolutions. Since the phase plane  $R^2$  rotates first and the sphere  $S^3$  rotates afterward, when the phase plane  $R^2$  rotates an additional  $k = 2 \times 1$  revolutions plus

$$C = 159.45683610888''$$

according to the formula  $\Delta\omega_1 = \Delta\omega_2$ , the sphere  $S^3$  would subsequently also have to rotate an additional  $k = 2 \times 1$  revolutions plus

$$159.45683610888$$

degrees. However, based on the calculations above, when the sphere  $S^3$  rotates  $k = 2 \times 1$

revolutions, it actually rotates an extra

$$2 \times 1 \times 88.577829092968 = 177.155658185936$$

degrees. Therefore, this situation cannot occur, and thus  $n$  cannot be the integer 1829.

$n$  can only be the integer 1828. But in this case, after rotating  $k = 2n = 2 \times 1828$  revolutions, the phase plane  $R^2$  has not yet rotated the full additional angle  $90^\circ$ . Hence, the phase plane  $R^2$  needs to continue rotating further. Since the phase plane  $R^2$  rotates first, the sphere  $S^3$  rotates afterward, and the sphere  $S^3$  lags behind the phase plane  $R^2$  by  $90^\circ$ , when the phase plane  $R^2$  has not yet completed a full additional revolution but has only rotated an additional

$$C = 159.45683610888'' (< 90^\circ)$$

the phase plane  $R^2$  will have just exactly rotated the additional angle  $90^\circ$  in total. Simultaneously, the electron will have completely transformed into a proton, and consequently, the sphere  $S^3$  will stop rotating.

In summary, during the asynchronous rotation process, the phase plane  $R^2$  rotates first, the sphere  $S^3$  rotates afterward, and the sphere  $S^3$  lags behind the phase plane  $R^2$  by  $90^\circ$ . After both the phase plane  $R^2$  and the sphere  $S^3$  have rotated  $k = 2n = 2 \times 1828$  revolutions, the phase plane  $R^2$  continues to rotate an additional angle of  $C = 159.45683610888''$ , while the sphere  $S^3$  stops rotating. At this point, the extra rotation angle of the phase plane  $R^2$  is exactly  $90^\circ$  (if parameters for the phase plane  $R^2$  are chosen such that one revolution requires only an angle of  $2\pi$ , then the extra rotation angle of the phase plane  $R^2$  is exactly  $45^\circ$ ), and the electron has completely transformed into a proton.

### (3) Calculating the Mass Ratio

Now, integrating the two results above, we calculate the mass ratio.

Through the above analysis and calculations, we find that the mapping  $f$  causes the phase plane  $R^2$  and the unit sphere  $S^3$  to exhibit two types of rotation: synchronous rotation and asynchronous rotation. In the case of synchronous rotation, the degree of the mapping  $f$  is 8. In the case of asynchronous rotation, the phase plane  $R^2$  rotates 1828 full turns plus an additional  $45^\circ$ . Therefore, for the phase plane  $R^2$ , with reference to Figure 20.8.18, the degree of the mapping  $f$  can be described as

$$\begin{cases} \deg_p f = 1836, & p \in S^1, S^1 \subset R^2, 45^\circ < \theta < 360^\circ; \\ \deg_p f = 1837, & p \in S^1, S^1 \subset R^2, 0^\circ \leq \theta \leq 45^\circ. \end{cases}$$

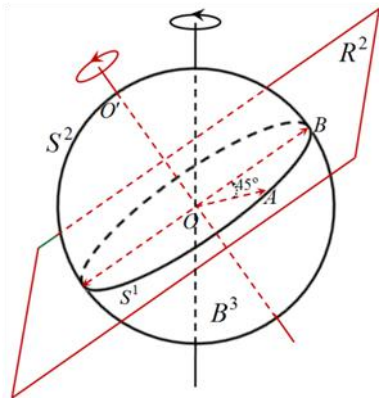


Figure 20.8.18 Degree of mapping  $f$

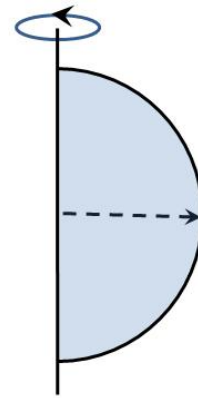


Figure 20.8.19 A three-dimensional solid open ball  $B^3$  formed by rotating a half-open disk  $D^2$  once

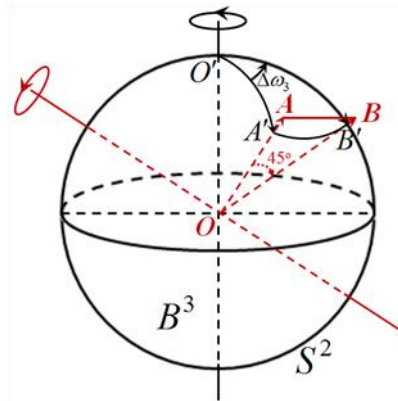
During the transformation from an electron to a proton, in the case of synchronous rotation,

the degree of the mapping  $f$  is 8, resulting in an eightfold increase in the flux through the integral surface  $S^2$  that passes through the proton. Similarly, the unit sphere  $S^3$  and the phase plane  $R^2$  undergo asynchronous rotation by 1828 full turns, causing the neighborhood  $B^3$  (the neighborhood  $O^3$ ) and the phase plane  $R^2$  to also rotate asynchronously by 1828 turns. The neighborhood  $B^3$  is an open ball, and its boundary is a two-dimensional sphere  $S^2$ , which is also the integral surface  $S^2$  in equation (20.8.8). The open ball  $B^3$  and the sphere  $S^2$  together form a three-dimensional closed solid ball  $\tilde{B}^3$ . The open ball  $B^3$  can be viewed as being generated by rotating a half-open disk  $D^2$  by one full turn, as illustrated in Figure 20.8.19. If this half-open disk  $D^2$  rotates 1828 times, it forms 1828 open balls  $B^3$ , all superimposed on one another. Consequently, the flux through the integral surface  $S^2$  passing through the proton becomes 1828 times that of the electron. Adding the eightfold increase in flux from synchronous rotation, the total flux through the integral surface  $S^2$  passing through the proton is 1836 times that of the electron.

This indicates that the rest mass of the proton is at least 1836 times that of the electron. The phrase "at least" is used because the rest mass of the proton is greater than 1836 times that of the electron, due to the fact that, compared to Figure 20.8.3, the phase plane shown in Figure 20.8.2 has not merely completed 1836 full turns but has undergone an additional rotation of  $45^\circ$ .

Since the integration surface  $S^2$  is a sphere and possesses spherical symmetry, regardless of whether the integration sphere  $S^2$  is stationary or how many turns it rotates, the total flux passing through the entire integration sphere  $S^2$  remains the same.

As shown in Figure 20.8.15, the rotation axes of the unit sphere  $S^3$ , the phase plane  $R^2$ , and the closed solid ball  $\tilde{B}^3$  are coincident, so their rotational motions are synchronized. As shown in Figure 20.8.20, if the rotation axes of the phase plane  $R^2$  and the closed solid ball  $\tilde{B}^3$  are not coincident but intersect, then the rotation axes of the phase plane  $R^2$  and the closed solid ball  $\tilde{B}^3$  are different, and they also differ from the rotation axis of the unit sphere  $S^3$ ; hence, their rotations are asynchronous.



**Figure 20.8.20** The additional rotation angle  $\Delta\omega_3 > 45^\circ$  of the integration surface  $S^2$

Due to the relativity of motion, the rotation of the phase plane  $R^2$  while the closed solid ball  $\tilde{B}^3$  remains stationary is equivalent to the phase plane  $R^2$  being stationary while the closed solid ball  $\tilde{B}^3$  rotates. In other words, the boundary (integral surface)  $S^2$  of the closed solid ball  $\tilde{B}^3$  rotates, but the rotation angle of the integral surface  $S^2$  is greater than  $45^\circ$ , because  $S^2$  is a two-dimensional curved spherical surface. Therefore, when calculating the flux through the integral surface  $S^2$ , we consider the phase plane  $R^2$  as stationary (Namely, we let the three-dimensional Euclidean coordinate system used for calculation be stationary, the phase



plane  $R^2$  is a plane within this coordinate system) while the integral surface  $S^2$  rotates. Suppose the additional rotation angle of  $S^2$  is  $\Delta\omega_3$ , as shown in Figure 20.8.20. Since the integral surface  $S^2$  must rotate further by a certain angle  $\Delta\omega_3$ , this results in an additional increase in the flux through the integral sphere  $S^2$  passing through the proton.

As shown in Figure 20.8.20, the phase plane  $R^2$  rotates by an angle  $\pi/4$ , which is equivalent to the unit vector  $\vec{OA}$  on the phase plane  $R^2$  rotating by an angle  $\pi/4$  to become unit vector  $\vec{OB}$ . Both are unit tangent vectors with length equal to 1, and the angle between them is  $\angle AOB = \pi/4$ . The vectors  $\vec{OA}$ ,  $\vec{OB}$ , and the line segment  $AB$  form a triangle  $\triangle AOB$ , lying on the phase plane  $R^2$ . The distance between points  $A$  and  $B$  is

$$d(A, B) = \sqrt{|\vec{OA}|^2 + |\vec{OB}|^2 - 2|\vec{OA}| \cdot |\vec{OB}| \cos \frac{\pi}{4}} = 0.765366864730179.$$

Let the tangent vectors  $\vec{OA}$  and  $\vec{OB}$  correspond to geodesic segments  $O'A'$  and  $O'B'$  on the integration sphere  $S^2$ , respectively, with the arc lengths of both  $O'A'$  and  $O'B'$  equal to 1. The arcs  $O'A'$ ,  $O'B'$ , and the arc  $A'B'$  form a spherical triangle  $\triangle A'O'B'$ , lying on the integration sphere  $S^2$ . According to Toponogov's comparison theorem, on the integration sphere  $S^2$  we can always find a triangle  $\triangle A'O'B'$  such that its corresponding sides are equal to those of  $\triangle AOB$ ; thus, we can set

$$d(A', B') = d(A, B),$$

where  $d(A, B)$  is the length between points  $A'$  and  $B'$ . By Toponogov's comparison theorem,  $\angle A'O'B' = \Delta\omega_3 > \angle AOB = \pi/4$ . Using the cosine theorem for spherical triangle sides from equation (11.11.1), we obtain

$$\cos(d(A', B')) = \cos 1 \cos 1 + \sin 1 \sin 1 \cos \Delta\omega_3.$$

So we obtain the magnitude of  $\angle A'O'B'$  as

$$\Delta\omega_3 = \arccos\left(\frac{\cos(d(A', B')) - \cos 1 \cos 1}{\sin 1 \sin 1}\right) = 0.919580099777479 = 52.6880586414687^\circ.$$

The proportion of  $\Delta\omega_3$  to a full circle  $360^\circ$  is

$$\frac{\Delta\omega_3}{360^\circ} = \frac{52.6880586414687^\circ}{360^\circ} = 0.14635571845.$$

Rotating by an additional angle  $\Delta\omega_3$  is equivalent to copying  $\frac{\Delta\omega_3}{360^\circ}$  of the flux passing through the integration sphere  $S^2$  in the tangent space of the point-like electron and pasting it into the solid closed ball  $\tilde{B}^3$  bounded by the integration sphere  $S^2$  within the tangent space of the point-like proton. This increases the proton's rest mass by  $\frac{\Delta\omega_3}{360^\circ}$  times the electron's mass.

Therefore, the proton's rest mass is

$$1836 + 0.14635571845 = 1836.14635571845$$

times the electron's rest mass. Experimentally, the proton's rest mass is measured to be 1836.152701(37) times that of the electron. The error is

$$\frac{1836.152701 - 1836.14635571845}{1836.152701} \times 100\% = 0.000345574828712222\%.$$

Recently, German scientists measured the proton's rest mass to be 1836.152673414(47) times that of the electron. The error is then

$$\frac{1836.152673414 - 1836.14635571845}{1836.152673414} \times 100\% = 0.00034407244821999\%.$$

The calculated value is very close to the experimental value.



#### **4.Equal Rest Masses of Particles and Antiparticles**

Since the expressions for the spin operators of a particle and its antiparticle are identical, the rest masses of a particle and its antiparticle are equal.

## Chapter 21 Transformation of Particles

### §21.1 Influence of Photons on the Energy of Other Particles

Let the Lie group corresponding to a particle be  $G$ , and let  $\varphi$  be a Lie transformation group acting on  $G$  from the left. Thus, there exists a smooth mapping  $f: \varphi \times G \rightarrow G$ . Therefore, the state of the particle before the action of  $\varphi$  (denoted as  $G_{\text{before}}$ ) and its state after the action (denoted as  $G_{\text{after}}$ ) are different. One can also consider that before the action of  $\varphi$ , the particle is the Lie group  $G$ , and after the action, it becomes the Lie group  $\varphi \times G$ , which is the direct product of these two Lie groups  $\varphi$  and  $G$ . Conversely, suppose the Lie group corresponding to a certain particle at present is  $G_{\text{after}}$ . Then, it can be considered that this Lie group  $G_{\text{after}}$  is obtained by the action of the Lie transformation group  $\varphi$  on the particle in its previous state  $G_{\text{before}}$ , i.e.,  $f: \varphi \times G_{\text{before}} \rightarrow G_{\text{after}}$ . Therefore, the particle in its current state ( $G_{\text{after}}$ ) can be viewed as the particle that was in the prior state  $G_{\text{before}}$  after being acted upon by the transformation group  $\varphi$ . The following examples illustrate this.

When an atom absorbs a photon, does this photon disappear? We can say no; the absorbed photon remains inside the atom. Where does the photon emitted by the atom come from? We can say that the emitted photon is precisely the one that was absorbed by the atom. The reasons are as follows:

1. The angular momentum operator describing the electron in the atom is consistent with the spin operator of the photon.

2. The selection rules for electric dipole transitions of the electron in the atom are:

$$\Delta l = l' - l = \pm 1, \\ \Delta m = m' - m = 0, \pm 1,$$

where  $l$  is the angular quantum number, and  $m$  is the magnetic quantum number, representing the projection of the electron's angular momentum on the  $z$ -axis,  $m = 0, \pm 1, \pm 2, \dots, \pm l$ .

These selection rules can be explained using the spin angular momentum of the photon. When an electron absorbs a photon, the electron and the photon form a composite particle. The angular quantum number of this composite particle is the electron's angular quantum number plus 1, because the photon's angular quantum number equals 1. Simultaneously, the electron's energy increases by the energy  $h\nu$  of the absorbed photon.

For example, when an electron with angular quantum number  $l = 1$  and magnetic quantum number  $m = 1$  absorbs a photon, the angular quantum number of the composite particle (electron plus photon) becomes

$$l' = 1 + 1 = 2,$$

and the magnetic quantum number  $m'$  can be one of  $m' = 0, \pm 1, \pm 2$ . This is because:

1. The electron's angular quantum number  $l' = 2$  indicates that the electron has absorbed 2 photons.

2. In its free state, a photon's spin  $z$ -component can be  $+\hbar$  or  $-\hbar$ , and inside an atom it can also be 0. If the spin  $z$ -components of both photons are  $+\hbar$  or  $-\hbar$ , their sum is  $+2\hbar$  or  $-2\hbar$ , and thus the magnetic quantum number  $m'$  equals  $+2$  or  $-2$ . If one photon has a spin  $z$ -component of  $+\hbar$  or  $-\hbar$  and the other has 0, their sum is  $+\hbar$  or  $-\hbar$ , making  $m' = +1$  or  $m' = -1$ . If one photon has  $+\hbar$  and the other has  $-\hbar$ , their sum is 0, making  $m' = 0$ .

When the electron's angular quantum number is  $l = 3$ , it can be understood that this electron is not a simple electron but a composite particle formed by combining with 3 photons. When this composite particle emits a photon, its angular quantum number decreases by 1, because the photon's angular quantum number equals 1. Thus:

$$l' = 3 - 1 = 2.$$

Simultaneously, the energy of this composite particle decreases by the energy of the emitted photon.

A composite particle formed by an electron ( $SU(2)$  group) absorbing a photon ( $SO(3)$  group) can be regarded as a direct inner product Lie group  $SO(3) \otimes SU(2)$ . This inner product Lie group is homeomorphic to the  $SU(2)$  group (see the content in §15.1). Therefore, an electron that has absorbed a large number of photons and increased its energy is still considered the same particle by us, but in reality, it is a composite particle combined with numerous photons.

In summary, after an electron absorbs photons, its energy increases, and the absorbed photons combine with the electron to form a composite particle. After an electron emits photons, its energy decreases, and the emitted photons are precisely those previously absorbed.

From the perspective of Lie transformation groups, the direct inner product Lie group  $SO(3) \otimes SU(2)$  can be viewed as the following mapping process:

$$\theta: SO(3) \otimes SU(2) \rightarrow SU(2). \quad (21.1.1)$$

Thus, an electron that has absorbed a photon can still be considered an electron, albeit one that has been acted upon (or transformed) by the photon (resulting in changes in angular momentum, energy, etc.). The process of an electron emitting a photon can also be seen as the mapping process (21.1.1), because after emitting the photon, the electron's energy, angular momentum, etc., have changed, all of which can be attributed to the influence of the emitted photon.

Can we consider that a free particle (now denoted by  $M$  representing either an  $SU(2)$  or  $SO(3)$  group) possesses a certain energy (greater than its rest energy) because it has absorbed photons, and that a decrease in energy is due to emitting photons? The answer is yes, because we can consider that a free particle  $M$  having a certain energy results from the following mapping process:

$$\varphi: SO(3) \otimes M \rightarrow M. \quad (21.1.2)$$

This process is the (left) action of the photon as a Lie transformation group on the particle  $M$ , which may increase or decrease the energy of  $M$ .  $SO(3) \otimes M$  can be viewed as the (left) action of the  $SO(3)$  group on particle  $M$ , or as the formation of a composite particle  $SO(3) \otimes M$  by the photon ( $SO(3)$  group) and particle  $M$ ; that is, the particle  $M$  absorbs a photon and thereby increases its own energy. Conversely,  $SO(3) \otimes M$  can also be seen as the process where particle  $M$  emits a photon and thereby decreases its own energy.

## §21.2 Mutual Transformation of Particles

The photon–antiphoton pair  $\gamma$  is an  $SO(3)$  group, while the electron–positron pair  $e^+$  and  $e^-$ , the proton–antiproton pair  $p$  and  $\bar{p}$ , and the neutrino–antineutrino pair  $\nu$  and  $\bar{\nu}$  are each an  $SU(2)$  group. These elementary particles can transform into one another, and in the process of transformation, the laws of conservation of energy, momentum, angular momentum, and charge must of course be obeyed.

According to the content of Chapter 12, there exists a homomorphic mapping from the  $SU(2)$  group to the  $SO(3)$  group. This homomorphism is in fact a 2:1 mapping: for every element of the  $SO(3)$  group, there are two elements of the  $SU(2)$  group that correspond to it. Therefore, two  $SU(2)$  groups with completely opposite orientations can combine to form one  $SO(3)$  group. Translating this to particles, two particle–antiparticle pairs—such as electron–positron, proton–antiproton, or neutrino–antineutrino—can combine to form a photon, i.e.,

$$e^+ + e^- \rightarrow \gamma, \quad (21.2.1)$$

$$p + \bar{p} \rightarrow \gamma, \quad (21.2.2)$$

$$\nu + \bar{\nu} \rightarrow \gamma. \quad (21.2.3)$$

Conversely, a photon can transform into an elementary spin-1/2 particle–antiparticle pair, i.e.,

$$\gamma \rightarrow e^+ + e^-, \quad (21.2.4)$$

$$\gamma \rightarrow p + \bar{p}, \quad (21.2.5)$$

$$\gamma \rightarrow \nu + \bar{\nu}. \quad (21.2.6)$$

According to equation (12.12.3), the direct product of two  $SO(3)$  groups equals the direct

product of two  $SU(2)$  groups, i.e.,

$$SO(3) \times SO(3) = SU(2) \times SU(2). \quad (21.2.7)$$

The direct product of groups essentially represents the mutual transformation relation between two groups. For particles, it describes the interaction between two particles; in other words, the interaction of two particles can be described by the direct product of the two Lie groups corresponding to them. The equality of the direct product of two  $SO(3)$  groups to the direct product of two  $SU(2)$  groups tells us that after two photons interact, they can transform into an interaction of two elementary spin-1/2 particles—that is, they can transform into two elementary spin-1/2 particles. Due to the need to obey the four conservation laws, they generally transform into a particle-antiparticle pair, i.e.,

$$\gamma + \gamma \rightarrow e^+ + e^-, \quad (21.2.8)$$

$$\gamma + \gamma \rightarrow p + \bar{p}, \quad (21.2.9)$$

$$\gamma + \gamma \rightarrow \nu + \bar{\nu}. \quad (21.2.10)$$

Conversely, a pair of elementary spin-1/2 particle and antiparticle can also transform into two photons, i.e.,

$$e^+ + e^- \rightarrow \gamma + \gamma,$$

$$p + \bar{p} \rightarrow \gamma + \gamma,$$

$$\nu + \bar{\nu} \rightarrow \gamma + \gamma.$$

Since

$$SU(2) \times SU(2) = SU(2) \times SU(2),$$

a pair of elementary spin-1/2 particle and antiparticle can also transform into another pair of elementary spin-1/2 particle and antiparticle, i.e.,

$$e^+ + e^- \rightarrow p + \bar{p}, \quad p + \bar{p} \rightarrow e^+ + e^-;$$

$$\nu + \bar{\nu} \rightarrow e^+ + e^-, \quad e^+ + e^- \rightarrow \nu + \bar{\nu};$$

$$\nu + \bar{\nu} \rightarrow p + \bar{p}, \quad p + \bar{p} \rightarrow \nu + \bar{\nu}.$$

According to the content of Chapter 14, under certain conditions, by reversing the orientation of particles, an antiproton and a positron can simultaneously transform into a proton and an electron, i.e.,

$$\bar{p} \rightarrow p, \quad e^+ \rightarrow e^-.$$

Conversely, a proton and an electron can simultaneously transform into an antiproton and a positron, i.e.,

$$p \rightarrow \bar{p}, \quad e^- \rightarrow e^+.$$

### §21.3 Analysis of Examples

Using the reaction formulas presented earlier, we can determine which reactions can occur and which cannot. Several examples are provided below for illustration.

**Example 21.3.1** A free photon cannot spontaneously transform into a pair of elementary spin-1/2 particle-antiparticle in isolation; external conditions are required for this conversion to occur. For instance, a photon can transform into a pair of positron and electron only when passing near an atomic nucleus, i.e.,

$$\gamma \xrightarrow{\text{(near an atomic nucleus)}} e^+ + e^-.$$

This is a phenomenon that has been experimentally observed.

**Example 21.3.2** In 1955, Emilio Segrè and his collaborators bombarded a hydrogen target with a 6.2 GeV proton beam from a high-energy accelerator, producing the following reaction:

$$p + p \rightarrow p + p + (\bar{p} + p).$$

In this reaction, a proton-antiproton pair  $\bar{p}$  and  $p$  is produced. This can be understood as resulting from the decomposition of photons within the proton beam, as described by reaction

(21.2.5). After observing this reaction, one might naturally ask: why does the following reaction not occur?

$$p + p \rightarrow p + p + \bar{p} + p + p + n.$$

Since the neutron is composed of a proton  $p$ , an electron  $e^-$ , and an antineutrino  $\bar{\nu}_e$ , for this reaction to occur, it must simultaneously be accompanied by the production of an antiproton  $\bar{p}$ , a positron  $e^+$ , and a neutrino  $\nu_e$ . It is impossible to produce only the proton  $p$ , electron  $e^-$ , and antineutrino  $\bar{\nu}_e$  without also producing the antiproton  $\bar{p}$ , positron  $e^+$ , and neutrino  $\nu_e$ .

**Example 21.3.3** In 1957, the antineutron  $\bar{n}$  was discovered in the following reaction:

$$\bar{p} + p \rightarrow \bar{n} + n.$$

This reaction can occur because during the collision process, photons within the high-energy antiproton  $\bar{p}$  beam undergo reactions described by (21.2.8) and (21.2.10), or alternatively by (21.2.4) and (21.2.6). The resulting positron  $e^+$  and neutrino  $\nu_e$  combine with the antiproton  $\bar{p}$  to form the antineutron  $\bar{n}$ , while the electron  $e^-$  and antineutrino  $\bar{\nu}_e$  combine with the proton  $p$  to form the neutron  $n$ .

**Example 21.3.4** In 1954, Fowler et al. directed a  $\pi^-$  beam with an energy of 1.5 GeV into a cloud chamber containing hydrogen gas and observed the following reaction and subsequent decay modes:

$$\pi^- + p \rightarrow \Lambda^0 + K^0,$$

$$\Lambda^0 \rightarrow \pi^- + p,$$

$$K^0 \rightarrow \pi^+ + \pi^-,$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu,$$

$$\pi^+ \rightarrow \mu^+ + \nu_\mu,$$

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu,$$

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu.$$

Using a reverse deduction method—tracing back from the final decay products—we find that photons within the 1.5 GeV  $\pi^-$  beam induced reactions (21.2.4) and (21.2.6). The produced  $e^+$ ,  $\nu_e$ , and  $\bar{\nu}_\mu$  combine to form a  $\mu^+$  particle, while  $e^-$ ,  $\bar{\nu}_e$ , and  $\nu_\mu$  combine to form a  $\mu^-$  particle. The produced  $\nu_\mu$  and  $\mu^+$  combine to form a  $\pi^+$  meson, and  $\bar{\nu}_\mu$  and  $\mu^-$  combine to form a  $\pi^-$  meson. The  $\pi^+$  and  $\pi^-$  mesons then combine to form a  $K^0$  meson, while the  $\pi^-$  and  $p$  together constitute the observed composite particle  $\Lambda^0$ .

## §21.4 Examination of the Solar Nuclear Reaction Model

To date, the solar nuclear reaction model has been extensively validated by observational experiments. In this section, we analyze the solar nuclear reaction model based on the two fundamental hypotheses proposed in this book and the mutual transformation relations of elementary particles mentioned in §21.2.

The nuclear reaction process inside the Sun is divided into the following four steps.

Step 1: Two protons  $p$  form deuterium  ${}^2_1\text{H}_1$ :

$$p + p \rightarrow {}^2_1\text{H}_1 + e^+ + \nu_e, \quad (21.4.1)$$

$$p + p + e^- \rightarrow {}^2_1\text{H}_1 + \nu_e. \quad (21.4.2)$$

Step 2: Deuterium plus a proton forms  ${}^3_2\text{He}_1$ :

$${}^2_1\text{H}_1 + p \rightarrow {}^3_2\text{He}_1 + \gamma. \quad (21.4.3)$$

Step 3: Formation of  ${}^4_2\text{He}_2$  or  ${}^7_4\text{Be}_3$  particles via  ${}^3_2\text{He}_1$ :

$${}^3_2\text{He}_1 + p \rightarrow {}^4_2\text{He}_2 + e^+ + \nu_e, \quad (21.4.4)$$

$${}^3_2\text{He}_1 + {}^3_2\text{He}_1 \rightarrow {}^4_2\text{He}_2 + p + p, \quad (21.4.5)$$

$${}^3_2\text{He}_1 + {}^4_2\text{He}_2 \rightarrow {}^7_4\text{Be}_3 + \gamma. \quad (21.4.6)$$

Step 4: Formation of  ${}^4_2\text{He}_2$  particles via Be:

$${}^7_4\text{Be}_3 + e^- \rightarrow {}^7_3\text{Li}_4 + \nu_e, \quad (21.4.7)$$

$${}^7_3\text{Li}_4 + p \rightarrow {}^4_2\text{He}_2 + {}^4_2\text{He}_2, \quad (21.4.8)$$

$${}^7_4\text{Be}_3 + p \rightarrow {}^8_5\text{B}_3 + \gamma, \quad (21.4.9)$$

$${}^8_5\text{B}_3 \rightarrow {}^8_4\text{Be}_4^* + e^+ + \nu_e, \quad (21.4.10)$$

$${}^8_4\text{Be}_4^* \rightarrow {}^4_2\text{He}_2 + {}^4_2\text{He}_2. \quad (21.4.11)$$

The deuterium nucleus  ${}^2_1\text{H}_1$  consists of one proton  $p$  and one neutron  $n$ . The neutron  $n$  is composed of a proton  $p$ , an electron  $e^-$ , and an antineutrino  $\bar{\nu}_e$ . The right-hand side of the first reaction (21.4.1)

$$p + p \rightarrow {}^2_1\text{H}_1 + e^+ + \nu_e$$

can be broken down into the particle set  $\{p, \underbrace{p, e^-, \bar{\nu}_e}_n, e^+, \nu_e\}$ . This suggests that when two protons collide, photons within the protons undergo the following reactions:

$$\gamma \rightarrow e^- + e^+, \quad \gamma \rightarrow \nu_e + \bar{\nu}_e,$$

or alternatively,

$$\gamma + \gamma \rightarrow e^- + e^+, \quad \gamma + \gamma \rightarrow \nu_e + \bar{\nu}_e.$$

The produced electron  $e^-$ , antineutrino  $\bar{\nu}_e$ , and a proton  $p$  form a neutron  $n$ , which then combines with another proton to form the deuterium nucleus  ${}^2_1\text{H}_1$ , while the positron  $e^+$  and neutrino  $\nu_e$  are released.

The right-hand side of the second reaction (21.4.2)

$$p + p + e^- \rightarrow {}^2_1\text{H}_1 + \nu_e$$

can be broken down into the particle set  $\{p, \underbrace{p, e^-, \bar{\nu}_e}_n, \nu_e\}$ . This suggests that when the two protons on the left side of the reaction collide with an electron, one of the photons they carry undergoes:

$$\gamma \rightarrow \bar{\nu}_e + \nu_e,$$

or two photons undergo:

$$\gamma + \gamma \rightarrow \bar{\nu}_e + \nu_e.$$

The produced antineutrino  $\bar{\nu}_e$ , electron  $e^-$ , and a proton  $p$  form a neutron  $n$ , which then combines with another proton to form the deuterium nucleus  ${}^2_1\text{H}_1$ , while the neutrino  $\nu_e$  is released.

The  ${}^3_2\text{He}_1$  nucleus contains two protons and one neutron. The left-hand side of the third reaction (21.4.3)

$${}^2_1\text{H}_1 + p \rightarrow {}^3_2\text{He}_1 + \gamma$$

can be broken down into the particle set  $\{p, \underbrace{p, e^-, \bar{\nu}_e}_n, p\}$ , and the right-hand side can be broken

down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \gamma\}$ . The right side has one additional photon  $\gamma$ . This

suggests that this photon is radiated when the deuterium nucleus  ${}^2_1\text{H}_1$  collides with a proton.

The  ${}^4_2\text{He}_2$  nucleus contains two protons and two neutrons. The left-hand side of the fourth reaction (21.4.4)

$${}^3_2\text{He}_1 + p \rightarrow {}^4_2\text{He}_2 + e^+ + \nu_e$$

can be broken down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, p\}$ , and the right-hand side can be broken

down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, e^+, \nu_e\}$ . The right side has four additional

particles:  $e^-, e^+, \nu_e$ , and  $\bar{\nu}_e$ . This suggests that when the  ${}^3_2\text{He}_1$  nucleus collides with a proton  $p$ , two photons undergo the following reactions:

$$\gamma \rightarrow e^- + e^+, \quad \gamma \rightarrow \nu_e + \bar{\nu}_e,$$

or four photons undergo:

$$\gamma + \gamma \rightarrow e^- + e^+, \quad \gamma + \gamma \rightarrow \nu_e + \bar{\nu}_e.$$

One electron  $e^-$  and antineutrino  $\bar{\nu}_e$  combine with a proton  $p$  to form a neutron  $n$ , which then joins the  ${}^3_2\text{He}_1$  nucleus to form the  ${}^4_2\text{He}_2$  nucleus, while the other positron  $e^+$  and neutrino  $\nu_e$  are released.

The left-hand side of the fifth reaction (21.4.5)

$${}^3_2\text{He}_1 + {}^3_2\text{He}_1 \rightarrow {}^4_2\text{He}_2 + p + p$$

can be broken down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n; p, p, \underbrace{p, e^-, \bar{\nu}_e}_n\}$ , and the right-hand side can

be broken down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; p, p\}$ . The number of particles on

both sides is equal. This suggests that this reaction involves two  ${}^3_2\text{He}_1$  nuclei colliding and merging into a  ${}^4_2\text{He}_2$  nucleus, releasing two protons  $p$ .

The beryllium nucleus  ${}^7_4\text{Be}_3$  consists of four protons and three neutrons. The left-hand side of the sixth reaction (21.4.6)

$${}^3_2\text{He}_1 + {}^4_2\text{He}_2 \rightarrow {}^7_4\text{Be}_3 + \gamma$$

can be broken down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n; p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n\}$ , and the

right-hand side can be broken down into the particle set  $\{p, p, p, p; \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \gamma\}$ . The right side has one additional photon. This suggests

that the  ${}^3_2\text{He}_1$  nucleus and  ${}^4_2\text{He}_2$  nucleus collide and merge to form the beryllium nucleus  ${}^7_4\text{Be}_3$ , while emitting a photon.

The lithium nucleus  ${}^7_3\text{Li}_4$  consists of three protons and four neutrons. The left-hand side of the seventh reaction (21.4.7)

$${}^7_4\text{Be}_3 + e^- \rightarrow {}^7_3\text{Li}_4 + \nu_e$$

can be broken down into the particle set  $\{p, p, p, p; \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, e^-\}$ , and the

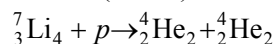
right-hand side can be broken down into the particle set  $\{p, p, p, p; \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \nu_e\}$ . This suggests that when the beryllium nucleus

${}^7_4\text{Be}_3$  collides with an electron  $e^-$ , one or two photons transform into a neutrino  $\nu_e$  and an antineutrino  $\bar{\nu}_e$ :

$$\gamma \rightarrow \nu_e + \bar{\nu}_e, \quad \text{or} \quad \gamma + \gamma \rightarrow \nu_e + \bar{\nu}_e,$$

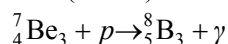
Then, a proton  $p$  in the beryllium nucleus  ${}^7_4\text{Be}_3$  combines with the electron  $e^-$  from the left side of the reaction and the produced antineutrino  $\bar{\nu}_e$  to form a neutron  $n$ , transforming  ${}^7_4\text{Be}_3$  into  ${}^7_3\text{Li}_4$ , while the neutrino  $\nu_e$  is released.

The left-hand side of the eighth reaction (21.4.8)



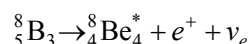
can be broken down into the particle set  $\{p, p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; p\}$ , and the right-hand side can be broken down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n\}$ . This suggests that the lithium nucleus  ${}^7_3\text{Li}_4$  collides and merges with a proton  $p$ , then splits into two  ${}^4_2\text{He}_2$  nuclei.

The left-hand side of the ninth reaction (21.4.9)



can be broken down into the particle set  $\{p, p, p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; p\}$ , and the right-hand side can be broken down into the particle set  $\{p, p, p, p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; \gamma\}$ . The right side has one additional photon  $\gamma$ . This suggests that the beryllium nucleus  ${}^7_4\text{Be}_3$  collides and merges with a proton  $p$  to form the boron nucleus  ${}^8_5\text{B}_3$ , emitting a photon.

The left-hand side (boron nucleus  ${}^8_5\text{B}_3$ ) of the tenth reaction (21.4.10)

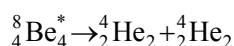


can be broken down into the particle set  $\{p, p, p, p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n\}$ , and the right-hand side can be broken down into the particle set  $\{p, p, p, p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; e^+, \nu_e\}$ . The right side has four additional particles:  $e^-$ ,  $e^+$ ,  $\nu_e$ , and  $\bar{\nu}_e$ . This suggests that the boron nucleus  ${}^8_5\text{B}_3$  undergoes decay, during which two photons within it produce the following reactions:

$$\gamma \rightarrow e^- + e^+, \quad \gamma \rightarrow \nu_e + \bar{\nu}_e.$$

The produced electron  $e^-$  and antineutrino  $\bar{\nu}_e$  combine with a proton  $p$  in  ${}^8_5\text{B}_3$  to form a neutron  $n$ , resulting in the  ${}^8_4\text{Be}_4^*$  nucleus, while the positron  $e^+$  and neutrino  $\nu_e$  are released.

The left-hand side (beryllium nucleus  ${}^8_4\text{Be}_4^*$ ) of the eleventh reaction (21.4.11)



can be broken down into the particle set  $\{p, p, p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n\}$ , and the right-hand side can be broken down into the particle set  $\{p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n; p, p, \underbrace{p, e^-, \bar{\nu}_e}_n, \underbrace{p, e^-, \bar{\nu}_e}_n\}$ . The particles are identical on both sides. This suggests that the beryllium nucleus  ${}^8_4\text{Be}_4^*$  undergoes  $\alpha$  decay, splitting into two helium nuclei  ${}^4_2\text{He}_2$ .

Through the above examination, we find that this solar nuclear reaction model is entirely consistent with the mutual transformation relations of elementary particles proposed in §21.2, as well as with the first hypothesis presented in this book. This is one of the reasons why the solar



nuclear reaction model has been successful.

## §21.5 The Internal Mechanism of Electron-to-Proton Transformation

This section builds on Chapter 20 to explore the internal mechanism by which an electron transforms into a proton in the reaction

$$e^+ + e^- \rightarrow p + \bar{p}.$$

On the phase plane  $\xi\eta$ , the three spin operators of the electron  $e^-$  can be expressed as

$$\hat{S}_x^{e^-} = \frac{1}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} + \xi\frac{\partial}{\partial\eta}\right), \quad \hat{S}_y^{e^-} = \frac{i}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right), \quad \hat{S}_z^{e^-} = \frac{1}{2}\hbar\left(\xi\frac{\partial}{\partial\xi} - \eta\frac{\partial}{\partial\eta}\right).$$

The proton is also an  $SU(2)$  group. Let its three spin operators be  $\hat{S}_x^p, \hat{S}_y^p, \hat{S}_z^p$ . We represent the proton's spin operators as vector fields on the phase plane  $uv$ .

The electron is a principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$  whose structure group  $G_{\hat{S}_3}$  is determined by the spin operator  $\hat{S}_z^{e^-}$ . The proton is a principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$  whose structure group  $G_{\hat{S}_1}$  is determined by the spin operator  $\hat{S}_x^p$ . If the spin operator  $\hat{S}_z^{e^-}$  that defines the structure group of the electron's principal bundle transforms into the spin operator  $\hat{S}_x^p$ , then the electron transforms into a proton. Therefore, by studying how the electron's spin operator transforms into the proton's spin operator, we can understand the internal mechanism of the electron-to-proton transformation. In what follows we assume that an electron transforms into a proton, and we derive how the three spin operators  $\hat{S}_x^{e^-}, \hat{S}_y^{e^-}, \hat{S}_z^{e^-}$  of the electron are transformed into the spin operators of the proton.

### 1. Transformation Formula

The transformation formula that converts the electron's spin operators into the proton's spin operators is given by equation (20.4.6). Replacing  $(\tilde{u}, \tilde{v})$  in (20.4.6) with  $(\xi, \eta)$ , we obtain

$$\begin{cases} u = \xi - \eta, \\ v = \xi + \eta. \end{cases} \quad (21.5.1)$$

Let  $F$  denote the mapping represented by equation (21.5.1). Written in matrix form, (21.5.1) becomes

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad (21.5.2)$$

and its inverse transformation is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \quad (21.5.3)$$

that is,

$$\begin{cases} \xi = \frac{1}{2}u + \frac{1}{2}v, \\ \eta = -\frac{1}{2}u + \frac{1}{2}v. \end{cases} \quad (21.5.4)$$

### 2. Transformation of the Electron's Spin Operators

#### (1) Transformation of the electron's spin operator $\hat{S}_x^{e^-}$

We transform the electron's spin operator  $\hat{S}_x^{e^-}$  using equation (21.5.4). From the expression

of  $\hat{S}_x^{e^-}$  we obtain the system

$$\begin{cases} \frac{d\xi}{dt} = \frac{1}{2}\hbar\eta, \\ \frac{d\eta}{dt} = \frac{1}{2}\hbar\xi. \end{cases} \quad (21.5.5)$$

Substituting (21.5.4) into (21.5.5) gives

$$\begin{cases} \frac{d\xi}{dt} = \frac{1}{2}\frac{du}{dt} + \frac{1}{2}\frac{dv}{dt} = \frac{1}{2}\hbar\left(-\frac{1}{2}u + \frac{1}{2}v\right), \\ \frac{d\eta}{dt} = -\frac{1}{2}\frac{du}{dt} + \frac{1}{2}\frac{dv}{dt} = \frac{1}{2}\hbar\left(\frac{1}{2}u + \frac{1}{2}v\right). \end{cases}$$

That is,

$$\begin{cases} \frac{du}{dt} + \frac{dv}{dt} = -\frac{1}{2}\hbar u + \frac{1}{2}\hbar v, \\ -\frac{du}{dt} + \frac{dv}{dt} = \frac{1}{2}\hbar u + \frac{1}{2}\hbar v. \end{cases} \quad (21.5.6)$$

Subtracting the two equations in (21.5.6) yields  $2\frac{du}{dt} = -\hbar u$ , i.e.,

$$\frac{du}{dt} = -\frac{1}{2}\hbar u. \quad (21.5.7)$$

Adding the two equations in (21.5.6) yields  $2\frac{dv}{dt} = \hbar v$ , i.e.,

$$\frac{dv}{dt} = \frac{1}{2}\hbar v. \quad (21.5.8)$$

From (21.5.7) and (21.5.8) we get

$$\hat{S}_x^p = \frac{1}{2}\hbar\left(-u\frac{\partial}{\partial u} + v\frac{\partial}{\partial v}\right),$$

or

$$\hat{S}_x^p = -\frac{1}{2}\hbar\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right). \quad (21.5.9)$$

## (2) Transformation of the electron's spin operator $\hat{S}_y^{e^-}$

Next we transform the electron's spin operator  $\hat{S}_y^{e^-}$ . From its expression we obtain the system

$$\begin{cases} \frac{d\xi}{dt} = \frac{i}{2}\hbar\eta, \\ \frac{d\eta}{dt} = -\frac{i}{2}\hbar\xi. \end{cases} \quad (21.5.10)$$

Substituting (21.5.4) into (21.5.10) gives

$$\begin{cases} \frac{d\xi}{dt} = \frac{1}{2}\frac{du}{dt} + \frac{1}{2}\frac{dv}{dt} = \frac{i}{2}\hbar\left(-\frac{1}{2}u + \frac{1}{2}v\right), \\ \frac{d\eta}{dt} = -\frac{1}{2}\frac{du}{dt} + \frac{1}{2}\frac{dv}{dt} = -\frac{i}{2}\hbar\left(\frac{1}{2}u + \frac{1}{2}v\right). \end{cases}$$

That is,

$$\begin{cases} \frac{du}{dt} + \frac{dv}{dt} = -\frac{i}{2}\hbar u + \frac{i}{2}\hbar v, \\ -\frac{du}{dt} + \frac{dv}{dt} = -\frac{i}{2}\hbar u - \frac{i}{2}\hbar v. \end{cases} \quad (21.5.11)$$

Subtracting the two equations in (21.5.11) yields  $2\frac{du}{dt} = i\hbar v$ , i.e.,

$$\frac{du}{dt} = \frac{i}{2} \hbar v. \quad (21.5.12)$$

Adding the two equations in (21.5.11) yields  $2 \frac{dv}{dt} = -i \hbar u$ , i.e.,

$$\frac{dv}{dt} = -\frac{i}{2} \hbar u. \quad (21.5.13)$$

From (21.5.12) and (21.5.13) we obtain

$$\hat{S}_y^p = \frac{i}{2} \hbar \left( v \frac{\partial}{\partial u} - u \frac{\partial}{\partial v} \right). \quad (21.5.14)$$

### (3) Transformation of the electron's spin operator $\hat{S}_z^{e^-}$

Finally, we transform the electron's spin operator  $\hat{S}_z^{e^-}$ . From its expression we obtain the system

$$\begin{cases} \frac{d\xi}{dt} = \frac{1}{2} \hbar \xi, \\ \frac{d\eta}{dt} = -\frac{1}{2} \hbar \eta. \end{cases} \quad (21.5.15)$$

Substituting (21.5.4) into (21.5.15) gives

$$\begin{cases} \frac{d\xi}{dt} = \frac{1}{2} \frac{du}{dt} + \frac{1}{2} \frac{dv}{dt} = \frac{1}{2} \hbar \left( \frac{1}{2} u + \frac{1}{2} v \right), \\ \frac{d\eta}{dt} = -\frac{1}{2} \frac{du}{dt} + \frac{1}{2} \frac{dv}{dt} = -\frac{1}{2} \hbar \left( -\frac{1}{2} u + \frac{1}{2} v \right). \end{cases}$$

That is,

$$\begin{cases} \frac{du}{dt} + \frac{dv}{dt} = \frac{1}{2} \hbar u + \frac{1}{2} \hbar v, \\ -\frac{du}{dt} + \frac{dv}{dt} = \frac{1}{2} \hbar u - \frac{1}{2} \hbar v. \end{cases} \quad (21.5.16)$$

Subtracting the two equations in (21.5.16) yields  $2 \frac{du}{dt} = \hbar v$ , i.e.,

$$\frac{du}{dt} = \frac{1}{2} \hbar v. \quad (21.5.17)$$

Adding the two equations in (21.5.16) yields  $2 \frac{dv}{dt} = \hbar u$ , i.e.,

$$\frac{dv}{dt} = \frac{1}{2} \hbar u. \quad (21.5.18)$$

From (21.5.17) and (21.5.18) we obtain

$$\hat{S}_z^p = \frac{1}{2} \hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \quad (21.5.19)$$

### (4) Summary

Combining the above calculations, we obtain the following three mapping relations:

$$\begin{aligned} \hat{S}_x^{e^-} &= \frac{1}{2} \hbar \left( \eta \frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial \eta} \right) \Rightarrow \hat{S}_x^p = -\frac{1}{2} \hbar \left( u \frac{\partial}{\partial v} - v \frac{\partial}{\partial u} \right), \\ \hat{S}_y^{e^-} &= \frac{i}{2} \hbar \left( \eta \frac{\partial}{\partial \xi} - \xi \frac{\partial}{\partial \eta} \right) \Rightarrow \hat{S}_y^p = \frac{i}{2} \hbar \left( v \frac{\partial}{\partial u} - u \frac{\partial}{\partial v} \right), \\ \hat{S}_z^{e^-} &= \frac{1}{2} \hbar \left( \xi \frac{\partial}{\partial \eta} - \eta \frac{\partial}{\partial \xi} \right) \Rightarrow \hat{S}_z^p = \frac{1}{2} \hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \end{aligned} \quad (21.5.20)$$

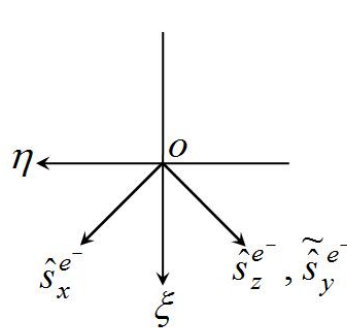
The computational results show that under the transformation (21.5.4), the three spin operators of the electron are transformed into the three spin operators of the proton. Consequently, the spin operator  $\hat{S}_z^{e^-}$  that defines the structure group  $G_{\hat{S}_3}$  of the electron is transformed into the spin

operator  $\hat{S}_z^p$  that defines the structure group  $G_{\hat{S}_1}$  of the proton. Therefore, under the transformation (21.5.4), an electron can transform into a proton.

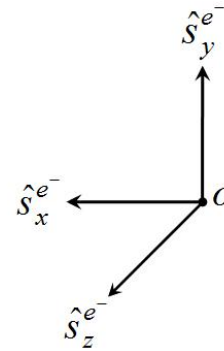
For visual clarity, we provide the following graphical explanation of this mapping process:

$\hat{S}_x^{e^-}$ ,  $\hat{S}_y^{e^-}$ ,  $\hat{S}_z^{e^-}$  are vector fields on the phase plane  $\xi\eta$ , but they can also be understood as vector fields in three-dimensional space. As shown in Figure 21.5.1, on the phase plane  $\xi\eta$ , the vector of the field  $\hat{S}_x^{e^-}$  at the point  $O(1,1)$  can be represented by an arrow  $\hat{S}_x^{e^-}$ ; the vector of the field  $\hat{S}_z^{e^-}$  at the point  $O(1,1)$  by an arrow  $\hat{S}_z^{e^-}$ ; and the vector of the field

$$\tilde{\hat{S}}_y^{e^-} = \frac{1}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right)$$



**Figure 21.5.1** Spin operators on the phase plane

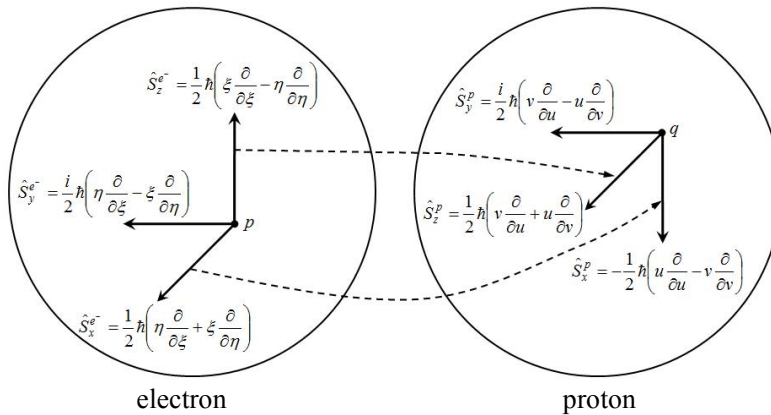


**Figure 21.5.2** Spin operators in  $R^3$

at the point  $O(1,1)$  by an arrow  $\tilde{\hat{S}}_y^{e^-}$ . The two arrows  $\hat{S}_z^{e^-}$  and  $\tilde{\hat{S}}_y^{e^-}$  coincide. Multiplying  $\tilde{\hat{S}}_y^{e^-}$  by the imaginary unit  $i$  gives

$$\hat{S}_y^{e^-} = i\tilde{\hat{S}}_y^{e^-} = \frac{i}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right).$$

Multiplying  $\tilde{\hat{S}}_y^{e^-}$  by the imaginary unit  $i$  is equivalent to rotating  $\tilde{\hat{S}}_y^{e^-}$  counterclockwise by  $90^\circ$ . Therefore, the vector of the field  $\hat{S}_y^{e^-}$  at the point  $O(1,1)$  can be drawn as a vector  $\hat{S}_y^{e^-}$  perpendicular to the phase plane  $\xi\eta$  at the point  $O(1,1)$ , as shown in Figure 21.5.2.



**Figure 21.5.3** Mapping from electron to proton

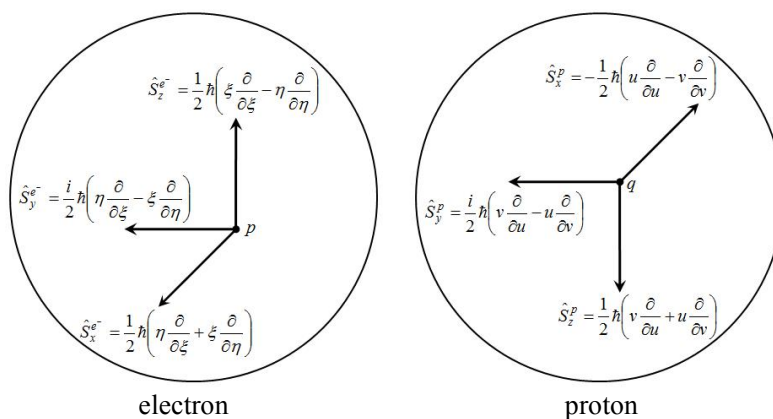
If we regard the phase plane  $\xi\eta$  and the plane  $uv$  as the same plane, then the vector  $\hat{S}_x^{e^-}$  is

perpendicular to  $\hat{S}_x^p$ , and the vector  $\hat{S}_z^{e-}$  is perpendicular to  $\hat{S}_z^p$ , as illustrated in Figure 21.5.3.

The electron's spin operator  $\hat{S}_x^{e-}$  transforms into the proton's spin operator  $\hat{S}_x^p$ . The electron's spin operator  $\hat{S}_z^{e-}$  determines the structure group  $G_{\hat{S}_3}$  of the electron's principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , while the proton's spin operator  $\hat{S}_z^p$  (note that the expression for  $\hat{S}_z^p$  here differs from the earlier one; the current operator  $\hat{S}_z^p$  corresponds to  $\hat{S}_x^p$  in previous chapters) determines the structure group  $G_{\hat{S}_1}$  of the proton's principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ .

The forms of the spin operators  $\hat{S}_y^{e-}$  and  $\hat{S}_y^p$  are identical, indicating that  $\hat{S}_y^{e-}$  remains formally unchanged under the transformation (21.5.4); that is, the vector field represented by  $\hat{S}_y^{e-}$  is invariant before and after the transformation. In contrast, the forms of the other two spin operators change, meaning the vector fields they represent are altered.

If we rotate both vector arrows representing the proton's spin operators  $\hat{S}_z^p$  and  $\hat{S}_x^p$  in Figure 21.5.3 by  $90^\circ$ , we obtain Figure 21.5.4. Comparing the electron and the proton in Figure 21.5.4, we find that the direction of the electron's spin operator  $\hat{S}_z^{e-}$  is exactly opposite to that of the proton's spin operator  $\hat{S}_z^p$ . These operators respectively determine the structure groups of the electron and the proton; hence, the rotation directions of the structure groups of the electron and the proton are exactly opposite, and consequently the directions of their integral curves are also opposite. For the electron, the rotation direction from the tangent vector  $\hat{S}_x^{e-}$  to  $\hat{S}_y^{e-}$  is clockwise, whereas for the proton, the rotation direction from  $\hat{S}_x^p$  to  $\hat{S}_y^p$  is counterclockwise. Therefore, the directions of the currents generated inside them are opposite, leading to opposite electric charges. These conclusions are consistent with those obtained earlier in Chapters 18.



**Figure 21.5.4** Opposite orientations of the structure groups of the electron and the proton

For the electron's spin operators  $\hat{S}_x^{e-}$ ,  $\hat{S}_y^{e-}$ , and  $\hat{S}_z^{e-}$ , their commutation relations are

$$[\hat{S}_x^{e-}, \hat{S}_y^{e-}] = i\hbar \hat{S}_z^{e-}, \quad [\hat{S}_y^{e-}, \hat{S}_z^{e-}] = i\hbar \hat{S}_x^{e-}, \quad [\hat{S}_z^{e-}, \hat{S}_x^{e-}] = i\hbar \hat{S}_y^{e-}.$$

But for the proton's spin operators  $\hat{S}_x^p$ ,  $\hat{S}_y^p$ , and  $\hat{S}_z^p$ , because  $\hat{S}_x^p$  and  $\hat{S}_z^{e-}$  differ by a minus sign in form, their commutation relations become

$$[\hat{S}_x^p, \hat{S}_y^p] = -i\hbar \hat{S}_z^p, \quad [\hat{S}_y^p, \hat{S}_z^p] = -i\hbar \hat{S}_x^p, \quad [\hat{S}_z^p, \hat{S}_x^p] = -i\hbar \hat{S}_y^p.$$

Obviously, the structure constants of the Lie groups for these two types of particles differ by a minus sign, yet the Lie algebras of the two particles are isomorphic. Because the structure constants differ by a minus sign, according to §12.2, among the two sets of operators—the

electron's spin operators  $\hat{S}_x^{e^-}, \hat{S}_y^{e^-}, \hat{S}_z^{e^-}$  and the proton's spin operators  $\hat{S}_x^p, \hat{S}_y^p, \hat{S}_z^p$ —one set must be **left-invariant tangent vector fields**, while the other set is **right-invariant tangent vector fields**. Through analysis below, a reasonable conclusion is: we regard the electron's spin operators  $\hat{S}_x^{e^-}, \hat{S}_y^{e^-}$ , and  $\hat{S}_z^{e^-}$  as right-invariant tangent vector fields, and the proton's spin operators  $\hat{S}_x^p, \hat{S}_y^p$ , and  $\hat{S}_z^p$  as left-invariant tangent vector fields.

According to our convention in classical physics, the counterclockwise (or right-handed) direction is defined as the positive direction, while the clockwise (or left-handed) direction is defined as the negative direction. Therefore, we may state: The proton carries a positive charge and generates a current in the positive direction when moving because the point-like particle inside the proton moves along the counter-clockwise (right-handed) direction, i.e., the point-like particle starts from the identity element and moves under the right translation of the group elements. Conversely, the electron carries a negative charge, and the direction of the current it produces when moving is opposite to its direction of motion because the point-like particle inside the electron moves along the clockwise (left-handed) direction, i.e., the point-like particle starts from the identity element and moves under the left translation of the group elements. The motion of a point-like particle from the identity element under left (or right) translation of the group elements also means that a tangent vector at the identity element, under left (or right) translation, generates a left-invariant (or right-invariant) tangent vector field.

For the electron, the motion of the point-like particle is driven by the one-parameter subgroup  $a_t^{e^-}$  determined by the spin operator  $\hat{S}_z^{e^-}$ . According to Theorem 12.5.9, any right-invariant vector field  $X$  on a Lie group  $G$  determines a one-parameter subgroup  $a_t^{e^-}$  of  $G$ , and the one-parameter transformation group  $\phi_t^{e^-}$  on  $G$  determined by the right-invariant vector field  $X$  is precisely the left translation determined by  $a_t^{e^-}$ . Hence, we may consider the electron's spin operators as right-invariant tangent vector fields. Since the three spin operators of the electron can serve as a basis for the electron's Lie algebra, all elements of the electron's Lie algebra are right-invariant tangent vector fields.

For the proton, the motion of the point-like particle is driven by the one-parameter subgroup  $a_t^p$  determined by the spin operator  $\hat{S}_z^p$ . According to Theorem 12.5.10, any left-invariant vector field  $X$  on a Lie group  $G$  determines a one-parameter subgroup  $a_t^p$  of  $G$ , and the one-parameter transformation group  $\phi_t^p$  on  $G$  determined by the left-invariant vector field  $X$  is precisely the right translation determined by  $a_t^p$ . Therefore, we may consider the proton's spin operators as left-invariant tangent vector fields. Since the three spin operators of the proton can serve as a basis for the proton's Lie algebra, all elements of the proton's Lie algebra are left-invariant tangent vector fields.

### 3.Implementation Method

Now let us explore the origin of equation (21.5.1).

Since a single electron spontaneously transforming into a proton violates the law of charge conservation, such a process cannot occur. However, if an electron collides with a positron, the following reaction can take place:

$$e^+ + e^- \rightarrow p + \bar{p},$$

which allows an electron to transform into a proton.

Assume that an electron  $e^-$  and a positron  $e^+$  interact. Their interaction can be represented by the direct product of two  $SU(2)$  groups:  $SU(2) \times SU(2)$ . This direct product can be viewed as a mapping process:

$$SU(2) \times SU(2) \rightarrow SU(2),$$

meaning that the left  $SU(2)$  group acts as a Lie transformation group on the right  $SU(2)$  group, or the right  $SU(2)$  group acts as a Lie transformation group on the left  $SU(2)$  group. In other words, one of the  $SU(2)$  groups is regarded as the transformation group. We take the left  $SU(2)$  group to

represent the positron (the transformation group) and the right  $SU(2)$  group to represent the electron (the smooth manifold being transformed).

According to equation (19.4.12), this positron  $SU(2)$  transformation group can be expressed in  $R^4$  as

$$GL(4, R^4)' = \exp[\hat{X}_4 s] \exp[\hat{X}_1 \theta + \hat{X}_2 \varphi + \hat{X}_3 \omega], \quad (21.5.21)$$

or in  $M^4$  as

$$GL(4, R^4)' = \exp[ic\hat{X}_4 t] \exp[\hat{X}_1 \theta + \hat{X}_2 \varphi + \hat{X}_3 \omega], \quad (21.5.22)$$

where

$$\hat{X}_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \hat{X}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \hat{X}_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \hat{X}_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (21.5.23)$$

The electron and positron have opposite orientations. Therefore, the electron's  $SU(2)$  group can be expressed in  $M^4$  as:

$$GL(4, R^4)' = \exp[-ic\hat{X}_4 t] \exp[\hat{X}_1 \theta + \hat{X}_2 \varphi + \hat{X}_3 \omega].$$

Since the tangent bundle of a Lie group is trivial, we can still denote the left-(or right-) invariant tangent vector field generated by the tangent vector  $\hat{X}_2$  at the identity element of the positron as  $\hat{X}_2$ . From the left-(or right-) invariant vector field  $\hat{X}_2$ , we can generate a one-parameter subgroup  $\exp[\hat{X}_2 \varphi]$ . Similarly, the one-parameter subgroup generated by  $\hat{X}_4$  is  $\exp[\hat{X}_4 s]$ . Because  $\hat{X}_4$  and  $\hat{X}_2$  commute, i.e.,

$$[\hat{X}_2, \hat{X}_4] = 0,$$

the product of these two one-parameter subgroups  $\exp[\hat{X}_2 \varphi]$  and  $\exp[\hat{X}_4 s]$ ,

$$G(\hat{X}_2, \hat{X}_4) = \exp[\hat{X}_4 s] \exp[\hat{X}_2 \varphi] \quad (21.5.24)$$

also forms a group.

Applying the subgroup  $G(\hat{X}_2, \hat{X}_4)$  from the left to the electron's  $SU(2)$  group gives

$$F: G(\hat{X}_2, \hat{X}_4) \times SU(2) \rightarrow SU(2). \quad (21.5.25)$$

In the local coordinate system at the identity element  $e \in SU(2)$ , equation (21.5.25) can be written as

$$p = \exp[\hat{X}_4 s] \exp[\hat{X}_2 \varphi] \cdot e = \exp[\hat{X}_4 s] \exp[\hat{X}_2 \varphi], \quad (21.5.26)$$

where the point  $p \in SU(2)$ .

Equation (21.5.26) is a coordinate transformation. According to the content of §5.1, the matrix of the tangent map and the cotangent map in the natural basis is the Jacobian matrix of equation (21.5.26). Therefore, we compute this mapping at the identity element  $e$ , i.e., at the origin ( $s = \theta = \varphi = \omega = 0$ ). The Jacobian matrix is

$$J = \left( \frac{\partial G(\hat{X}_2, \hat{X}_4)}{\partial s} \quad \frac{\partial G(\hat{X}_2, \hat{X}_4)}{\partial \varphi} \right)_e = \begin{pmatrix} \hat{X}_4 & \hat{X}_2 \end{pmatrix}_e, \quad (21.5.27)$$

that is

$$J = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}_e. \quad (21.5.28)$$

Our goal is to analyze the process of electron-to-proton conversion. Therefore, we regard the  $G(\hat{X}_2, \hat{X}_4)$  group on the left side of equation (21.5.25) as the Lie transformation group generated by the positron, and the  $SU(2)$  group on the right side as the electron's  $SU(2)$  group. The Jacobian matrix  $J$  is the transformation matrix for the components of tangent vectors. Assume the electron's spin operators are vector fields on the phase plane  $\xi\eta$  (which lies in the tangent space). The coordinates of this phase plane are taken as  $(\xi, \eta)$ , and the vector coordinates in this plane can also be represented as  $(\xi, \eta)$ . Further, assume the spin operators of the particle produced after the transformation (i.e., the proton) are vector fields on the phase plane  $uv$ . The vector coordinates in

this plane are taken as  $(u, v)$ , which can also represent vectors in this plane. Taking any two identical vectors  $(\xi, \eta)$  to form a four-dimensional vector  $(\xi, \eta, \xi, \eta)$  and multiplying by the matrix  $J$ , we obtain

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}_e \begin{pmatrix} \xi \\ \eta \\ \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \xi - \eta \\ \xi + \eta \end{pmatrix}, \quad (21.5.29)$$

which yields equation (21.5.1), i.e.,

$$\begin{cases} u = \xi - \eta, \\ v = \xi + \eta. \end{cases}$$

Thus, at the identity element  $e$  of the  $SU(2)$  group (i.e., at the origin), the transformation formula (21.5.1) exists. We have already computed earlier to obtain the transformation relations (21.5.20).

The determinant of the matrix of the mapping  $G(\hat{X}_2, \hat{X}_4)$  generated by the positron is

$$\begin{aligned} \text{del}(G(\hat{X}_2, \hat{X}_4)) &= \text{del}(\exp[\hat{X}_4 s] \exp[\hat{X}_2 \varphi]) \\ &= \text{del}(\exp[\hat{X}_4 s]) \text{del}(\exp[\hat{X}_2 \varphi]) = \exp[s(\text{tr} \hat{X}_4)] \exp[\varphi(\text{tr} \hat{X}_2)] > 0, \end{aligned}$$

which, according to Theorem 9.4.1, shows that the mapping  $G(\hat{X}_2, \hat{X}_4)$  generated by the positron is **orientation-preserving under pushforward**. Consequently, the positron maps the electron's orientation (left-handed orientation) onto the newly produced particle, endowing the new particle with a left-handed orientation as well. The newly produced particle is the proton, and the proton indeed possesses a left-handed orientation. If it is stipulated that the three spin operators of the electron, in the order  $\hat{S}_x^{e^-} \rightarrow \hat{S}_y^{e^-} \rightarrow \hat{S}_z^{e^-}$ , form a left-handed orientation, then the three spin operators of the proton, in the corresponding order  $\hat{S}_x^p \rightarrow \hat{S}_y^p \rightarrow \hat{S}_z^p$ , will also form a left-handed orientation, as shown in Figure 21.5.3.

Therefore, to transform an electron  $e^-$  into a proton  $p$ , we let this electron  $e^-$  collide with a positron  $e^+$ . This collision process is precisely the mutual transformation process of two  $SU(2)$  groups:  $SU(2) \times SU(2)$ . Because it is a mutual collision, the point-like particles of these two particles—their identity elements—become sufficiently close or even in complete contact. Hence, equation (21.5.26) can come into play, and consequently equation (21.5.1) can exist. As a result, the electron may transform into a proton  $p$ , and the positron may transform into an antiproton  $\bar{p}$ . That is, the following process can occur:

$$e^+ + e^- \rightarrow p + \bar{p}.$$

Furthermore, since the  $SU(2)$  group is a connected Lie group, according to Theorem 12.1.2 or Theorem 12.1.3, if the transformation occurs in a neighborhood of the identity element  $e$ , the entire Lie group will consequently transform. These are the reasons why the collision of an electron and a positron can relatively easily lead to their conversion into a proton-antiproton pair.

It is important to note that the electron's spin operators  $\hat{S}_x^{e^-}$ ,  $\hat{S}_y^{e^-}$ ,  $\hat{S}_z^{e^-}$  correspond one-to-one with the matrices  $\hat{X}_1$ ,  $\hat{X}_2$ ,  $\hat{X}_3$ , i.e.,  $\hat{S}_x^{e^-}$  corresponds to  $\hat{X}_1$ ,  $\hat{S}_y^{e^-}$  to  $\hat{X}_2$ , and  $\hat{S}_z^{e^-}$  to  $\hat{X}_3$ . In this electron-to-proton conversion process, only the spin operator  $\hat{S}_y^{e^-}$  (or  $\hat{X}_2$ ) remains unchanged, while the other two spin operators undergo transformation. Precisely because the spin operator  $\hat{S}_y^{e^-}$  (or  $\hat{X}_2$ ) remains invariant throughout this transformation, the transformation formula (21.5.1) persists from beginning to end, enabling the conversion process to be completed entirely.

#### 4. Elastic Collision

If we modify  $G(\hat{X}_2, \hat{X}_4)$  in equation (21.5.24) to



$$G(\hat{X}_2) = \exp[\hat{X}_2\varphi], \quad (21.5.30)$$

then equation (21.5.25) becomes

$$F : G(\hat{X}_2) \times SU(2) \rightarrow SU(2). \quad (21.5.31)$$

We now examine the effect of this transformation  $G(\hat{X}_2)$  on the spin operators  $\hat{S}_x^{e-}, \hat{S}_y^{e-}, \hat{S}_z^{e-}$  of the electron, which is an  $SU(2)$  group.

First, compute the Jacobian matrix of this mapping at the identity element  $e$ , i.e., at the origin ( $s = \theta = \varphi = \omega = 0$ ):

$$J = \left( \frac{\partial G(\hat{X}_2)}{\partial \varphi} \right)_e = \hat{X}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (21.5.32)$$

Assume the transformed spin operator of the electron is a vector field on the phase plane  $uv$ , where  $(u, v)$  represents a vector in this plane. Before the transformation, the electron's spin operators are vector fields on the phase plane  $\xi\eta$ , whose coordinates are taken as  $(\xi, \eta)$ . The vector coordinates in this plane can also be represented as  $(\xi, \eta)$ . Multiplying  $(\xi, \eta)$  by the matrix  $J$  gives

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad (21.5.33)$$

that is, we obtain

$$\begin{cases} u = -\eta, \\ v = \xi. \end{cases} \quad (21.5.34)$$

Or equivalently

$$\begin{cases} \xi = v, \\ \eta = -u. \end{cases} \quad (21.5.35)$$

Under this transformation (21.5.35), the spin operators

$$\hat{S}_x^{e-} = \frac{1}{2}\hbar \left( \eta \frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial \eta} \right), \quad \hat{S}_y^{e-} = \frac{i}{2}\hbar \left( \eta \frac{\partial}{\partial \xi} - \xi \frac{\partial}{\partial \eta} \right), \quad \hat{S}_z^{e-} = \frac{1}{2}\hbar \left( \xi \frac{\partial}{\partial \xi} - \eta \frac{\partial}{\partial \eta} \right)$$

transform respectively into

$$\begin{aligned} \hat{S}_x'^{e-} &= \tilde{F}(\hat{S}_x^{e-}) = \frac{1}{2}\hbar \left( -u \frac{\partial}{\partial v} + v \frac{\partial}{\partial(-u)} \right) = -\frac{1}{2}\hbar \left( u \frac{\partial}{\partial v} + v \frac{\partial}{\partial u} \right), \\ \hat{S}_y'^{e-} &= \tilde{F}(\hat{S}_y^{e-}) = \frac{i}{2}\hbar \left( -u \frac{\partial}{\partial v} - v \frac{\partial}{\partial(-u)} \right) = -\frac{i}{2}\hbar \left( u \frac{\partial}{\partial v} - v \frac{\partial}{\partial u} \right), \\ \hat{S}_z'^{e-} &= \tilde{F}(\hat{S}_z^{e-}) = \frac{1}{2}\hbar \left( v \frac{\partial}{\partial v} - (-u) \frac{\partial}{\partial(-u)} \right) = \frac{1}{2}\hbar \left( v \frac{\partial}{\partial v} - u \frac{\partial}{\partial u} \right). \end{aligned}$$

The commutation relations of the spin operators  $\hat{S}_x'^{e-}, \hat{S}_y'^{e-}, \hat{S}_z'^{e-}$  are

$$[\hat{S}_x'^{e-}, \hat{S}_y'^{e-}] = i\hbar \hat{S}_z'^{e-}, \quad [\hat{S}_y'^{e-}, \hat{S}_z'^{e-}] = i\hbar \hat{S}_x'^{e-}, \quad [\hat{S}_z'^{e-}, \hat{S}_x'^{e-}] = i\hbar \hat{S}_y'^{e-}.$$

After the transformation, the commutation relations of  $\hat{S}_x'^{e-}, \hat{S}_y'^{e-}, \hat{S}_z'^{e-}$  remain identical to those of  $\hat{S}_x^{e-}, \hat{S}_y^{e-}, \hat{S}_z^{e-}$ . Therefore, if the original spin operators were left-invariant tangent vector fields, the transformed ones are still left-invariant; if they were right-invariant tangent vector fields, they remain right-invariant. Moreover, the spin operators before and after the transformation are either parallel or antiparallel. Consequently, the mapping (21.5.31) can describe particle interaction processes such as the elastic collision of two electrons,

$$e^- + e^- \rightarrow e^- + e^-,$$

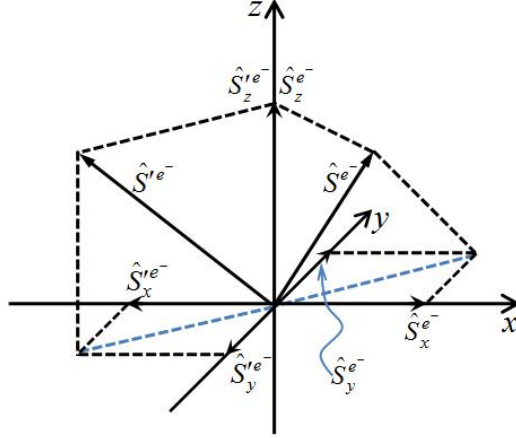
where no transformation into other particles occurs.

Figure 21.5.5 helps us understand why  $\hat{S}_x^{e-}$  and  $\hat{S}_x'^{e-}$ , as well as  $\hat{S}_y^{e-}$  and  $\hat{S}_y'^{e-}$ , have

opposite signs. The operators  $\hat{S}^{e-} = (\hat{S}_x^{e-}, \hat{S}_y^{e-}, \hat{S}_z^{e-})$  and  $\hat{S}'^{e-} = (\hat{S}_x'^{e-}, \hat{S}_y'^{e-}, \hat{S}_z'^{e-})$  belong to different Lie algebras and different smooth manifolds (or Lie groups). Hence, with reference to equation (11.10.1), we can stipulate that their inner product is zero, i.e.,

$$\hat{S}^{e-} \cdot \hat{S}'^{e-} = 0.$$

Thus, if  $\hat{S}^{e-}$  and  $\hat{S}'^{e-}$  are represented in the same coordinate system, they can be regarded as mutually perpendicular vectors, as illustrated in Figure 21.5.5.



**Figure 21.5.5** Reason why  $\hat{S}_x^{e-}$  and  $\hat{S}_x'^{e-}$ ,  $\hat{S}_y^{e-}$  and  $\hat{S}_y'^{e-}$  have opposite signs

## §21.6 Internal Mechanism of Electron-to-Neutrino Transformation

In this section we study the internal mechanism by which an electron transforms into a neutrino in the reaction

$$e^+ + e^- \rightarrow \nu + \bar{\nu}.$$

The electron is an  $SU(2)$  group. On the phase plane  $\xi\eta$ , the three spin operators of the electron can be expressed as

$$\hat{S}_x^{e-} = \frac{1}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} + \xi\frac{\partial}{\partial\eta}\right), \quad \hat{S}_y^{e-} = \frac{i}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right), \quad \hat{S}_z^{e-} = \frac{1}{2}\hbar\left(\xi\frac{\partial}{\partial\xi} - \eta\frac{\partial}{\partial\eta}\right).$$

The neutrino is also an  $SU(2)$  group. Let its three spin operators be  $\hat{S}_x^\nu, \hat{S}_y^\nu, \hat{S}_z^\nu$ . We represent the neutrino's spin operators as vector fields on the same phase plane  $uv$ . The electron is a principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$  whose structure group  $G_{\hat{S}_3}$  is determined by the spin operator  $\hat{S}_z^{e-}$ . The neutrino is a principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$  whose structure group  $G_{\hat{S}_2}$  is determined by the spin operator  $\hat{S}_y^\nu$ . If the spin operator  $\hat{S}_z^{e-}$  that defines the structure group of the electron's principal bundle transforms into the spin operator  $\hat{S}_y^\nu$ , then the electron transforms into a neutrino.

### 1. Transformation Formula

Assume that an electron and a positron interact. Their interaction can be represented by the direct product of two  $SU(2)$  groups:  $SU(2) \times SU(2)$ . This direct product can be viewed as a mapping process

$$SU(2) \times SU(2) \rightarrow SU(2).$$

We may regard one of the  $SU(2)$  groups in  $SU(2) \times SU(2)$  as the Lie transformation group.

The matrix corresponding to  $\hat{S}_x^{e-}$  is the one given in equation (21.5.23):

$$\hat{X}_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

The one-parameter subgroup generated by  $\hat{X}_1$  is  $\exp[\hat{X}_1\theta]$ . Because  $\hat{X}_4$  commutes with  $\hat{X}_1$ , i.e.,

$$[\hat{X}_1, \hat{X}_4] = 0,$$

the product of the one-parameter subgroup  $\exp[\hat{X}_1\theta]$  and the one-parameter subgroup  $\exp[\hat{X}_4s]$  generated by  $\hat{X}_4$ ,

$$G(\hat{X}_1, \hat{X}_4) = \exp[\hat{X}_4s]\exp[\hat{X}_1\theta] \quad (21.6.1)$$

is also a group.

The subgroup  $G(\hat{X}_1, \hat{X}_4)$  can be regarded either as a subgroup of the left  $SU(2)$  factor in  $SU(2) \times SU(2)$  or as a subgroup of the right  $SU(2)$  factor. In what follows we take  $G(\hat{X}_1, \hat{X}_4)$  as a subgroup of the left  $SU(2)$ :

$$F : G(\hat{X}_1, \hat{X}_4) \times SU(2) \rightarrow SU(2). \quad (21.6.2)$$

According to Theorems 12.5.9 and 12.5.10, the group  $G(\hat{X}_1, \hat{X}_4)$  can act as a transformation group, producing a mapping effect on a smooth manifold. Following §5.1, the matrix of the tangent map and the cotangent map in the natural basis is the Jacobian matrix of the mapping  $G(\hat{X}_1, \hat{X}_4)$ . Therefore, we compute the Jacobian matrix of this mapping at the identity element  $e$ , i.e., at the origin ( $s = \theta = \varphi = \omega = 0$ ). The Jacobian matrix is

$$J = \left( \frac{\partial G(\hat{X}_1, \hat{X}_4)}{\partial s} \quad \frac{\partial G(\hat{X}_1, \hat{X}_4)}{\partial \theta} \right)_e = \begin{pmatrix} \hat{X}_4 & \hat{X}_1 \end{pmatrix}_e, \quad (21.6.3)$$

that is,

$$J = \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \end{pmatrix}_e. \quad (21.6.4)$$

Our purpose is to analyze the process of electron-to-neutrino conversion. Hence we regard the  $G(\hat{X}_1, \hat{X}_4)$  group on the left side of equation (21.6.2) as the Lie transformation group generated by the positron, and the  $SU(2)$  group on the right side as the electron's  $SU(2)$  group. Let the electron's spin operators be vector fields on the phase plane  $\xi\eta$  (lying in the tangent space). The coordinates of this phase plane are taken as  $(\xi, \eta)$ , and the vector coordinates in this plane can also be represented as  $(\xi, \eta)$ . Suppose the particle produced after the transformation (i.e., the neutrino) has its spin operators as vector fields on the phase plane  $uv$ , and let  $(u, v)$  denote a vector in this plane. Taking any two identical vectors  $(\xi, \eta)$  to form a four-dimensional vector  $(\xi, \eta, \xi, \eta)$  and multiplying by the matrix  $J$  yields

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \end{pmatrix}_e \begin{pmatrix} \xi \\ \eta \\ \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \xi + i\eta \\ i\xi + \eta \end{pmatrix}, \quad (21.6.5)$$

that is,

$$\begin{cases} u = \xi + i\eta, \\ v = i\xi + \eta. \end{cases} \quad (21.6.6)$$

The neutrino is also an  $SU(2)$  group; its three spin operators  $\hat{S}_x^\nu, \hat{S}_y^\nu, \hat{S}_z^\nu$  are vector fields on the phase plane  $uv$ . Assuming that an electron transforms into a neutrino, we now derive how the three spin operators  $\hat{S}_x^{e-}, \hat{S}_y^{e-}, \hat{S}_z^{e-}$  of the electron are transformed into the spin operators of the neutrino.

The transformation formula that converts the electron's spin operators into the neutrino's spin

operators is equation (21.6.6). Its Jacobian matrix is

$$J = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \quad (21.6.7)$$

The transpose of matrix  $J$  is itself, i.e.,  $J^T = J$ . Thus, we obtain

$$\begin{pmatrix} f_* \frac{\partial}{\partial \xi} \\ f_* \frac{\partial}{\partial \eta} \end{pmatrix} = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix},$$

i.e.,

$$f_* \frac{\partial}{\partial \xi} = \frac{\partial}{\partial u} + i \frac{\partial}{\partial v}, \quad f_* \frac{\partial}{\partial \eta} = i \frac{\partial}{\partial u} + \frac{\partial}{\partial v}. \quad (21.6.8)$$

## 2. Transformation of the Electron's Spin Operators

### (1) Transformation of the electron's spin operator $\hat{S}_x^{e^-}$

We transform the electron's spin operator  $\hat{S}_x^{e^-}$  using equation (21.6.8). From the expression of  $\hat{S}_x^{e^-}$  we obtain

$$\begin{aligned} \hat{S}_x^v &= \frac{1}{2} \hbar \left( \eta \left( f_* \frac{\partial}{\partial \xi} \right) + \xi \left( f_* \frac{\partial}{\partial \eta} \right) \right) = \frac{1}{2} \hbar \eta \left( \frac{\partial}{\partial u} + i \frac{\partial}{\partial v} \right) + \frac{1}{2} \hbar \xi \left( i \frac{\partial}{\partial u} + \frac{\partial}{\partial v} \right) \\ &= \left( \frac{1}{2} \hbar \eta + \frac{1}{2} \hbar \xi i \right) \frac{\partial}{\partial u} + \left( \frac{1}{2} \hbar \eta i + \frac{1}{2} \hbar \xi \right) \frac{\partial}{\partial v} \\ &= \frac{1}{2} \hbar (\eta + \xi i) \frac{\partial}{\partial u} + \frac{1}{2} \hbar (\eta i + \xi) \frac{\partial}{\partial v} = \frac{1}{2} \hbar v \frac{\partial}{\partial u} + \frac{1}{2} \hbar u \frac{\partial}{\partial v}. \end{aligned}$$

That is,

$$\hat{S}_x^v = \frac{1}{2} \hbar \left( v \frac{\partial}{\partial u} + u \frac{\partial}{\partial v} \right). \quad (21.6.9)$$

### (2) Transformation of the electron's spin operator $\hat{S}_y^{e^-}$

Next we transform the electron's spin operator  $\hat{S}_y^{e^-}$ . From its expression we obtain

$$\begin{aligned} \hat{S}_y^v &= \frac{i}{2} \hbar \left( \eta \left( f_* \frac{\partial}{\partial \xi} \right) - \xi \left( f_* \frac{\partial}{\partial \eta} \right) \right) = \frac{i}{2} \hbar \eta \left( \frac{\partial}{\partial u} + i \frac{\partial}{\partial v} \right) - \frac{i}{2} \hbar \xi \left( i \frac{\partial}{\partial u} + \frac{\partial}{\partial v} \right) \\ &= \left( \frac{i}{2} \hbar \eta - \frac{i}{2} \hbar \xi i \right) \frac{\partial}{\partial u} + \left( \frac{i}{2} \hbar \eta i - \frac{i}{2} \hbar \xi \right) \frac{\partial}{\partial v} \\ &= \frac{1}{2} \hbar (i \eta + \xi) \frac{\partial}{\partial u} - \frac{1}{2} \hbar (\eta + i \xi) \frac{\partial}{\partial v} \\ &= \frac{1}{2} \hbar u \frac{\partial}{\partial u} - \frac{1}{2} \hbar v \frac{\partial}{\partial v}. \end{aligned}$$

That is,

$$\hat{S}_y^v = \frac{1}{2} \hbar \left( u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v} \right). \quad (21.6.10)$$

### (3) Transformation of the electron's spin operator $\hat{S}_z^{e^-}$

Finally, we transform the electron's spin operator  $\hat{S}_z^{e^-}$ . From its expression we obtain

$$\begin{aligned} \hat{S}_z^v &= \frac{1}{2} \hbar \left( \xi \left( f_* \frac{\partial}{\partial \xi} \right) - \eta \left( f_* \frac{\partial}{\partial \eta} \right) \right) = \frac{1}{2} \hbar \xi \left( \frac{\partial}{\partial u} + i \frac{\partial}{\partial v} \right) - \frac{1}{2} \hbar \eta \left( i \frac{\partial}{\partial u} + \frac{\partial}{\partial v} \right) \\ &= \left( \frac{1}{2} \hbar \xi - \frac{1}{2} \hbar \eta i \right) \frac{\partial}{\partial u} + \left( \frac{1}{2} \hbar \xi i - \frac{1}{2} \hbar \eta \right) \frac{\partial}{\partial v} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}\hbar(\xi - \eta i)\frac{\partial}{\partial u} + \frac{1}{2}\hbar(\xi i - \eta)\frac{\partial}{\partial v} \\
&= \frac{i}{2}\hbar(-\xi i - \eta)\frac{\partial}{\partial u} + \frac{i}{2}\hbar(\xi + \eta i)\frac{\partial}{\partial v} \\
&= -\frac{i}{2}\hbar v\frac{\partial}{\partial u} + \frac{i}{2}\hbar u\frac{\partial}{\partial v}.
\end{aligned}$$

That is,

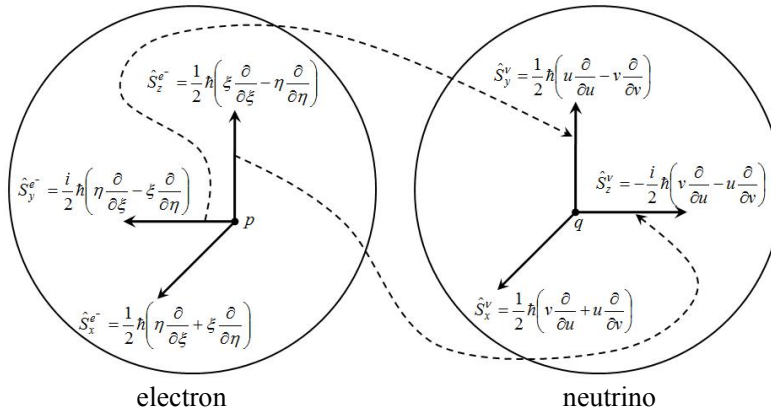
$$\hat{S}_z^\nu = -\frac{i}{2}\hbar\left(v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}\right). \quad (21.6.11)$$

#### (4)Summary

Combining the above calculations, we obtain the following three mapping relations:

$$\begin{aligned}
\hat{S}_x^{e^-} &= \frac{1}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} + \xi\frac{\partial}{\partial\eta}\right) \Rightarrow \hat{S}_x^\nu = \frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right), \\
\hat{S}_y^{e^-} &= \frac{i}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right) \Rightarrow \hat{S}_y^\nu = \frac{1}{2}\hbar\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right), \\
\hat{S}_z^{e^-} &= \frac{1}{2}\hbar\left(\xi\frac{\partial}{\partial\xi} - \eta\frac{\partial}{\partial\eta}\right) \Rightarrow \hat{S}_z^\nu = -\frac{i}{2}\hbar\left(v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}\right).
\end{aligned}$$

The computational results show that under the mapping  $F$ , the three spin operators of the electron are transformed into the three spin operators of the neutrino. Consequently, the spin operator  $\hat{S}_z^{e^-}$  that defines the structure group  $G_{\hat{S}_3}$  of the electron is transformed into the spin operator  $\hat{S}_z^\nu$  that defines the structure group  $G_{\hat{S}_2}$  of the neutrino. Therefore, under the mapping  $F$ , an electron can transform into a neutrino. The electron's spin operator  $\hat{S}_x^{e^-}$  remains unchanged under the mapping  $F$ , while  $\hat{S}_y^{e^-}$  and  $\hat{S}_z^{e^-}$  both change. We illustrate the mapping process from the electron's spin operators to the neutrino's spin operators in Figure 21.6.1.



**Figure 21.6.1** Mapping from electron to neutrino

The determinant of the matrix of the mapping  $G(\hat{X}_1, \hat{X}_4)$  generated by the positron is

$$\text{del}(G(\hat{X}_1, \hat{X}_4)) = \exp[s(\text{tr}\hat{X}_4)]\exp[\theta(\text{tr}\hat{X}_1)] > 0,$$

According to Theorem 9.4.1, the mapping  $G(\hat{X}_1, \hat{X}_4)$  is **orientation-preserving under pushforward**. Hence the positron maps the electron's orientation (left-handed orientation) onto the newly produced particle, endowing the new particle with a left-handed orientation as well. The newly produced particle is the neutrino, and the neutrino indeed possesses a left-handed orientation. If we stipulate that the three spin operators of the electron, in the order  $\hat{S}_x^{e^-} \rightarrow \hat{S}_y^{e^-} \rightarrow \hat{S}_z^{e^-}$ , form a left-handed orientation, then the three spin operators of the neutrino,

in the corresponding order  $\hat{S}_x^\nu \rightarrow \hat{S}_y^\nu \rightarrow \hat{S}_z^\nu$ , also form a left-handed orientation, as shown in Figure 21.6.1.

## §21.7 Internal Mechanism of Proton-to-Neutrino Transformation

This section studies the internal mechanism by which a proton transforms into a neutrino in the reaction

$$\bar{p} + p \rightarrow \nu + \bar{\nu}.$$

The proton is an  $SU(2)$  group. On the phase plane  $\xi\eta$ , the three spin operators of the proton can be expressed as

$$\hat{S}_x^p = \frac{1}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} + \xi\frac{\partial}{\partial\eta}\right), \quad \hat{S}_y^p = \frac{i}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right), \quad \hat{S}_z^p = \frac{1}{2}\hbar\left(\xi\frac{\partial}{\partial\xi} - \eta\frac{\partial}{\partial\eta}\right).$$

The corresponding spin operators in matrix form are

$$\hat{S}_1 = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_2 = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{S}_3 = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

respectively (see Equation (16.11.2)).

The neutrino is also an  $SU(2)$  group. Let its three spin operators be  $\hat{S}_x^\nu, \hat{S}_y^\nu, \hat{S}_z^\nu$ . We represent the neutrino's spin operators as vector fields on the same phase plane  $uv$ . The proton is a principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_1})$  whose structure group  $G_{\hat{S}_1}$  is determined by the spin operator  $\hat{S}_x^p$ . The neutrino is a principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$  whose structure group  $G_{\hat{S}_2}$  is determined by the spin operator  $\hat{S}_y^\nu$ . If the spin operator  $\hat{S}_x^p$  that defines the structure group of the proton's principal bundle transforms into the spin operator  $\hat{S}_y^\nu$ , then the proton transforms into a neutrino.

### 1. Transformation Formula

Assume a proton  $p$  and an antiproton  $\bar{p}$  interact. Their interaction can be represented by the direct product of two  $SU(2)$  groups:  $SU(2) \times SU(2)$ . This direct product can be viewed as a mapping process

$$SU(2) \times SU(2) \rightarrow SU(2).$$

We may regard one of the  $SU(2)$  groups in  $SU(2) \times SU(2)$  as the Lie transformation group.

The matrix corresponding to  $\hat{S}_z^p$  is the one given in equation (21.5.23):

$$\hat{X}_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

The one-parameter subgroup generated by  $\hat{X}_3$  is  $\exp[\hat{X}_3\omega]$ . Because  $\hat{X}_4$  commutes with  $\hat{X}_3$ , i.e.,

$$[\hat{X}_3, \hat{X}_4] = 0,$$

the product of the one-parameter subgroup  $\exp[\hat{X}_3\omega]$  and the one-parameter subgroup  $\exp[\hat{X}_4s]$  generated by  $\hat{X}_4$ ,

$$G(\hat{X}_3, \hat{X}_4) = \exp[\hat{X}_4s]\exp[\hat{X}_3\omega] \quad (21.7.1)$$

is also a group.

The subgroup  $G(\hat{X}_3, \hat{X}_4)$  can be regarded either as a subgroup of the left  $SU(2)$  factor in  $SU(2) \times SU(2)$  or as a subgroup of the right  $SU(2)$  factor. In what follows we take  $G(\hat{X}_3, \hat{X}_4)$  as a subgroup of the left  $SU(2)$ :

$$F : G(\hat{X}_1, \hat{X}_4) \times SU(2) \rightarrow SU(2). \quad (21.7.2)$$

According to Theorems 12.5.9 and 12.5.10, the group  $G(\hat{X}_3, \hat{X}_4)$  can act as a transformation group, producing a mapping effect on a smooth manifold. Following §5.1, the matrix of the tangent map and the cotangent map in the natural basis is the Jacobian matrix of the mapping  $G(\hat{X}_3, \hat{X}_4)$ . Therefore, we compute the Jacobian matrix of this mapping at the identity element  $e$ , i.e., at the origin ( $s = \theta = \varphi = \omega = 0$ ). The Jacobian matrix is

$$J = \left( \frac{\partial G(\hat{X}_3, \hat{X}_4)}{\partial s} \quad \frac{\partial G(\hat{X}_3, \hat{X}_4)}{\partial \omega} \right)_e = \begin{pmatrix} \hat{X}_4 & \hat{X}_3 \end{pmatrix}_e, \quad (21.7.3)$$

that is,

$$J = \begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 1 & 0 & -i \end{pmatrix}_e. \quad (21.7.4)$$

Our purpose is to analyze the proton-to-neutrino conversion process. Hence we regard the  $G(\hat{X}_3, \hat{X}_4)$  group on the left side of equation (21.7.2) as the Lie transformation group generated by the antiproton, and the  $SU(2)$  group on the right side as the proton's  $SU(2)$  group. Let the proton's spin operators be vector fields on the phase plane  $\xi\eta$  (lying in the tangent space). The coordinates of this phase plane are taken as  $(\xi, \eta)$ , and the vector coordinates in this plane can also be represented as  $(\xi, \eta)$ . Suppose the particle produced after the transformation (i.e., the neutrino) has its spin operators as vector fields on the phase plane  $uv$ , and let  $(u, v)$  denote a vector in this plane. Taking any two identical vectors  $(\xi, \eta)$  to form a four-dimensional vector  $(\xi, \eta, \xi, \eta)$  and multiplying by the matrix  $J$  yields

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 1 & 0 & -i \end{pmatrix}_e \begin{pmatrix} \xi \\ \eta \\ \xi \\ \eta \end{pmatrix} = \begin{pmatrix} (1+i)\xi \\ (1-i)\eta \end{pmatrix}, \quad (21.7.5)$$

that is,

$$\begin{cases} u = (1+i)\xi, \\ v = (1-i)\eta. \end{cases} \quad (21.7.6)$$

The neutrino is also an  $SU(2)$  group; its three spin operators  $\hat{S}_x^\nu, \hat{S}_y^\nu, \hat{S}_z^\nu$  are vector fields on the phase plane  $uv$ . Assuming that a proton transforms into a neutrino, we now derive how the three spin operators  $\hat{S}_x^p, \hat{S}_y^p, \hat{S}_z^p$  of the proton are transformed into the spin operators of the neutrino.

The transformation formula that converts the proton's spin operators into the neutrino's spin operators is equation (21.7.6). Its Jacobian matrix is

$$J = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}. \quad (21.7.7)$$

The transpose of matrix  $J$  is itself, i.e.,  $J^T = J$ . Thus, we obtain

$$\begin{pmatrix} f_* \frac{\partial}{\partial \xi} \\ f_* \frac{\partial}{\partial \eta} \end{pmatrix} = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix},$$

i.e.,

$$f_* \frac{\partial}{\partial \xi} = (1+i) \frac{\partial}{\partial u}, \quad f_* \frac{\partial}{\partial \eta} = (1-i) \frac{\partial}{\partial v}. \quad (21.7.8)$$

## 2. Transformation of the Proton's Spin Operators

### (1) Transformation of the proton's spin operator $\hat{S}_x^p$

We transform the proton's spin operator  $\hat{S}_x^p$  using equation (21.7.8). From the expression of  $\hat{S}_x^p$  we obtain

$$\begin{aligned}\hat{S}_x^p &= \frac{1}{2}\hbar\left(\eta(f_*\frac{\partial}{\partial\xi}) + \xi(f_*\frac{\partial}{\partial\eta})\right) = \frac{1}{2}\hbar\left(\eta(1+i)\frac{\partial}{\partial u} + \xi(1-i)\frac{\partial}{\partial v}\right) \\ &= \frac{1}{2}\hbar\left(\frac{1+i}{1-i}v\frac{\partial}{\partial u} + \frac{1-i}{1+i}u\frac{\partial}{\partial v}\right) = \frac{1}{2}\hbar\left(\frac{(1+i)^2}{(1-i)(1+i)}v\frac{\partial}{\partial u} + \frac{(1-i)^2}{(1+i)(1-i)}u\frac{\partial}{\partial v}\right) \\ &= \frac{1}{2}\hbar\left(\frac{2i}{2}v\frac{\partial}{\partial u} + \frac{-2i}{2}u\frac{\partial}{\partial v}\right) = \frac{i}{2}\hbar\left(v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}\right).\end{aligned}$$

That is,

$$\hat{S}_x^p = \frac{i}{2}\hbar\left(v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}\right). \quad (21.7.9)$$

### (2) Transformation of the proton's spin operator $\hat{S}_y^p$

Next we transform the proton's spin operator  $\hat{S}_y^p$ . From its expression we obtain

$$\begin{aligned}\hat{S}_y^p &= \frac{i}{2}\hbar\left(\eta(f_*\frac{\partial}{\partial\xi}) - \xi(f_*\frac{\partial}{\partial\eta})\right) = \frac{i}{2}\hbar\left(\eta(1+i)\frac{\partial}{\partial u} - \xi(1-i)\frac{\partial}{\partial v}\right) \\ &= \frac{i}{2}\hbar\left(\frac{1+i}{1-i}v\frac{\partial}{\partial u} - \frac{1-i}{1+i}u\frac{\partial}{\partial v}\right) = \frac{i}{2}\hbar\left(iv\frac{\partial}{\partial u} + iu\frac{\partial}{\partial v}\right) \\ &= -\frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right).\end{aligned}$$

That is,

$$\hat{S}_y^p = -\frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right). \quad (21.7.10)$$

### (3) Transformation of the proton's spin operator $\hat{S}_z^p$

Finally, we transform the proton's spin operator  $\hat{S}_z^p$ . From its expression we obtain

$$\begin{aligned}\hat{S}_z^p &= \frac{1}{2}\hbar\left(\xi(f_*\frac{\partial}{\partial\xi}) - \eta(f_*\frac{\partial}{\partial\eta})\right) = \frac{1}{2}\hbar\left(\xi(1+i)\frac{\partial}{\partial u} - \eta(1-i)\frac{\partial}{\partial v}\right) \\ &= \frac{1}{2}\hbar\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right).\end{aligned}$$

That is,

$$\hat{S}_z^p = \frac{1}{2}\hbar\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right). \quad (21.7.11)$$

### (4) Summary

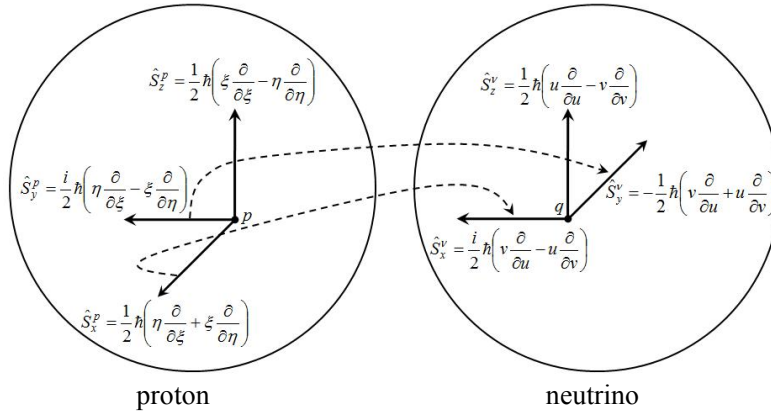
Combining the above calculations, we obtain the following three mapping relations:

$$\begin{aligned}\hat{S}_x^p &= \frac{1}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} + \xi\frac{\partial}{\partial\eta}\right) \Rightarrow \hat{S}_x^p = \frac{i}{2}\hbar\left(v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}\right), \\ \hat{S}_y^p &= \frac{i}{2}\hbar\left(\eta\frac{\partial}{\partial\xi} - \xi\frac{\partial}{\partial\eta}\right) \Rightarrow \hat{S}_y^p = -\frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right), \\ \hat{S}_z^p &= \frac{1}{2}\hbar\left(\xi\frac{\partial}{\partial\xi} - \eta\frac{\partial}{\partial\eta}\right) \Rightarrow \hat{S}_z^p = \frac{1}{2}\hbar\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right).\end{aligned}$$

The computational results show that under the mapping  $F$ , the three spin operators of the proton



are transformed into the three spin operators of the neutrino. Consequently, the spin operator  $\hat{S}_x^p$  that defines the structure group  $G_{\hat{S}_1}$  of the proton is transformed into the spin operator  $\hat{S}_x^\nu$  that defines the structure group  $G_{\hat{S}_2}$  of the neutrino. Therefore, under the mapping  $F$ , a proton can transform into a neutrino. The proton's spin operator  $\hat{S}_z^p$  remains unchanged under the mapping  $F$ , while  $\hat{S}_x^p$  and  $\hat{S}_y^p$  both change. We illustrate the mapping process from the proton's spin operators to the neutrino's spin operators in Figure 21.7.1.



**Figure 21.7.1** Mapping from proton to neutrino

The determinant of the matrix of the mapping  $G(\hat{X}_3, \hat{X}_4)$  generated by the antiproton is

$$\det(G(\hat{X}_3, \hat{X}_4)) = \exp[s(\text{tr}\hat{X}_4)] \exp[\omega(\text{tr}\hat{X}_3)] > 0,$$

According to Theorem 9.4.1, the mapping  $G(\hat{X}_3, \hat{X}_4)$  is **orientation-preserving under pushforward**. Hence the antiproton maps the proton's orientation (left-handed orientation) onto the newly produced particle, endowing the new particle with a left-handed orientation as well. The newly produced particle is the neutrino, and the neutrino indeed possesses a left-handed orientation. If we stipulate that the three spin operators of the proton, in the order  $\hat{S}_x^p \rightarrow \hat{S}_y^p \rightarrow \hat{S}_z^p$ , form a left-handed orientation, then the three spin operators of the neutrino, in the corresponding order  $\hat{S}_x^\nu \rightarrow \hat{S}_y^\nu \rightarrow \hat{S}_z^\nu$ , also form a left-handed orientation, as shown in Figure 21.7.1.

From the definition of one-parameter subgroups, it follows that the transformation of the  $SU(2)$  group given by

$$F_1 : \exp[\hat{X}_3\omega] \times SU(2) \rightarrow SU(2)$$

is a diffeomorphism. Since  $\hat{X}_4$  is the identity matrix, the following transformation is also a diffeomorphism:

$$F_2 : \exp[\hat{X}_4s] \times ([\hat{X}_3\omega] \times SU(2)) \rightarrow SU(2) :$$

Thus, the transformation

$$F : \exp[\hat{X}_4s] \exp[\hat{X}_3\omega] \times SU(2) \rightarrow SU(2)$$

i.e.,

$$F : G(\hat{X}_3, \hat{X}_4) \times SU(2) \rightarrow SU(2)$$

is also a diffeomorphism. Therefore, by Theorem 6.4.2, the tangent map  $F_*$  establishes an isomorphism between the Lie algebra of smooth tangent vector fields on the proton and the Lie algebra of the neutrino. Furthermore, since both the proton and the neutrino are spheres  $S^3$ , which are simply connected, it follows from Theorem 12.5.2 that the mapping  $F$  is an isomorphic mapping.

Because the spin operator

$$\hat{S}_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

commutes with the matrix

$$\hat{X}_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

the one-parameter subgroup  $\exp[\hat{S}_3\lambda]$  commutes with the one-parameter subgroup  $\exp[\hat{X}_3\omega]$ , i.e.,

$$\exp[\hat{S}_3\lambda] \exp[\hat{X}_3\omega] = \exp[\hat{X}_3\omega] \exp[\hat{S}_3\lambda].$$

Moreover, since the spin operator  $\hat{S}_3$  also commutes with the identity matrix  $\hat{X}_4$ , the one-parameter subgroup  $\exp[\hat{S}_3\lambda]$  commutes with the one-parameter subgroup  $\exp[\hat{X}_4s]$  as well. Hence, the one-parameter subgroup  $\exp[\hat{S}_3\lambda]$  commutes with the transformation group  $G(\hat{X}_3, \hat{X}_4) = \exp[\hat{X}_4s] \exp[\hat{X}_3\omega]$ , and consequently, by Theorem 6.6.3, the spin operator  $\hat{S}_3$ , i.e.,  $\hat{S}_z^P$ , is invariant under the diffeomorphism  $G(\hat{X}_3, \hat{X}_4)$ .

## §21.8 Internal Mechanism of Photon Transformation into Other elementary

### Particles

According to Theorem 12.5.3, a Lie group determines its Lie algebra, not vice versa. Two Lie algebras being isomorphic does not imply that the corresponding Lie groups are isomorphic; conversely, if two Lie groups are isomorphic, their Lie algebras must be isomorphic. Therefore, to study the mutual transformation of particles, one must first identify the Lie groups corresponding to the particles before and after the transformation, and then determine their Lie algebras. The mutual transformations among electrons, protons, and neutrinos studied earlier are transformations between different principal bundles within the same Lie group. The mutual transformations between photons and spin-1/2 elementary particles are transformations between two different Lie groups.

Experimentally, it is observed that high-energy photons passing near an atomic nucleus can undergo the following reaction:

$$\gamma \rightarrow e^+ + e^-.$$

According to the discussion in §21.2, photons can also undergo reactions such as

$$\gamma \rightarrow p + \bar{p}, \quad \gamma \rightarrow \nu + \bar{\nu}.$$

Since the photon is an  $SO(3)$  group, while spin-1/2 elementary particles (such as electron-positron pairs, proton-antiproton pairs, and neutrino-antineutrino pairs) are each an  $SU(2)$  group, the process of a photon transforming into spin-1/2 elementary particles is essentially the transformation of an  $SO(3)$  group into an  $SU(2)$  group. We already know that two distinct elements  $(Eg, -Eg)$  of the  $SU(2)$  group, correspond to the same element in the  $SO(3)$  group, where

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad -E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad g \in SU(2).$$

In §16.13, we demonstrated that two  $SU(2)$  groups with opposite orientations can combine to form a projective space  $RP^3$  or an  $SO(3)$  group, or equivalently, two spin-1/2 elementary particles with opposite orientations can combine to form a photon. Conversely, a photon can also transform into two spin-1/2 elementary particles.

Because the  $SO(3)$  group is determined by its three spin operators  $\hat{L}_x, \hat{L}_y, \hat{L}_z$ , and the  $SU(2)$  group is determined by its three spin operators  $\hat{S}_x, \hat{S}_y, \hat{S}_z$ , the conversion of the spin operators  $\hat{L}_x, \hat{L}_y, \hat{L}_z$  into  $\hat{S}_x, \hat{S}_y, \hat{S}_z$  leads to the transformation of an  $SO(3)$  group into an  $SU(2)$  group.

This section investigates how the three spin operators  $\hat{L}_x, \hat{L}_y, \hat{L}_z$  of the photon can be transformed into the spin operators  $\hat{S}_x, \hat{S}_y, \hat{S}_z$ .

A photon is an  $SO(3)$  group; its three spin operators can be expressed as

$$\hat{L}_x = -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right), \quad \hat{L}_y = -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right), \quad \hat{L}_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right).$$

Their commutation relations are

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z, \quad [\hat{L}_y, \hat{L}_z] = i\hbar \hat{L}_x, \quad [\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y.$$

If the point-like particle of a photon collides with or comes sufficiently close to the point-like particle of a spin-1/2 elementary particle, then the one-parameter subgroup determined by the element (tangent vector) of the Lie algebra at the identity element of the spin-1/2 particle acts as a transformation group and will transform the photon's spin operators. For example, a photon passing near an atomic nucleus inevitably interacts with a proton or neutron inside the nucleus. As another example, a high-energy proton (or antiproton) possesses high energy precisely because it contains high-energy photons. Consequently, the point-like particle of a photon inside a high-energy proton comes sufficiently close to the proton's identity element, making it possible for the photon to be acted upon by the transformation group determined by the tangent vector at the proton's identity element, and thus possibly transform into a spin-1/2 elementary particle, as in the following reaction:

$$p + p \rightarrow p + p + (\bar{p} + p).$$

In the following, we use the proton as an example to analyze this transforming effect of a spin-1/2 elementary particle on a photon.

The interaction between a photon and a certain proton can be represented as the transformation effect of the proton's  $SU(2)$  group acting as a Lie transformation group on the photon's  $SO(3)$  group:  $SU(2) \times SO(3)$ .

We express the spin operators  $\hat{S}_x, \hat{S}_y, \hat{S}_z$  of the spin-1/2 elementary particle as vector fields on the phase plane  $uv$ . Therefore, we need to analyze how  $\hat{L}_x, \hat{L}_y, \hat{L}_z$  are transformed into  $\hat{S}_x, \hat{S}_y, \hat{S}_z$  on the phase plane  $uv$ .

### 1. Transformation of the photon's spin operator $\hat{L}_x$

We transform the operator  $\hat{L}_x$  using the transformation formula (21.5.1). Changing the symbols in (21.5.1), we obtain

$$\begin{cases} u = y - z, \\ v = y + z. \end{cases}$$

Its Jacobian matrix is

$$J = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

The transpose of matrix  $J$  is

$$J^T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Thus, we obtain

$$\begin{pmatrix} f_* \frac{\partial}{\partial y} \\ f_* \frac{\partial}{\partial z} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix},$$

i.e.,

$$f_* \frac{\partial}{\partial y} = \frac{\partial}{\partial u} + \frac{\partial}{\partial v}, \quad f_* \frac{\partial}{\partial z} = -\frac{\partial}{\partial u} + \frac{\partial}{\partial v}.$$

From the expression of  $\hat{L}_x$ , we obtain

$$\begin{aligned}\hat{S}'_x &= -i\hbar \left( y(f_* \frac{\partial}{\partial z}) - z(f_* \frac{\partial}{\partial y}) \right) = -i\hbar \left( y(-\frac{\partial}{\partial u} + \frac{\partial}{\partial v}) - z(\frac{\partial}{\partial u} + \frac{\partial}{\partial v}) \right) \\ &= -i\hbar \left( (-y-z)\frac{\partial}{\partial u} + (y-z)\frac{\partial}{\partial v} \right) = -i\hbar \left( -v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v} \right).\end{aligned}$$

Thus, we obtain the vector field on the phase plane  $uv$  that results from transforming  $\hat{L}_x$ :

$$\hat{S}'_x = i\hbar \left( v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v} \right). \quad (21.8.1)$$

This can also be written as

$$\hat{S}'_x = \frac{i}{2}\hbar \left( v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v} \right) + \frac{i}{2}\hbar \left( v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v} \right) = \hat{S}_x + \hat{S}_x, \quad (21.8.2)$$

where

$$\hat{S}_x = \frac{i}{2}\hbar \left( v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v} \right).$$

Equation (21.8.2) tells us that the vector field  $\hat{S}'_x$  obtained from transforming  $\hat{L}_x$  can be regarded as the sum of two vector fields  $\hat{S}_x$ ; that is, the photon's spin operator  $\hat{L}_x$  can be transformed into the sum of two tangent vector fields  $\hat{S}_x$ , and  $\hat{S}_x$  is exactly one of the spin operators of a spin-1/2 elementary particle.

## 2. Transformation of the photon's spin operator $\hat{L}_y$

We transform the operator  $\hat{L}_y$  using the transformation formula (21.6.6). Changing the symbols in (21.6.6), we obtain

$$\begin{cases} u = x + iz, \\ v = ix + z. \end{cases}$$

Its Jacobian matrix is

$$\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}.$$

Thus, we obtain

$$f_* \frac{\partial}{\partial x} = \frac{\partial}{\partial u} + i \frac{\partial}{\partial v}, \quad f_* \frac{\partial}{\partial z} = i \frac{\partial}{\partial u} + \frac{\partial}{\partial v}.$$

From the expression of  $\hat{L}_y$ , we obtain

$$\begin{aligned}\hat{S}'_y &= -i\hbar \left( z(f_* \frac{\partial}{\partial x}) - x(f_* \frac{\partial}{\partial z}) \right) = -i\hbar \left( z\frac{\partial}{\partial u} + zi\frac{\partial}{\partial v} - xi\frac{\partial}{\partial u} - x\frac{\partial}{\partial v} \right) \\ &= -i\hbar \left( (z-xi)\frac{\partial}{\partial u} + (zi-x)\frac{\partial}{\partial v} \right) = -\hbar \left( (iz+x)\frac{\partial}{\partial u} + (-z-ix)\frac{\partial}{\partial v} \right) \\ &= -\hbar \left( u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v} \right).\end{aligned}$$

Thus, we obtain the vector field on the phase plane  $uv$  that results from transforming  $\hat{L}_y$ :

$$\hat{S}'_y = -\hbar \left( u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v} \right). \quad (21.8.3)$$

This can also be written as

$$\hat{S}'_y = -\frac{1}{2}\hbar \left( u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v} \right) - \frac{1}{2}\hbar \left( u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v} \right) = \hat{S}_y + \hat{S}_y, \quad (21.8.4)$$

where

$$\hat{S}_y = -\frac{1}{2}\hbar\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right).$$

Equation (21.8.4) tells us that the vector field  $\hat{S}'_y$  obtained from transforming  $\hat{L}_y$  can be regarded as the sum of two vector fields  $\hat{S}_y$ ; that is, the photon's spin operator  $\hat{L}_y$  can be transformed into the sum of two tangent vector fields  $\hat{S}_y$ , and  $\hat{S}_y$  is exactly one of the spin operators of a spin-1/2 elementary particle.

### 3. Transformation of the photon's spin operator $\hat{L}_z$

We transform the operator  $\hat{L}_z$  using the transformation formula (21.7.6). Changing the symbols in (21.7.6), we obtain

$$\begin{cases} u = (1+i)x, \\ v = (1-i)y. \end{cases}$$

Thus, we obtain

$$f_*\frac{\partial}{\partial x} = (1+i)\frac{\partial}{\partial u}, \quad f_*\frac{\partial}{\partial y} = (1-i)\frac{\partial}{\partial v}.$$

From the expression of  $\hat{L}_z$ , we obtain

$$\begin{aligned} \hat{S}'_z &= -i\hbar\left(x(f_*\frac{\partial}{\partial y}) - y(f_*\frac{\partial}{\partial x})\right) = -i\hbar\left(x(1-i)\frac{\partial}{\partial v} - y(1+i)\frac{\partial}{\partial u}\right) \\ &= -\hbar\left(x(i+1)\frac{\partial}{\partial v} - y(i-1)\frac{\partial}{\partial u}\right) = -\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right). \end{aligned}$$

Thus, we obtain the vector field on the phase plane  $uv$  that results from transforming  $\hat{L}_z$ :

$$\hat{S}'_z = -\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right). \quad (21.8.5)$$

This can also be written as

$$\hat{S}'_z = -\frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right) - \frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right) = \hat{S}_z + \hat{S}_z, \quad (21.8.6)$$

where

$$\hat{S}_z = -\frac{1}{2}\hbar\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right).$$

Equation (21.8.6) tells us that the vector field  $\hat{S}'_z$  obtained from transforming  $\hat{L}_z$  can be regarded as the sum of two vector fields  $\hat{S}_z$ ; that is, the photon's spin operator  $\hat{L}_z$  can be transformed into the sum of two tangent vector fields  $\hat{S}_z$ , and  $\hat{S}_z$  is exactly one of the spin operators of a spin-1/2 elementary particle.

The commutation relations of  $\hat{S}_x, \hat{S}_y, \hat{S}_z$  are

$$[\hat{S}_x, \hat{S}_y] = i\hbar\hat{S}_z, \quad [\hat{S}_y, \hat{S}_z] = i\hbar\hat{S}_x, \quad [\hat{S}_z, \hat{S}_x] = i\hbar\hat{S}_y,$$

which are identical to those of  $\hat{L}_x, \hat{L}_y, \hat{L}_z$ . From the commutation relations of  $\hat{S}_x, \hat{S}_y, \hat{S}_z$ , we obtain the commutation relations of  $\hat{S}'_x, \hat{S}'_y, \hat{S}'_z$ :

$$[\hat{S}'_x, \hat{S}'_y] = 2i\hbar\hat{S}'_z, \quad [\hat{S}'_y, \hat{S}'_z] = 2i\hbar\hat{S}'_x, \quad [\hat{S}'_z, \hat{S}'_x] = 2i\hbar\hat{S}'_y.$$

The calculation results in this section can serve as an example of Theorem 12.6.2. The proton, as a Lie transformation group  $SU(2)$ , acts on the photon on the right, forming an isomorphism as follows:

$$\theta: SU(2) \times SO(3) \rightarrow SO(3).$$

This isomorphic mapping induces an isomorphism from the Lie algebra  $\text{Lie}(G)$  of the  $SU(2)$

group to the Lie algebra  $\chi(M)$  of the  $SO(3)$  group:

$$\tilde{\theta} : \text{Lie}(G) \rightarrow \chi(M).$$

Therefore, for any  $X, Y \in \text{Lie}(G)$ , the equality  $\tilde{\theta}[X, Y] = [\tilde{\theta}(X), \tilde{\theta}(Y)]$  holds. This is precisely why the commutation relations of  $\hat{S}'_x, \hat{S}'_y, \hat{S}'_z$  are identical to those of  $\hat{L}_x, \hat{L}_y, \hat{L}_z$ , and also to those of  $\hat{S}_x, \hat{S}_y, \hat{S}_z$ .

#### 4. Summary

Combining the above calculations, we find that all three spin operators of the photon can be transformed into the sum of two spin operators of a spin-1/2 elementary particle. This is consistent with the earlier conclusion: an  $SO(3)$  group can be viewed as the inner direct product of two  $SU(2)$  groups; that is, an  $SO(3)$  group can be decomposed into two  $SU(2)$  groups. Alternatively, two  $SU(2)$  groups with opposite orientations can combine to form one  $SO(3)$  group. Therefore, with the aid of external conditions, a photon can decompose into two spin-1/2 elementary particles. The calculation process above provides one such external condition, namely the three transformation formulas (21.5.2), (21.6.6), and (21.7.6). That is, the point-like particle of the photon collides with or comes sufficiently close to the point-like particle of a spin-1/2 elementary particle (i.e., their identity elements collide or become sufficiently close). Of course, this decomposition process must also satisfy the laws of conservation of energy, momentum, angular momentum, and charge.

## §21.9 The Structure and Composition of the Universe

### 1. The Structure of the Universe

Let the coordinates of a four-dimensional Euclidean space  $R^4$  be  $(x, y, z, w)$ . Set  $a = w + iz$ ,  $b = -y + ix$ , where  $i$  is the imaginary unit, and define

$$g(x, y, z, w) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} = \begin{pmatrix} w + iz & -y + ix \\ y + ix & w - iz \end{pmatrix}. \quad (21.9.1)$$

Denote the set of all such matrices  $g(x, y, z, w)$  by  $M(4, R^4)$ . Although  $M(4, R^4)$  is not a Lie group, because the elements of  $M(4, R^4)$  are in one-to-one correspondence with the points of  $R^4$ , we may also regard  $M(4, R^4)$  as  $R^4$ .

Let  $GL(4, R^4)$  be the set of all invertible matrices  $g(x, y, z, w)$ , i.e.,

$$GL(4, R^4) = \{Q \in M(4, R^4) \mid Q \text{ is invertible}\}. \quad (21.9.2)$$

Clearly,  $GL(4, R^4)$  is a subset of  $R^4$ .

We may view the operation of taking the determinant  $\det Q$  of the matrix  $Q$  as a mapping

$$\det : Q \rightarrow R. \quad (21.9.3)$$

Obviously,  $\det Q$  is a continuous function because

$$\lim_{\substack{(x, y, z, w) \rightarrow \\ (x_0, y_0, z_0, w_0)}} \begin{pmatrix} w + iz & -y + ix \\ y + ix & w - iz \end{pmatrix} = \begin{pmatrix} w_0 + iz_0 & -y_0 + ix_0 \\ y_0 + ix_0 & w_0 - iz_0 \end{pmatrix}.$$

Hence, the mapping (21.9.3) is continuous.

Since the determinant of the matrix  $Q$  satisfies  $\det Q \neq 0$ , the mapping (21.9.3) can be written as

$$\det : Q \rightarrow (-\infty, 0) \cup (0, \infty). \quad (21.9.4)$$

Because both  $(-\infty, 0)$  and  $(0, \infty)$  are open subsets of  $R$ , according to Theorem 1.5.1,  $\det^{-1}(-\infty, 0)$  and  $\det^{-1}(0, \infty)$  are open subsets of  $GL(4, R^4)$ . Therefore,  $GL(4, R^4)$  is an open

set. Since  $GL(4, R^4)$  is a subset of  $R^4$ , it follows that  $GL(4, R^4)$  is an open subset of  $R^4$ .

Define a subset  $SU(2)$  of  $GL(4, R^4)$  by

$$SU(2) = \{Q \in GL(4, R^4) \mid Q^+ = Q^{-1}, |Q| = 1\}, \quad (21.9.5)$$

where  $+$  denotes the complex conjugate transpose (Hermitian conjugate). The set  $SU(2)$  is a group and is a Lie subgroup of the general linear group  $GL(4, R^4)$ .

From  $|Q| = 1$  and (21.9.1) we obtain

$$x^2 + y^2 + z^2 + w^2 = 1, \quad (21.9.6)$$

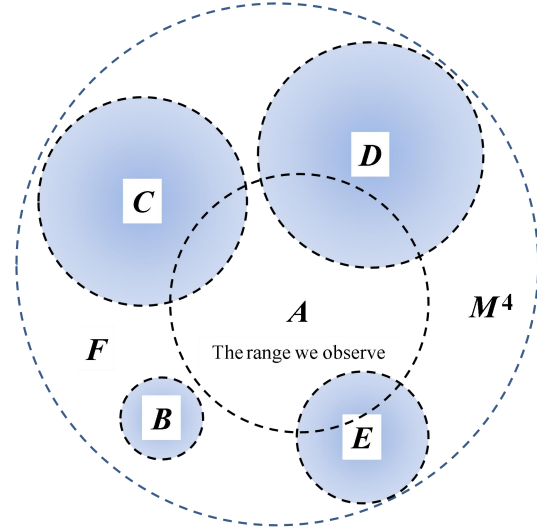
so the  $SU(2)$  group is homeomorphic to the three-dimensional unit sphere  $S^3$  of radius 1. Denote

$$S^3 = \{(x, y, z, w) \in R^4 \mid x^2 + y^2 + z^2 + w^2 = 1\}. \quad (21.9.7)$$

The  $SU(2)$  group (or  $S^3$ ) is a subset of  $R^4$ . Identifying antipodal points of  $S^3$  yields the real projective space  $RP^3$ ; hence  $RP^3$  is also a subset of  $R^4$ . That is, both  $S^3$  and  $RP^3$  can be embedded into  $R^4$  as hypersurfaces, and in fact as hypersurfaces that lie in an open subset of  $R^4$ .

Now set  $w = ict$  (i.e., change the coordinates  $(x, y, z, w)$  of the space  $R^4$  to  $(x, y, z, ict)$ ). Then the space  $R^4$  becomes the Minkowski spacetime  $M^4$ . Consequently,  $S^3$  and  $RP^3$  become hypersurfaces contained in an open set of  $M^4$ , and naturally they are also hypersurfaces in  $M^4$ .

Based on the analysis above, we can envision the structure of the universe. As shown in Figure 21.9.1, we regard the entirety of Minkowski spacetime  $M^4$  as the entire universe. This universe is covered by numerous open subsets  $A, B, C, D, E, F, \dots$ . These open subsets can vary in size and may intersect or be disjoint. All these open subsets together form an open cover of the  $M^4$  smooth manifold (the entire universe). Since any two open subsets in Minkowski spacetime  $M^4$  are homeomorphic — indeed, each open subset can be homeomorphic to the entire Minkowski spacetime  $M^4$  — the entire universe can be divided into several parts. Each part is an open subset, can be viewed as a small universe, and contains many  $S^3$  and  $RP^3$ , i.e., contains many elements particles and the particles composed of these elements particles. Each particle is a part of a small universe or of the entire universe. The universe we observe is merely an open set within  $M^4$ .



**Figure 21.9.1** The Structure of the Universe

Each particle is a part of a small universe or of the entire universe. The universe we observe is merely an open set within  $M^4$ .

## 2.Types of Elementary Particles

### (1)Spin-1/2 Elementary Particles

Let  $V$  be an  $m$ -dimensional linear space over the complex field. For any elements

$$\alpha = (a^1, a^2, \dots, a^m), \quad \beta = (b^1, b^2, \dots, b^m) \in V$$

define the inner product

$$\langle \alpha, \beta \rangle = a^1 b^{1*} + a^2 b^{2*} + \dots + a^m b^{m*},$$

where  $*$  denotes complex conjugation. Then  $V$  is called a **unitary space**.

A linear transformation  $U$  on  $V$  that preserves the inner product, i.e., satisfies

$$\langle U\alpha, U\beta \rangle = \langle \alpha, \beta \rangle,$$

is called a **unitary transformation**. The matrix of a unitary transformation with respect to an

orthonormal basis is called a **unitary matrix**. A unitary matrix  $U$  satisfies

$$U^{*T} = U^+ = U^{-1}, \quad \det U = |U| = \pm 1,$$

where  $T$  denotes the transpose operation on the matrix  $U$ , and the complex conjugate transpose is also called the Hermitian conjugate, denoted by the symbol  $+$ . An  $SU(2)$  matrix is a three-dimensional unitary matrix with determinant 1, i.e.,

$$SU(2) = \{U \mid U \in GL(2, \mathbb{C}), U^+ = U^{-1}, |U| = 1\}.$$

All such  $SU(2)$  matrices form the  $SU(2)$  group. We already know that every elementary spin-1/2 particle corresponds to an  $SU(2)$  group.

### (2)Photons

Let  $V$  be an  $m$ -dimensional linear space over the real field. For any elements

$$\alpha = (a^1, a^2, \dots, a^m), \quad \beta = (b^1, b^2, \dots, b^m) \in V$$

define the inner product

$$\langle \alpha, \beta \rangle = a^1 b^1 + a^2 b^2 + \dots + a^m b^m,$$

then  $V$  is called a **Euclidean space**  $R^m$ .

A linear transformation  $A$  on  $V$  that preserves the inner product, i.e., satisfies

$$\langle A\alpha, A\beta \rangle = \langle \alpha, \beta \rangle,$$

is called an **orthogonal transformation**. The matrix of an orthogonal transformation with respect to an orthonormal basis is called an **orthogonal matrix**. An orthogonal matrix  $A$  satisfies

$$A^T = A^{-1}, \quad \det A = |A| = \pm 1.$$

An  $SO(3)$  matrix is a three-dimensional orthogonal matrix with determinant 1, i.e.,

$$SO(3) = \{A \mid A \in GL(3, \mathbb{R}), A^T = A^{-1}, |A| = 1\}.$$

All such  $SO(3)$  matrices form the  $SO(3)$  group. We already know that every photon corresponds to an  $SO(3)$  group.

### (3)Elementary Dark Particles

Let  $V$  be an  $m$ -dimensional linear space over the real field. For any elements

$$\alpha = (a^1, a^2, \dots, a^m), \quad \beta = (b^1, b^2, \dots, b^m) \in V$$

define the inner product

$$\langle \alpha, \beta \rangle = -a^1 b^1 + a^2 b^2 + \dots + a^m b^m,$$

then  $V$  is called a **Minkowski space**  $M^m$ .

A linear transformation  $A$  on  $V$  that preserves this inner product, i.e., satisfies

$$\langle A\alpha, A\beta \rangle = \langle \alpha, \beta \rangle,$$

is called a **Lorentz transformation**. The matrix of a Lorentz transformation with respect to an orthonormal basis is called a **Lorentz matrix**. A Lorentz matrix  $A$  satisfies

$$\det A = |A| = \pm 1.$$

If  $V$  is four-dimensional and we set

$$\alpha = (t^1, x^1, y^1, z^1), \quad \beta = (t^2, x^2, y^2, z^2),$$

where  $t$  is the time coordinate and  $x, y, z$  are the spatial coordinates, then

$$\langle \alpha, \beta \rangle = -t^1 t^2 + x^1 x^2 + y^1 y^2 + z^1 z^2,$$

and  $V$  becomes the four-dimensional Minkowski spacetime  $M^4$ .

If  $V$  is three-dimensional, the set of all real  $3 \times 3$  transformation matrices  $A = (a_\mu^\nu)$ ,  $\mu, \nu = 0, 1, 2$ , that preserve the quadratic form  $-t^2 + x^2 + y^2$  ( $t, x, y \in \mathbb{R}$ ) forms a three-dimensional group called the **general Lorentz group in three dimensions**. If, in addition,  $A = (a_\mu^\nu)$  satisfies  $\det A = 1$ ,  $a_0^0 > 0$ , then the set of all such matrices  $A = (a_\mu^\nu)$  forms the **proper Lorentz group in three dimensions**, denoted by  $L_+^\uparrow$ :

$$L_+^\uparrow = \{A \mid A^T \eta A = \eta, \det A = 1, a_0^0 > 0\}, \quad (21.9.8)$$

where

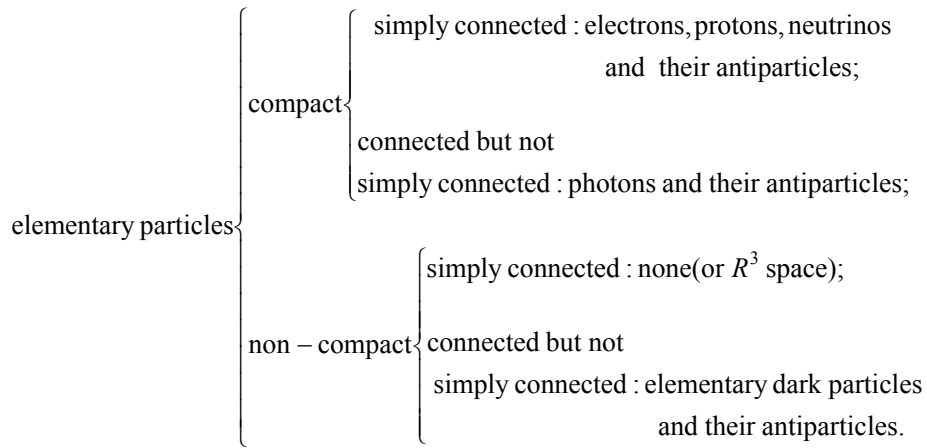


$$\eta = (\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mu, \nu = 0, 1, 2. \quad (21.9.9)$$

The  $L_+^\uparrow$  group is a connected Lie group.

Now we recognize that three-dimensional Lie groups include not only  $SU(2)$  and  $SO(3)$ , but also  $L_+^\uparrow$ . Since the  $SU(2)$  group and  $SO(3)$  group correspond respectively to a type of particle, the  $L_+^\uparrow$  group should also correspond to a type of particle—the **elementary dark particle**. We take the  $L_+^\uparrow$  group as the model for the elementary dark particle; i.e., we assume that an elementary dark particle is homeomorphic to the three-dimensional proper Lorentz group  $L_+^\uparrow$ .

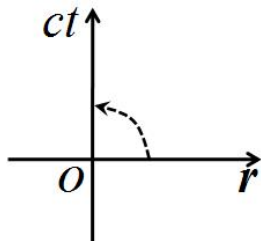
Because the  $SU(2)$  group and  $SO(3)$  group are compact, photons and spin-1/2 elementary particles are compact. The  $L_+^\uparrow$  group is non-compact; therefore, elementary dark particles are non-compact. We can classify elementary particles as follows:



## §21.10 The Origin of Matter–Antimatter Asymmetry

### 1.Orientation of Minkowski Spacetime

In Minkowski spacetime  $M^4$ , a real coordinate system  $(x, y, z, ct)$  can be established, where  $x, y, z$  are spatial coordinates,  $t$  is the time coordinate, and  $c$  is the speed of light. Denote  $\mathbf{r} = (x, y, z)$ .



**Figure 21.10.1** Right-handed orientation

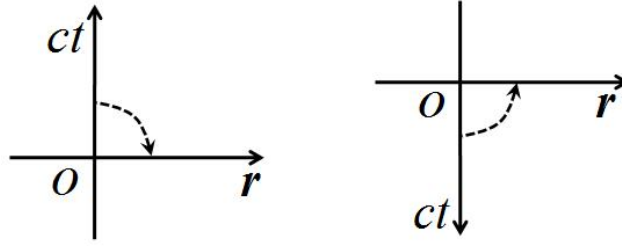
$M^4$  is a connected, orientable smooth manifold, thus it admits two distinct orientations. The transition from the spatial coordinates  $\mathbf{r}$  to the time coordinate  $ct$  is called a **right-handed orientation**, represented by the 4-form

$$cdx \wedge dy \wedge dz \wedge dt, \quad (21.10.1)$$

as shown in Figure 21.10.1. Conversely, the transition from the time coordinate  $ct$  to the spatial coordinates  $\mathbf{r}$  is called a **left-handed orientation**, represented by the 4-form

$$cdt \wedge dx \wedge dy \wedge dz, \quad (21.10.2)$$

as shown in Figure 21.10.2.



**Figure 21.10.2** Left-handed orientation

Since

$$cdx \wedge dy \wedge dz \wedge dt = -cdx \wedge dy \wedge dt \wedge dz = cdx \wedge dt \wedge dy \wedge dz = -cdt \wedge dx \wedge dy \wedge dz,$$

we have

$$cdx \wedge dy \wedge dz \wedge dt = -cdt \wedge dx \wedge dy \wedge dz.$$

Moreover,

$$cdx \wedge dy \wedge dz \wedge d(-t) = -cdx \wedge dy \wedge dz \wedge dt = cdt \wedge dx \wedge dy \wedge dz.$$

Therefore, if in a right-handed real coordinate system we keep the spatial coordinates unchanged but invert the time coordinate, i.e.,  $t \rightarrow -t$ , then the orientation of  $M^4$  changes from right-handed to left-handed (compare Figure 21.10.1 with the right panel of Figure 21.10.2). Equivalently, If we let the signs of the spatial and time coordinates in a right-handed oriented real coordinate system be opposite, then its orientation becomes left-handed.

Since  $M^4$  admits two different orientations, which one should we adopt? In practice, we have already taken  $M^4$  to be **left-handed oriented**. The reason is that in classical physics, we typically set up a three-dimensional spatial coordinate system  $(x, y, z)$  that follows the right-handed rule for force analysis. When studying the motion of objects, we then introduce the time coordinate  $t$ . This might seem to imply a right-handed orientation; however, in many physical formulas and equations, the signs of the spatial and time coordinates are opposite. According to the analysis above, this means that in spacetime  $M^4$  we actually adopt a left-handed orientation. Several examples illustrate this point.

In the Lorentz transformation, the inertial frame  $\Sigma$  and the inertial frame  $\Sigma'$  have the same orientation. Assume the  $x$ -axis of frame  $\Sigma$  and the  $x'$ -axis of frame  $\Sigma'$  are parallel and point in the same direction, and  $\Sigma'$  moves relative to  $\Sigma$  with constant velocity  $v$  along the positive  $x$ -direction. The Lorentz transformation is

$$\begin{cases} x' = \gamma(x - vt), \\ y' = y, \\ z' = z, \\ t' = \gamma\left(t - \frac{v}{c^2}x\right), \end{cases} \quad (21.10.3)$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Equation (21.10.3) describes the motion of frame  $\Sigma'$  as seen from frame  $\Sigma$ . Because  $v > 0$ , it is clear from (21.10.3) that in frame  $\Sigma$  the coordinate  $x$  and the time  $t$  have opposite signs; hence frame  $\Sigma$  is left-handed oriented.

The inertial frames  $\Sigma$  and  $\Sigma'$  have the same orientation because

$$\begin{aligned} dx' \wedge dy' \wedge dz' \wedge dt' &= \gamma(dx - vdt) \wedge dy \wedge dz \wedge \gamma\left(dt - \frac{v}{c^2}dx\right) \\ &= \gamma^2(dx \wedge dy \wedge dz - vdt \wedge dy \wedge dz) \wedge \left(dt - \frac{v}{c^2}dx\right) \\ &= \gamma^2(dx \wedge dy \wedge dz \wedge dt + \frac{v^2}{c^2}dt \wedge dy \wedge dz \wedge dx) \end{aligned}$$

$$\begin{aligned}
&= \gamma^2 (\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}t - \frac{v^2}{c^2} \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}t) \\
&= \gamma^2 \left( 1 - \frac{v^2}{c^2} \right) \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}t \\
&= \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}t.
\end{aligned}$$

Since  $\Sigma$  and  $\Sigma'$  share the same orientation, frame  $\Sigma'$  is also left-handed oriented:

$$\mathrm{d}x' \wedge \mathrm{d}y' \wedge \mathrm{d}z' \wedge \mathrm{d}(-t') = \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}(-t).$$

In classical physics, the wave function of a sinusoidal wave propagating along the positive  $x$ -direction is

$$E(x, t) = A \cos(kx - \omega t),$$

where  $k > 0, \omega > 0$ . Because  $k > 0$  and  $\omega > 0$ , the variables  $x$  and  $t$  again appear with opposite signs. Consequently, the coordinate system in which the wave function is described is also left-handed oriented.

Comparing the momentum and energy operators

$$-i\hbar \frac{\partial}{\partial x}, \quad -i\hbar \frac{\partial}{\partial y}, \quad -i\hbar \frac{\partial}{\partial z}, \quad i\hbar \frac{\partial}{\partial t},$$

we observe that the spatial coordinates  $x, y, z$  and the time coordinate  $t$  once more carry opposite signs. Therefore, the coordinate system in which these operators are defined is also left-handed oriented.

The reason these coordinate systems are taken to be left-handed oriented is that the conclusions derived from them agree with experimental facts, which in turn indicates that in reality spacetime  $M^4$  itself is left-handed oriented. Hence, in the preceding chapters as well as in the following discussion, we shall consistently adopt the left-handed orientation for spacetime  $M^4$ .

## 2. The Reason for the Scarcity of Antimatter

A particle is an orientable smooth manifold (i.e., a hypersurface)  $S$  embedded in spacetime  $M^4$ .  $S$  admits two distinct orientations. The orientation of spacetime  $M^4$  can induce an orientation on the particle; according to Theorem 9.4.4, this is indeed possible. On  $S$  choose a frame  $(p; n, e_1, e_2, e_3)$ , called a **Darboux frame**, where  $p$  is a point on  $S$ ,  $(e_1, e_2, e_3)$  are tangent vectors at  $p$  on  $S$ , and  $n$  is the normal vector at  $p$ .  $(e_1, e_2, e_3)$  is a basis of the tangent space  $T_p S$  at point  $p$  on  $S$ , and  $(n, e_1, e_2, e_3)$  is a basis of the tangent space  $T_p M^4$  at point  $p$  on  $M^4$ . If the orientation of  $(n, e_1, e_2, e_3)$  coincides with the chosen orientation of  $M^4$ , then  $(e_1, e_2, e_3)$  gives the orientation of the particle's smooth manifold  $S$ .

A particle is a three-dimensional smooth manifold  $S$ . Suppose originally  $S$  is not a submanifold of  $M^4$ ; for  $S$  to become a submanifold of  $M^4$ , there must be a mapping that embeds  $S$  into  $M^4$ . This mapping is the **inclusion map**; let  $f: S \rightarrow M^4$  be this inclusion. Let the 4-form  $\omega$  determine the orientation of  $M^4$ , then  $i_n \omega = n \lrcorner \omega$  is a 3-form on  $M^4$ . And then  $\sigma = f_S^*(n \lrcorner \omega)$  is a 3-form on  $S$  obtained via the **pullback map**  $f_S^*$  (or equivalently, by restriction to the submanifold  $S$ ). That is,

$$\sigma = f_S^*(n \lrcorner \omega) = n \lrcorner \omega(e_1, e_2, e_3) = \omega(n, e_1, e_2, e_3).$$

According to Theorem 9.4.4,  $\sigma = f_S^*(n \lrcorner \omega)$  determines an orientation on the submanifold  $S$ , and this orientation is **consistent** with the orientation of  $M^4$ .

We can now begin to understand why there are more particles (or matter) than antiparticles (or antimatter) in the universe. Earlier we defined the photon, electron, proton, and neutrino as **left-handed oriented** particles, while the antiphoton, positron, antiproton, and antineutrino are **right-handed oriented** particles. Particles generated in spacetime  $M^4$  are inevitably influenced by the orientation of  $M^4$ . Since the orientation of spacetime  $M^4$  is left-handed, particles

created within  $M^4$  will most likely also be left-handed oriented—that is, they will most likely be **ordinary particles**. Nevertheless, exceptions can occur, leading to the production of antiparticles. The specific circumstances will be discussed in the next section.

## §21.11 Properties of Dark Particles and the Birth of Particles

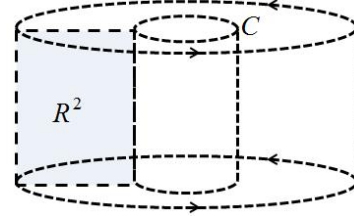
We refer to the elementary dark particles and the matter composed of them as dark matter. Astronomical observations over the past century have revealed the existence of dark matter in the universe. Dark matter possesses the following characteristics<sup>6</sup>:

- 1)It cannot be baryonic matter; it carries no electric charge or color charge.
- 2)It cannot be composed of Standard Model particles.
- 3)It does not participate in electromagnetic interactions; therefore, it neither emits nor absorbs light.
- 4)Most of it is cold and/or warm, non-relativistic or at least not extremely relativistic.
- 5)It does interact via gravity. The gravitational effects of dark matter are ubiquitous in all aspects of cosmic structure.

This section presents several important properties of dark particles.

### 1.Topological Structure of Elementary Dark Particles

The manifold structure of the three-dimensional group  $L_+^\uparrow$  is  $R^2 \times S^1$  (the manifold structure of the four-dimensional group  $L_+^\uparrow$  in four-dimensional Minkowski spacetime is  $R^3 \times S^3$ , which is six-dimensional). Therefore, an elementary dark particle is an  $R^2 \times S^1$  smooth manifold, which can be visualized as rotating  $R^2$  along  $S^1$ , forming a hollow three-dimensional solid cylindrical surface, as shown in Figure 21.11.1.



**Figure 21.11.1**  $R^2 \times S^1$  formed by rotating  $R^2$  along  $S^1$

### 2.Dark Particles Have Antiparticles

Since an elementary dark particle is an  $R^2 \times S^1$  smooth manifold, and both  $R^2$  and  $S^1$  are connected, orientable smooth manifolds, according to Theorem 9.5.1,  $R^2 \times S^1$  is orientable, and according to Theorem 1.9.1,  $R^2 \times S^1$  is connected. Moreover, because it is connected, it admits two distinct orientations; hence, elementary dark particles possess antiparticles. Dark matter composed of elementary dark particles also has antiparticles.

### 3.The Spin of an Elementary Dark Particle is 1

#### (1)The $L_+^\uparrow$ Group

The three-dimensional  $L_+^\uparrow$  group represents rotational transformations between the  $x$ - and  $y$ -axes, as well as boost transformations between the  $t$ -axis and the  $x$ -axis, and between the  $t$ -axis and the  $y$ -axis. The rotational transformation between the  $x$ - and  $y$ -axes is represented by the matrix

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \quad (21.11.1)$$

Setting  $x = \cos \theta$ ,  $y = \sin \theta$ , we have

<sup>6</sup>Yang Binglin, translated by Liu Guoli, Wang Wenyu, Wang Fei. *Dark Matter and Related Cosmology* [M]. Beijing: Science Press, 2019,p.7.

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & x & -y \\ 0 & y & x \end{pmatrix}.$$

The determinant of the matrix  $A_1$  is

$$|A_1| = x^2 + y^2 = 1,$$

so the matrix  $A_1$  is homeomorphic to the circle  $S^1$ . The boost transformation between the  $t$ -axis and the  $x$ -axis is represented by the matrix

$$A_2 = \begin{pmatrix} \cosh \sigma & -\sinh \sigma & 0 \\ -\sinh \sigma & \cosh \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (21.11.2)$$

where

$$\cosh \sigma = \frac{1}{2}(e^\sigma + e^{-\sigma}), \quad \sinh \sigma = \frac{1}{2}(e^\sigma - e^{-\sigma}), \quad \sigma \in (-\infty, \infty).$$

The matrix  $A_2$  is homeomorphic to a line. The boost transformation between the  $t$ -axis and the  $y$ -axis is represented by the matrix

$$A_3 = \begin{pmatrix} \cosh \tau & 0 & -\sinh \tau \\ 0 & 1 & 0 \\ -\sinh \tau & 0 & \cosh \tau \end{pmatrix}, \quad (21.11.3)$$

where  $\tau \in (-\infty, \infty)$ . The matrix  $A_3$  is also homeomorphic to a line.

## (2) Matrix Form of the Spin Operators of the $L_+^\uparrow$ Group

We now compute the infinitesimal generators of the  $L_+^\uparrow$  group:

$$\hat{X}_1 = \left. \frac{\partial A_1}{\partial \theta} \right|_{\{\theta=0\}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (21.11.4)$$

$$\hat{X}_2 = \left. \frac{\partial A_2}{\partial \sigma} \right|_{\{\sigma=0\}} = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (21.11.5)$$

$$\hat{X}_3 = \left. \frac{\partial A_3}{\partial \tau} \right|_{\{\tau=0\}} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \quad (21.11.6)$$

The commutation relations of  $\hat{X}_1, \hat{X}_2, \hat{X}_3$  are

$$[\hat{X}_1, \hat{X}_2] = \hat{X}_3, \quad [\hat{X}_2, \hat{X}_3] = -\hat{X}_1, \quad [\hat{X}_3, \hat{X}_1] = \hat{X}_2. \quad (21.11.7)$$

Therefore, from equation (12.3.6) we obtain

$$\begin{aligned} A = g(\theta, \sigma, \tau) &= \exp[\theta \hat{X}_1 + \sigma \hat{X}_2 + \tau \hat{X}_3] = \exp\left[\frac{1}{\hbar}(-i\theta \hbar \hat{X}_1 + \sigma \hbar \hat{X}_2 + \tau \hbar \hat{X}_3)\right] \\ &= \exp\left[\frac{1}{\hbar}(-i\theta \hat{J}_1 + \sigma \hat{J}_2 + \tau \hat{J}_3)\right], \end{aligned} \quad (21.11.8)$$

where

$$\hat{J}_1 = i\hbar \hat{X}_1 = \hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \hat{J}_2 = \hbar \hat{X}_2 = \hbar \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \hat{J}_3 = \hbar \hat{X}_3 = \hbar \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

The commutation relations of  $\hat{J}_1, \hat{J}_2, \hat{J}_3$  are

$$[\hat{J}_1, \hat{J}_2] = i\hbar \hat{J}_3, \quad [\hat{J}_2, \hat{J}_3] = i\hbar \hat{J}_1, \quad [\hat{J}_3, \hat{J}_1] = i\hbar \hat{J}_2. \quad (21.11.9)$$

We have already found that the eigenvalues of  $\hat{J}_1$  are  $0, \pm\hbar$ , three in total. Let us now find the eigenvalues and eigenfunctions of  $\hat{J}_2$ . The eigenvalue equation for  $\hat{J}_2$  is

$$\hbar \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \end{pmatrix} = \lambda \begin{pmatrix} t \\ x \\ y \end{pmatrix}. \quad (21.11.10)$$

Simplifying,

$$\begin{cases} -\hbar x = \lambda t, \\ -\hbar t = \lambda x, \\ 0y = \lambda y. \end{cases}$$

From the first and second equations we get  $\lambda = \pm\hbar$ . The solution of the system, i.e., the eigenfunction, is a column matrix

$$\begin{pmatrix} \mp c \\ c \\ 0 \end{pmatrix},$$

where  $c$  is an arbitrary constant. When  $\lambda = \hbar$ , the first row takes the minus sign; when  $\lambda = -\hbar$ , the first row takes the plus sign.

If the eigenvalue  $\lambda$  in equation (21.11.10) equals 0, then equation (21.11.10) becomes

$$\hbar \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \end{pmatrix} = 0. \quad (21.11.11)$$

Simplifying,

$$\begin{cases} -x = 0, \\ -t = 0, \\ 0y = 0. \end{cases}$$

The solution of the eigenvalue equation, i.e., the eigenfunction, is a column matrix

$$\begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix},$$

where  $c$  is an arbitrary constant.

The eigenvalues of  $\hat{J}_3$  are also  $0, \pm\hbar$ , the same as those of  $\hat{J}_1$  and  $\hat{J}_2$ . Therefore, the spin of an elementary dark particle, like that of a photon, has a spin quantum number of 1.

### (3) Differential Form of the Spin Operators of the $L_+^\uparrow$ Group

Next, we derive the differential form of the spin operators of the  $L_+^\uparrow$  group.

We have already found that the spin operator corresponding to matrix (21.11.4) is

$$\hat{L}_1 = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right).$$

If  $\sigma$  is an infinitesimal number, then

$$\begin{aligned} \cosh \sigma &= \frac{1}{2}(e^\sigma + e^{-\sigma}) = \frac{1}{2} \left( \left( 1 + \sigma + \frac{1}{2!}\sigma^2 + \dots \right) + \left( 1 - \sigma + \frac{1}{2!}\sigma^2 + \dots \right) \right) \approx 1, \\ \sinh \sigma &= \frac{1}{2}(e^\sigma - e^{-\sigma}) = \frac{1}{2} \left( \left( 1 + \sigma + \frac{1}{2!}\sigma^2 + \dots \right) - \left( 1 - \sigma + \frac{1}{2!}\sigma^2 + \dots \right) \right) \approx \sigma. \end{aligned}$$

Therefore, assuming  $\sigma$  is an infinitesimal number, then the matrix (21.11.2)

$$A_2 = \begin{pmatrix} \cosh \sigma & -\sinh \sigma & 0 \\ -\sinh \sigma & \cosh \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix} \approx \begin{pmatrix} 1 & -\sigma & 0 \\ -\sigma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The boost by an infinitesimal  $\sigma$  induces the coordinate transformation

$$\begin{pmatrix} t' \\ x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & -\sigma & 0 \\ -\sigma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \end{pmatrix}. \quad (21.11.12)$$

Let  $\psi(t, x, y)$  be a smooth function. Under the coordinate transformation (21.11.12),  $\psi(t, x, y)$  transforms as

$$\psi(t', x', y') = \psi(t - \sigma x, -\sigma t + x, y).$$

According to Theorem 3.1.1, we obtain

$$\begin{aligned} \psi(t', x', y') &= \psi(t - \sigma x, -\sigma t + x, y) \\ &= \psi(t, x, y) - \sigma x \frac{\partial}{\partial t} \psi(t, x, y) - \sigma t \frac{\partial}{\partial x} \psi(t, x, y) \\ &= \psi(t, x, y) - \sigma \frac{1}{\hbar} \left( \hbar x \frac{\partial}{\partial t} + \hbar t \frac{\partial}{\partial x} \right) \psi(t, x, y). \end{aligned}$$

Let

$$\hat{L}_2 = \hbar \left( x \frac{\partial}{\partial t} + t \frac{\partial}{\partial x} \right).$$

Similarly, assuming  $\tau$  is an infinitesimal number, then

$$A_3 = \begin{pmatrix} \cosh \tau & 0 & -\sinh \tau \\ 0 & 1 & 0 \\ -\sinh \tau & 0 & \cosh \tau \end{pmatrix} \approx \begin{pmatrix} 1 & 0 & -\tau \\ 0 & 1 & 0 \\ -\tau & 0 & 1 \end{pmatrix}.$$

The boost by an infinitesimal  $\tau$  induces the coordinate transformation

$$\begin{pmatrix} t' \\ x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 & -\tau \\ 0 & 1 & 0 \\ -\tau & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \end{pmatrix}. \quad (21.11.13)$$

The smooth function  $\psi(t, x, y)$  under this coordinate transformation (21.11.13) transforms as

$$\psi(t', x', y') = \psi(t - \tau y, x, -\tau t + y).$$

According to Theorem 3.1.1, we obtain

$$\begin{aligned} \psi(t', x', y') &= \psi(t - \tau y, x, -\tau t + y) \\ &= \psi(t, x, y) - \tau y \frac{\partial}{\partial t} \psi(t, x, y) - \tau t \frac{\partial}{\partial y} \psi(t, x, y) \\ &= \psi(t, x, y) - \tau \frac{1}{\hbar} \left( \hbar y \frac{\partial}{\partial t} + \hbar t \frac{\partial}{\partial y} \right) \psi(t, x, y). \end{aligned}$$

Let

$$\hat{L}_3 = \hbar \left( y \frac{\partial}{\partial t} + t \frac{\partial}{\partial y} \right).$$

The operators  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  are the **spin angular-momentum operators** of the elementary dark particle. Their commutation relations are

$$[\hat{L}_1, \hat{L}_2] = i\hbar \hat{L}_3, \quad [\hat{L}_2, \hat{L}_3] = i\hbar \hat{L}_1, \quad [\hat{L}_3, \hat{L}_1] = i\hbar \hat{L}_2. \quad (21.11.14)$$

#### 4.Dark Particles Possess a Gravitational Field

On the three-dimensional  $L_+^\uparrow$  group, there exists a bi-invariant metric. For example, we present a bi-invariant metric on the  $L_+^\uparrow$  group here. First, define a function: for any  $\hat{V} \in \text{Lie}(L_+^\uparrow)$ ,

$$\theta(\hat{V}) = -\hat{V}^T. \quad (21.11.15)$$

Based on this function, we obtain

$$\theta(\hat{X}_1) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \theta(\hat{X}_2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \theta(\hat{X}_3) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Then define a left-invariant metric

$$\langle \hat{V}_i, \hat{V}_j \rangle = -\text{tr}(\hat{V}_i \theta(\hat{V}_j)). \quad (21.11.16)$$

where  $\hat{V}_i$  and  $\hat{V}_j$  are left-invariant tangent vector fields in matrix form, and  $\text{tr}$  denotes the trace of the matrix  $\hat{V}_i \theta(\hat{V}_j)$ . For example,

$$\hat{X}_1 \theta(\hat{X}_1) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \text{tr}(\hat{X}_1 \theta(\hat{X}_1)) = -2,$$

so

$$\langle \hat{X}_1, \hat{X}_1 \rangle = -\text{tr}(\hat{X}_1 \theta(\hat{X}_1)) = 2. \quad (21.11.17)$$

Similarly, we can compute

$$\begin{aligned} \langle \hat{X}_2, \hat{X}_2 \rangle &= -\text{tr}(\hat{X}_2 \theta(\hat{X}_2)) = 2, \\ \langle \hat{X}_3, \hat{X}_3 \rangle &= -\text{tr}(\hat{X}_3 \theta(\hat{X}_3)) = 2, \\ \langle \hat{X}_i, \hat{X}_j \rangle &= -\text{tr}(\hat{X}_i \theta(\hat{X}_j)) = 0 \quad (i \neq j). \end{aligned} \quad (21.11.18)$$

The metric (21.11.16) is bi-invariant because for any  $g \in L_+^\uparrow$ ,

$$\begin{aligned} \langle \text{Ad}_g \hat{V}_i, \text{Ad}_g \hat{V}_j \rangle &= \langle g \hat{V}_i g^{-1}, g \hat{V}_j g^{-1} \rangle = -\text{tr}(g \hat{V}_i g^{-1} \theta(g \hat{V}_j g^{-1})) \\ &= -\text{tr}(g \hat{V}_i g^{-1}) \text{tr}(\theta(g \hat{V}_j g^{-1})) = -\text{tr}(\hat{V}_i) \text{tr}(-(g \hat{V}_j g^{-1})^T) \\ &= -\text{tr}(\hat{V}_i) (-\text{tr}(g \hat{V}_j g^{-1})^T) = -\text{tr}(\hat{V}_i) (-\text{tr}(g \hat{V}_j g^{-1})) \\ &= -\text{tr}(\hat{V}_i) (-\text{tr}(\hat{V}_j)) = -\text{tr}(\hat{V}_i) (-\text{tr}(\hat{V}_j^T)) = -\text{tr}(\hat{V}_i) (\text{tr}(-\hat{V}_j^T)) \\ &= -\text{tr}(\hat{V}_i) (\text{tr}(\hat{V}_j)) = -\text{tr}(\hat{V}_i \theta(\hat{V}_j)) = \langle \hat{V}_i, \hat{V}_j \rangle, \end{aligned}$$

according to Theorem 12.8.3, the left-invariant metric  $\langle \hat{V}_i, \hat{V}_j \rangle$  is bi-invariant.

Since a bi-invariant metric exists on the three-dimensional  $L_+^\uparrow$  group, by Theorem 12.8.4, the sectional curvature

$$K(\hat{V}_i, \hat{V}_j) > 0, \quad i \neq j. \quad (21.11.19)$$

Therefore, as a Lie group  $L_+^\uparrow$ , the elementary dark particle is curved and consequently possesses a gravitational field. Dark matter composed of elementary dark particles naturally also possesses a gravitational field.

### 5. Classification of Elementary Dark Particles

Although an elementary dark particle is a three-dimensional Lie group  $L_+^\uparrow$  and is a smooth manifold  $R^2 \times S^1$ , from the viewpoint of principal bundles, elementary dark particles can be divided into three classes.

Since in local coordinates the three-dimensional proper Lorentz group  $L_+^\uparrow$  can be expressed as the product of three matrices,

$$A_3 A_2 A_1 = \begin{pmatrix} \cosh \tau & 0 & -\sinh \tau \\ 0 & 1 & 0 \\ -\sinh \tau & 0 & \cosh \tau \end{pmatrix} \begin{pmatrix} \cosh \sigma & -\sinh \sigma & 0 \\ -\sinh \sigma & \cosh \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}, \quad (21.11.20)$$

analogous to the  $SO(3)$  group and  $SU(2)$  group, according to Theorem 12.13.4 the following three principal bundles exist:



**First**, the principal bundle  $(L_+^\uparrow, A_3A_2, \pi, G_{\hat{L}_1})$ , where the projection map is  $\pi: L_+^\uparrow \rightarrow A_3A_2$ , the base space  $A_3A_2$  is the product space of the matrices  $A_3$  and  $A_2$  in (21.11.20) (this base space is a plane  $R^2$ ), and  $G_{\hat{L}_1}$  denotes the one-dimensional Lie subgroup generated by the spin operator  $\hat{L}_1$ , called the fiber type or structure group. The elementary dark particle corresponding to this principal bundle is denoted as  $\chi_0$ , and its antiparticle as  $\bar{\chi}_0$ .

**Second**, the principal bundle  $(L_+^\uparrow, A_3A_1, \pi, G_{\hat{L}_2})$ , where the projection map is  $\pi: L_+^\uparrow \rightarrow A_3A_1$ , the base space  $A_3A_1$  is the product space of the matrices  $A_3$  and  $A_1$  in (21.11.20) (this base space is the product of a line and a circle, i.e., a two-dimensional cylinder, which is isometric to the plane  $R^2$ ), and  $G_{\hat{L}_2}$  denotes the one-dimensional Lie subgroup generated by the spin operator  $\hat{L}_2$ , called the fiber type or structure group. The elementary dark particle corresponding to this principal bundle is denoted as  $\chi_1$ , and its antiparticle as  $\bar{\chi}_1$ .

**Third**, the principal bundle  $(L_+^\uparrow, A_2A_1, \pi, G_{\hat{L}_3})$ , where the projection map is  $\pi: L_+^\uparrow \rightarrow A_2A_1$ , the base space  $A_2A_1$  is the product space of the matrices  $A_2$  and  $A_1$  in (21.11.20) (this base space is the product of a line and a circle, i.e., a two-dimensional cylinder isometric to the plane  $R^2$ ), and  $G_{\hat{L}_3}$  denotes the one-dimensional Lie subgroup generated by the spin operator  $\hat{L}_3$ , called the fiber type or structure group. The elementary dark particle corresponding to this principal bundle is denoted as  $\chi_2$ , and its antiparticle as  $\bar{\chi}_2$ .

## 6. Rest Mass and Short-Range Forces of Elementary Dark Particles

From the commutation relations (21.11.14) of the spin angular-momentum operators  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  of an elementary dark particle, we see that its Lie algebra is isomorphic to those of the proton and the photon. Therefore, according to Theorem 12.5.13, the local Lie groups (i.e., the neighborhoods of the identity element) of the elementary dark particle, the proton, and the photon are isomorphic. Consequently, their tangent spaces at the identity element are isomorphic. Moreover, because the expressions for the spin operators that generate the structure groups of the elementary dark particles  $\chi_1, \chi_2$  and their antiparticles are identical to those that generate the structure group of the proton, it follows from the definition of rest mass given in equation (20.8.7) that the rest masses of the elementary dark particles  $\chi_1, \chi_2$  and their antiparticles are all equal to the rest mass of the proton:

$$m_{\chi_1} = m_{\chi_1^-} = m_{\chi_2} = m_{\chi_2^-} = m_p,$$

even though the proton is compact while the dark particles  $\chi_1, \chi_2$  are non-compact. Similarly, just like the photon, according to the rest-mass definition (20.8.7), the rest masses of the elementary dark particle  $\chi_0$  and its antiparticle  $\bar{\chi}_0$  are both zero:

$$m_{\chi_0} = m_{\bar{\chi}_0} = 0.$$

The expression for the spin operator  $\hat{L}_1$  is identical to that of the photon's spin operator; hence the integral curves of the spin operator for the elementary dark particle  $\chi_0$  coincide with those of the photon's spin operator. As a result, **an attractive force** exists near its point-like particle. The expressions for the spin operators  $\hat{L}_2$  and  $\hat{L}_3$  are identical to that of the proton's spin operator; therefore, the integral curves of the spin operators for the elementary dark particles  $\chi_1$  and  $\chi_2$  coincide with those of the proton's spin operator. Consequently, **a repulsive force** exists near their point-like particles.

Although both the electron and the proton exhibit a repulsive force near their point-like particles—so that when these two point-like particles come sufficiently close they repel each other—if they are combined with an antineutrino, which produces an attractive force near its point-like particle, these three particles can form a neutron. Similarly, the elementary dark

particles  $\chi_1, \chi_2$  and their antiparticles generate a repulsive force near their point-like particles, making it impossible for them alone to form composite particles. However, when combined with the elementary dark anti-particle  $\bar{\chi}_0$ , which produces an attractive force near its point-like particle, they can coalesce into new particles. Here the dark anti-particle  $\bar{\chi}_0$  acts as a **binding agent**.

## 7. Dark Particles Do Not Generate Electromagnetic Fields

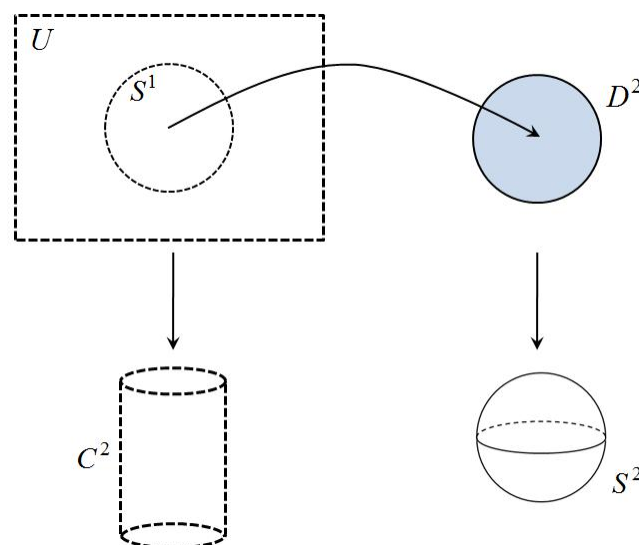
The base manifolds of the three types of elementary dark particles are either planes or cylinders, both of which are flat; hence the curvature (Gaussian curvature) of the base manifolds is zero. Let  $\Omega_\alpha$  be the curvature 2-form of the base manifold, then  $\Omega_\alpha = 0$ . Substituting  $\Omega_\alpha$  into equation (12.13.10) yields the curvature forms  $\tilde{\Omega} = 0$  of the three principal bundles  $(L_+^\uparrow, A_3 A_2, \pi, G_{\tilde{L}_1}), (L_+^\uparrow, A_3 A_1, \pi, G_{\tilde{L}_2})$  and  $(L_+^\uparrow, A_2 A_1, \pi, G_{\tilde{L}_3})$ ; consequently, the curvatures of these three principal bundles are all zero. Therefore, elementary dark particles carry no electric charge and thus do not participate in electromagnetic interactions. Dark matter composed of elementary dark particles also carries no charge and does not engage in electromagnetic interactions.

We refer to photons, spin-1/2 elementary particles, and the particles or matter formed from them as **visible particles** or **visible matter**. Below we analyze the relationship between these two classes of particles—the elementary visible particles and the elementary dark particles.

## 8. Similar Properties of Dark Particles and Their Antiparticles

Since the gravitational field is independent of orientation, a photon and its antiphoton produce identical gravitational fields. Furthermore, because neither photons nor their antiphotons carry electric charge, the electromagnetic fields they generate are also identical. Additionally, the spin and mass of a photon and its antiphoton are the same. Thus, many important properties of a photon and its antiphoton are identical. Consequently, we often regard a photon and its antiphoton as the same particle. Similarly, because the gravitational field is independent of orientation, a dark particle and its antiparticle produce identical gravitational fields. Moreover, dark particles and their antiparticles generate no electromagnetic fields. Furthermore, the spin and mass of a dark particle and its antiparticle are also identical. Hence, many important properties of a dark particle and its antiparticle are the same. Therefore, it is natural for us to **regard a dark particle and its antiparticle as the same particle**.

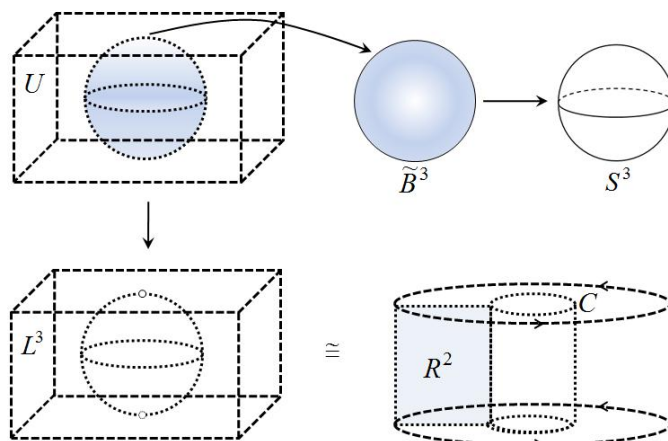
## 9. Conception and Birth of Particles



**Figure 21.11.2** Generating the sphere  $S^2$  and the cylinder  $C^2$  from a plane

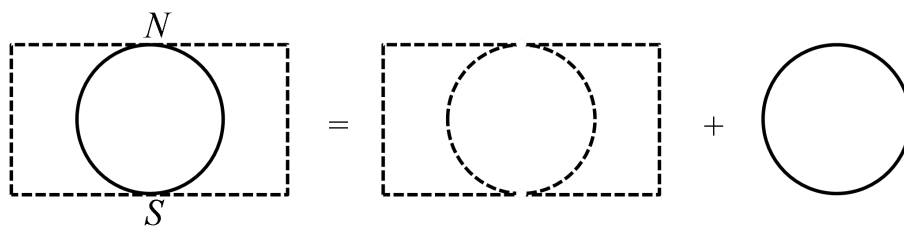
As shown in Figure 21.11.2, take an open set  $U$  in the plane  $R^2$ . Clearly,  $U$  is smoothly

homeomorphic to  $R^2$ . If a disk  $D^2$  is removed from  $U$  and the boundary circle  $S^1$  of  $D^2$  is **pinched to a point**, we obtain the two-sphere  $S^2$ ; that is,  $S^2$  is the quotient space  $D^2/S^1$  (with the boundary circle collapsed to a point), denoted as  $S^2 = D^2/S^1$ . If every pair of antipodal points on the boundary  $S^1$  of the disk  $D^2$  is **identified to a single point**, we obtain the real projective plane  $RP^2$ . After removing the disk  $D^2$ , the open set  $U$  is homeomorphic to a cylindrical surface  $C^2$  (i.e.,  $R \times S^1$ ).



**Figure 21.11.3** A three-dimensional open set  $U$  can be decomposed into a solid closed ball and a hollow solid cylinder

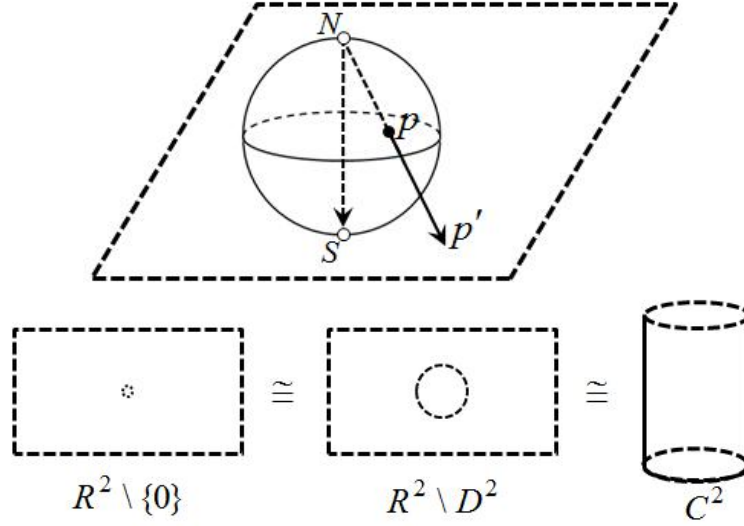
Similarly, as shown in Figure 21.11.3, take an open set  $U$  (e.g., an open solid rectangular box) in the three-dimensional space  $R^3$ . Clearly,  $U$  is smoothly homeomorphic to  $R^3$ . If a solid closed ball  $\tilde{B}^3$  is removed from  $U$  and the boundary sphere  $S^2$  of  $\tilde{B}^3$  is **pinched to a point**, we obtain the three-sphere  $S^3$ ; that is,  $S^3$  is the quotient space  $\tilde{B}^3/S^2$  (with the boundary sphere collapsed to a point), denoted as  $S^3 = \tilde{B}^3/S^2$ . If every pair of antipodal points on the boundary  $S^2$  of the solid ball  $\tilde{B}^3$  is **identified to a single point**, we obtain the real projective space  $RP^3$ . After removing a solid closed ball  $\tilde{B}^3$  from the open set  $U$ , the remaining set is denoted as  $L^3$ ;  $L^3$  is homeomorphic to a hollow solid cylinder (i.e.,  $R^2 \times S^1$ ), and  $L^3 \cong R^2 \times S^1$ .



**Figure 21.11.4** Cross-section through the north and south poles

Figure 21.11.4 is a cross-section through the north and south poles of the solid ball  $\tilde{B}^3$  inside the open set  $U$  of Figure 21.11.3. Although the open set  $U$ , as a manifold, has no boundary, as a topological space it does possess a boundary—consisting of six faces. Regarding  $U$  as a topological space, its boundary exists. The top and bottom faces of  $U$  are tangent respectively to the north pole  $N$  and the south pole  $S$  of the solid ball. When the solid ball is removed from  $U$ , a point is thereby deleted from each of the top and bottom faces. The open set  $U$  after removal of the solid ball,  $U \setminus \tilde{B}^3$ , is homeomorphic to a hollow solid cylinder. Figure 21.11.5 helps illustrate why this homeomorphism holds. The inner boundary of  $U \setminus \tilde{B}^3$  is a two-sphere with its north and south poles removed, denoted  $S^2 \setminus \{N, S\}$ . This sphere  $S^2 \setminus \{N, S\}$  is homeomorphic to a plane

with a point removed,  $R^2 \setminus \{0\}$  (see Theorem 1.6.1), and  $R^2 \setminus \{0\}$  is homeomorphic to a plane with a disk removed,  $R^2 \setminus D^2$ . Since  $R^2 \setminus D^2$  is homeomorphic to the cylinder  $C^2$  (extending infinitely in both directions, i.e., without top or bottom boundaries), transitivity of homeomorphism implies that  $S^2 \setminus \{N, S\}$  is homeomorphic to  $C^2$ .



**Figure 21.11.5** A sphere with its north and south poles removed is homeomorphic to a cylindrical surface

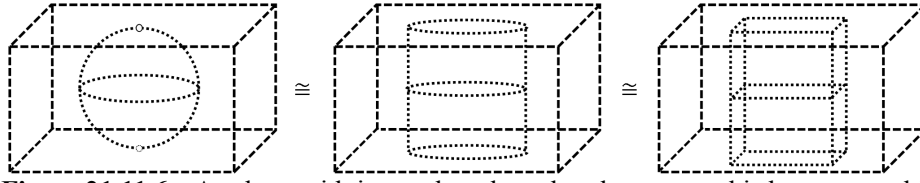
If the boundary sphere  $S^2$  of the solid ball  $\tilde{B}^3$  is pinched to a point, the solid ball  $\tilde{B}^3$  becomes a three-sphere  $S^3$ . However, within the space  $R^3$  it is impossible to pinch the boundary sphere  $S^2$  to a point; that is,  $S^3$  cannot be embedded in  $R^3$ . Nevertheless, if the boundary sphere  $S^2$  expands with time as illustrated in Figure 15.5.1, following the expansion law given by equation (15.5.11), then the solid ball  $\tilde{B}^3$  with its interior region **removed** (i.e., excluded) can be regarded as belonging to one of the spin-1/2 elementary particles. Similarly, if every pair of antipodal points on the boundary sphere  $S^2$  of the solid ball  $\tilde{B}^3$  are identified, the solid ball  $\tilde{B}^3$  becomes the real projective space  $RP^3$ . Yet, within  $R^3$  it is also impossible to identify antipodal points on  $S^2$  to a single point; that is,  $RP^3$  cannot be embedded in  $R^3$ . However, if the radius of the boundary sphere  $S^2$  expands according to equation (15.5.10), then the solid ball  $\tilde{B}^3$  with its **center point removed** can be regarded as a photon.

Based on the analysis above, any open set  $U$  in the space  $R^3$  can, in the manner described, give rise to a photon and an elementary dark particle, or to a spin-1/2 elementary particle and an elementary dark particle.

Conversely, it can also be stated that whenever a visible particle is created, there always exists an open set  $U$  that embeds this particle in the manner shown in the upper-left diagram of Figure 21.11.3. The part of this open set  $U$  after removing the visible particle precisely corresponds to an elementary dark particle. Therefore, **the creation of every visible particle is accompanied by the corresponding creation of an elementary dark particle.**

Although the open set  $U$  has only two point gaps (the removed poles), there is no need to worry that the particles inside cannot escape. This is because the sphere with its north and south poles removed,  $S^2 \setminus \{N, S\}$ , is homeomorphic to an infinite cylinder  $C^2$  (which extends without top or bottom boundaries) and also homeomorphic to a hollow rectangular box without upper or lower boundaries, as illustrated in Figure 21.11.6. Consequently,  $S^2 \setminus \{N, S\}$  can be continuously deformed into a cylinder  $C^2$ , and likewise into a hollow rectangular box. Thus, particles inside the open set  $U$  can indeed escape. This continuous deformation process is analogous to the birth of

an infant: the contraction of the uterus corresponds to the contraction of the central part of the sphere  $S^2$ , while the dilation of the cervical opening corresponds to the expansion of the south and north poles.



**Figure 21.11.6** A sphere with its north and south poles removed is homeomorphic to a hollow cylinder and to a hollow rectangular box

## 10. Obeying Conservation Laws

Since spacetime  $M^4$  is left-handed oriented, according to §21.10 the open set  $U$  is also left-handed oriented. By Theorem 9.4.5, the manifolds  $B^3$  and  $R^2 \times S^1$  obtained by decomposing  $U$  are likewise left-handed oriented. Consequently, in general, the photon, the spin-1/2 elementary particle, and the elementary dark particle produced from the decomposition of  $U$  are all left-handed oriented particles—i.e., they are **positive particles**. Exceptions can occur, however, because the decomposition process of  $U$  must **obey the conservation laws**.

The decomposition of one open set  $U$  into an elementary dark particle and a photon obeys **charge conservation**, because neither particle carries electric charge. Assuming the total angular momentum before decomposition is zero, and after decomposition the two particles move in opposite directions, if the photon's spin- $z$  component is  $\hbar$ , then the elementary dark particle's spin- $z$  component must be  $-\hbar$ ; thus the total angular momentum after decomposition is also zero. Therefore, this process can also satisfy **angular-momentum conservation**. Of course, it must also satisfy **energy conservation** and **momentum conservation** to occur.

The decomposition of one open set  $U$  into an elementary dark particle and a proton would violate both charge conservation and angular-momentum conservation. Hence **two open sets must decompose simultaneously**. If one open set  $U$  decomposes into an elementary dark particle and a proton, and another open set  $U'$  decomposes into another elementary dark particle and an electron, then only when these two processes occur together can charge conservation and angular-momentum conservation be satisfied. The proton carries one unit of positive charge, the electron carries one unit of negative charge, and their sum is exactly zero—charge conservation is fulfilled. Assume the total angular momentum of the two open sets before decomposition is zero. Let the decomposition of  $U$  yield an elementary dark particle moving left with spin- $z$  component  $-\hbar$  and a proton moving right with spin- $z$  component  $\frac{1}{2}\hbar$ . Let the decomposition of  $U'$  yield an elementary dark particle moving right with spin- $z$  component  $\hbar$  and an electron moving left with spin- $z$  component  $-\frac{1}{2}\hbar$ . Then the total angular momentum before and after decomposition remains zero; hence when the two decompositions occur simultaneously, **angular-momentum conservation** is obeyed.

The decomposition of one open set  $U$  into an elementary dark particle and a neutrino obeys charge conservation but violates angular-momentum conservation; therefore, again **two open sets must decompose together**. Assume the total angular momentum of the two open sets before decomposition is zero. One open set  $U$  decomposes into an elementary dark particle and a neutrino; the other open set  $U'$  initially also decomposes into an elementary dark particle and a neutrino, but when the two particles separate they must both reverse their orientations, becoming an elementary anti-dark particle and an antineutrino, respectively. Only when these two processes occur together can angular-momentum conservation be satisfied. Let the decomposition of  $U$  give an elementary dark particle moving right with spin- $z$  component  $\hbar$  and a neutrino moving left with spin- $z$  component  $-\frac{1}{2}\hbar$ . Let the decomposition of  $U'$  give an elementary anti-dark particle moving left with spin- $z$  component  $-\hbar$  and an antineutrino moving right with spin- $z$

component  $\frac{1}{2}\hbar$ . The total angular momentum before and after decomposition is zero; thus when the two decompositions occur simultaneously, **angular-momentum conservation** is obeyed.

Let us examine in more detail how conservation laws can force a particle to transform into its antiparticle. Suppose before a transformation there are two vectors  $A_1$  and  $B_1$ , and after the transformation they become  $A_2$  and  $B_2$ . Conservation requires

$$A_1 + B_1 = A_2 + B_2,$$

or

$$B_2 - B_1 = -(A_2 - A_1).$$

Set

$$\Delta A = A_2 - A_1, \quad \Delta B = B_2 - B_1,$$

then we obtain

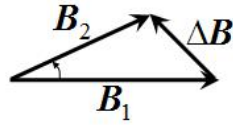
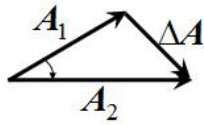
$$\Delta B = -\Delta A.$$

Thus, if before the transformation  $\Delta A = \Delta B$ , then to satisfy conservation, after the transformation  $\Delta A$  and  $\Delta B$  must be **opposite vectors**. Suppose  $\Delta B$  is the opposite of  $\Delta A$ , i.e.,  $\Delta B = -\Delta A$ ; this is equivalent to multiplying  $\Delta A$  by  $-1$ . If  $\Delta A$  is two-dimensional, we may regard  $-1$  as the  $2 \times 2$  determinant

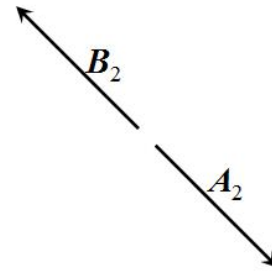
$$\begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix}, \quad (21.11.21)$$

so multiplying by  $-1$  means multiplying by this determinant, i.e.,

$$-\Delta A = \begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix} \Delta A.$$



**Figure 21.11.7** Conservation requires that the two planes have opposite orientations



**Figure 21.11.8** Conservation requires  $A_2$  and  $B_2$  to be parallel but opposite in direction

As shown in Figure 21.11.7, because after the transformation  $\Delta B = -\Delta A$ , the vectors  $\Delta A$  and  $\Delta B$  are parallel but opposite. Transforming vector  $A_1$  into  $A_2$  requires a clockwise rotation, while transforming vector  $B_1$  into  $B_2$  requires a counterclockwise rotation, in order to satisfy conservation. This is equivalent to saying that conservation demands the orientation of the plane (or space) containing  $A_1$  and  $A_2$  to be **opposite** to that of the plane (or space) containing  $B_1$  and  $B_2$ . If  $A_1$  and  $A_2$  are vectors of a particle  $G$ , and  $B_1$  and  $B_2$  are vectors of the antiparticle  $\bar{G}$  of  $G$ , then meeting this requirement means satisfying the conservation laws. Equivalently, there exists a diffeomorphism  $f: G \rightarrow \bar{G}$  whose Jacobian matrix has determinant  $|J| < 0$ . Like (21.11.21), because the determinant of  $f$  is  $|J| < 0$ , the mapping  $f$  reverses the orientation of particle  $G$ , turning particle  $G$  into antiparticle  $\bar{G}$ . This forces  $\Delta B$  to become the opposite of  $\Delta A$ , i.e.,  $\Delta A$  is multiplied by the determinant (21.11.21).

If  $A_1 = B_1 = 0$ , then

$$\Delta A = A_2, \quad \Delta B = B_2.$$

Since  $\Delta B = -\Delta A$ , as illustrated in Figure 21.11.8,



$$\mathbf{B}_2 = -\mathbf{A}_2.$$

If  $|\mathbf{A}_2| = -\frac{1}{2}\hbar$  is the spin- $z$  component of a neutrino, then  $|\mathbf{B}_2| = -|\mathbf{A}_2| = \frac{1}{2}\hbar$  is the spin- $z$  component of an antineutrino. Therefore, in the particle-creation process described above, producing two neutrinos simultaneously would violate angular-momentum conservation. Hence, angular-momentum conservation generates a mapping  $f$  that forces **one neutrino to transform into an antineutrino**.

Consequently, if only an elementary dark particle  $\chi_0$  and a photon are produced by the decomposition of an open set in  $R^3$ , there is no need for **paired** open sets to decompose simultaneously. However, if the decomposition is to yield elementary dark particles  $\chi_1, \chi_2$  together with electrons, protons, and neutrinos, then **pairwise simultaneous decomposition of open sets in  $R^3$  is mandatory** in order to satisfy both charge conservation and angular-momentum conservation. Naturally, the decomposition must also obey energy conservation and momentum conservation.

Based on the above analysis, when elementary visible particles and elementary dark particles are born, **positive (left-handed oriented) particles dominate**. This explains why the observable universe is **matter-dominated**, with antimatter being extremely scarce.

### 11. The Abundance of Dark Matter in the Universe

The production of a spin-1/2 elementary visible particle corresponds to the creation of an  $SU(2)$  group. The  $SU(2)$  group can serve as one of three principal bundles. If it serves as the principal bundle  $(SU(2), BC, \pi, G_{\hat{S}_1})$ , the corresponding particle is a proton. If it serves as the principal bundle  $(SU(2), A'C', \pi, G_{\hat{S}_2})$ , the corresponding particle is a neutrino. If it serves as the principal bundle  $(SU(2), AB, \pi, G_{\hat{S}_3})$ , the corresponding particle is an electron. We should assume that **which principal bundle the  $SU(2)$  group becomes is equally probable and random**. If this were not the case—for instance, if the probability of becoming  $(SU(2), BC, \pi, G_{\hat{S}_1})$  were highest—it would lead to an excess of protons and a shortage of electrons, causing the universe not to be electrically neutral. There are three types of neutrinos  $\nu_e, \nu_\mu, \nu_\tau$ ; we also assume that **which type of neutrino the principal bundle represents is equally probable and random**.

The production of a photon corresponds to the creation of an  $SO(3)$  group. The  $SO(3)$  group can likewise serve as one of three principal bundles. According to the different principal bundles, we classify photons into three types: principal-bundle  $(SO(3), BC, \pi, G_{\hat{L}_1})$  photon  $\gamma_1$ , principal-bundle  $(SO(3), AC, \pi, G_{\hat{L}_2})$  photon  $\gamma_2$ , and principal-bundle  $(SO(3), AB, \pi, G_{\hat{L}_3})$  photon  $\gamma_3$ . We should also assume that **which principal bundle the  $SO(3)$  group becomes is equally probable and random**.

Let the events of producing the three types of photons be denoted by  $E_1, E_2, E_3$ ; the event of producing a proton by  $E_4$ ; the event of producing an electron by  $E_5$ ; and the events of producing neutrinos  $\nu_e, \nu_\mu, \nu_\tau$  by  $E_6, E_7, E_8$ , respectively. At least one of these eight events must occur, forming a **complete set of events  $E$** :

$$E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_7 + E_8 = E.$$

The set  $E$  contains all elementary events; therefore, the probability that the event set  $E$  occurs equals 1, i.e.,  $p(E) = 1$ .

The production of an elementary dark particle corresponds to the creation of an  $L_+^\uparrow$  group. The  $L_+^\uparrow$  group can also serve as one of three principal bundles. If it serves as the principal bundle  $(L_+^\uparrow, A_3A_2, \pi, G_{\hat{L}_1})$ , the corresponding elementary dark particle is  $\chi_0$ . If it serves as the principal bundle  $(L_+^\uparrow, A_3A_1, \pi, G_{\hat{L}_2})$ , the corresponding elementary dark particle is  $\chi_1$ . If it serves as the

principal bundle  $(L_+^\uparrow, A_2 A_1, \pi, G_{\hat{L}_3})$ , the corresponding elementary dark particle is  $\chi_2$ . We again assume that **which principal bundle the  $L_+^\uparrow$  group becomes is equally probable and random.**

Let the events of producing the elementary dark particles  $\chi_0, \chi_1, \chi_2$  be denoted by  $F_0, F_1, F_2$ , respectively. At least one of these three events must occur, forming a complete set of events  $F$ :

$$F_0 + F_1 + F_2 = F.$$

The probability that the event set  $F$  occurs is also 1, i.e.,  $p(F) = 1$ .

$$E = \{\gamma_1, \gamma_2, \gamma_3; p, e^-, \nu_e, \nu_\mu, \nu_\tau\} \Leftrightarrow \{\chi_0, \chi_1, \chi_2\} = F$$

**Figure 21.11.7** Correspondence between the complete event set  $E$  and the complete event set  $F$

As shown in Figure 21.11.7, the number of events in the complete set  $E$  is 8, while that in  $F$  is 3. The numbers of events in the two sets are not equal. Because the least common multiple of 8 and 3 is 24, only **three** complete sets  $E$  (total events = 24) and **eight** complete sets  $F$  (total events = 24) can be put into a one-to-one correspondence. More generally, let  $n$  be an integer; then the production of  $8 \times 3n$  elementary visible particles will be accompanied by the production of  $3 \times 8n$  elementary dark particles. If  $n = 2$ , then 48 elementary visible particles are produced together with 48 elementary dark particles. Among these 48 elementary visible particles, there are 6 neutrinos  $\nu_e$ , of which 3 neutrinos  $\nu_e$  transform into antineutrinos  $\bar{\nu}_e$ . Similarly, there are 6 neutrinos  $\nu_\mu$ , of which 3 transform into antineutrinos  $\bar{\nu}_\mu$ ; and 6 neutrinos  $\nu_\tau$ , of which 3 transform into antineutrinos  $\bar{\nu}_\tau$ . Suppose the universe contained only these 48 elementary visible particles, and that during the formation of atoms, 3 protons, 3 electrons, and 3 antineutrinos combine to form 3 neutrons; these 3 neutrons then combine with another 3 protons to form 3 deuterium nuclei; and the 3 deuterium nuclei combine with another 3 electrons to form 3 deuterium atoms. Alternatively, one could suppose that 2 protons and 2 electrons form 2 hydrogen molecules; 2 protons, 2 electrons, and 2 antineutrinos form 2 neutrons; these 2 neutrons then combine with another 2 protons to form 2 helium nuclei; and these 2 helium nuclei combine with 2 electrons to form 2 helium molecules. Clearly, after the elementary visible particles form atoms, very few redundant elementary visible particles remain. A similar analysis can be performed for other even values of  $n > 2$ .

Let the rest mass of a proton be  $k$  times the rest mass of an electron, i.e.,

$$m_p = km_e.$$

Recently, German scientists measured  $k=1836.152673414(47)$ . In Chapter 20, we calculated  $k=1836.14635571845$ .

Since the rest masses of dark particles  $\chi_0$  and photons are both zero, and the rest masses of dark particles  $\chi_1, \chi_2$  equal the proton rest mass, i.e.,

$$m_{\chi_1} = m_{\chi_2} = m_p,$$

in the collection consisting of 6 ( $= 3n = 3 \times 2$ ) complete event sets  $E$  and 16 ( $= 8n = 8 \times 2$ ) complete event sets  $F$ , the total rest mass  $m_{light}$  of visible matter is

$$m_{light} = 6m_p + 6m_e,$$

and the total rest mass  $m_{all}$  is

$$m_{all} = 16m_{\chi_1} + 16m_{\chi_2} + 6m_p + 6m_e = 38m_p + 6m_e.$$

The total mass of the universe is some multiple of  $m_{all}$ , say  $n$  times (clearly  $n$  is an extremely large natural number). Similarly, the total rest mass of visible matter in the universe is  $n$  times  $m_{light}$ . From this we can compute the fraction of visible matter in the total mass of the universe:

$$\lambda_{visible} = \frac{(6m_p + 6m_e) \times n}{(38m_p + 6m_e) \times n} = \frac{3m_p + 3m_e}{19m_p + 3m_e} = \frac{3km_e + 3m_e}{19km_e + 3m_e} = \frac{3k + 3}{19k + 3}.$$



Substituting the recent German scientists' measured value  $k=1836.152673414(47)$  yields

$$\lambda_{\text{visible, ger}} = \frac{3k+3}{19k+3} = \frac{3 \times 1836.152673414 + 3}{19 \times 1836.152673414 + 3} = 0.157967145067177,$$

and the fraction of dark matter in the total mass of the universe:

$$\lambda_{\text{dark, ger}} = 1 - \lambda_{\text{visible, ger}} = 1 - 0.157967145067177 = 0.842032854932823.$$

If we substitute the value  $k=1836.14635571845$  calculated in Chapter 20, we obtain

$$\lambda_{\text{visible, we}} = \frac{3k+3}{19k+3} = \frac{3 \times 1836.14635571845 + 3}{19 \times 1836.14635571845 + 3} = 0.157967145316293.$$

Since 1989, humanity has launched the COBE, WMAP, and Planck satellites to probe the composition of the universe. The 2014 *Review of Particle Physics* reported the observed fractions: dark energy =  $(68.5 \pm 1.7 - 1.6)\%$ , ordinary baryonic matter =  $(4.99 \pm 0.22)\%$ . Ordinary baryonic matter, i.e., normal matter, “refers to atoms and their constituents, including leptons. However, the observed baryonic matter in the form of luminous galaxies and diffuse gas accounts for less than 1%. Where the majority of baryonic matter resides remained a mystery for a long time. Recent X-ray observations, however, have found that this ordinary matter exists as tenuous intergalactic hydrogen atoms, about 6 atoms per cubic meter. In contrast, interstellar space contains about one million atoms per cubic meter.”<sup>7</sup> We all know that when protons and neutrons combine to form atomic nuclei, a mass defect occurs, but this defect is relatively small and can be neglected in our approximation. Likewise, we neglect the mass defect that may arise when two or more dark particles combine to form a new dark particle. Moreover, since ordinary matter in the universe exists mainly in the form of hydrogen atoms, and dark particles lack electromagnetic interactions, in the above calculation we have approximated all particles in the universe as free particles.

From the latest observational results, the fraction of dark matter in the total mass-energy of the universe is calculated as

$$100\% - 4.99\% - 68.5\% = 26.51\%.$$

If we set aside dark energy (we do not need the dark-energy hypothesis to explain cosmic expansion), the experimentally measured fraction of total visible matter in the total mass is

$$\lambda'_{\text{visible}} = \frac{0.0499}{0.0499 + 0.2651} = \frac{0.0499}{0.315} = 0.158412698412698,$$

and the fraction of total dark matter in the total mass is

$$\lambda'_{\text{dark}} = \frac{0.2651}{0.0499 + 0.2651} = \frac{0.2651}{0.315} = 0.841587301587302.$$

The difference between the measured value  $\lambda'_{\text{visible}}$  and the calculated value  $\lambda_{\text{visible, ger}}$  is

$$\lambda'_{\text{visible}} - \lambda_{\text{visible, ger}} = 0.158412698412698 - 0.157967145067177 = 0.000445553345521321.$$

The error is

$$\text{error}_{\text{ger}} = \left( \frac{\lambda'_{\text{visible}} - \lambda_{\text{visible, ger}}}{\lambda'_{\text{visible}}} \right) \times 100\% = 0.28126112993831\%.$$

The calculated and measured values agree quite well.

The difference between the measured value  $\lambda'_{\text{visible}}$  and the calculated value  $\lambda_{\text{visible, we}}$  is

$$\lambda'_{\text{visible}} - \lambda_{\text{visible, we}} = 0.158412698412698 - 0.157967145316293 = 0.000445553096405149.$$

The error is

$$\text{error}_{\text{we}} = \left( \frac{\lambda'_{\text{visible}} - \lambda_{\text{visible, we}}}{\lambda'_{\text{visible}}} \right) \times 100\% = 0.281260972680606\%.$$

<sup>7</sup>Yang Binglin, translated by Liu Guoli, Wang Wenyu, Wang Fei. *Dark Matter and Related Cosmology* [M]. Beijing: Science Press, 2019:5.

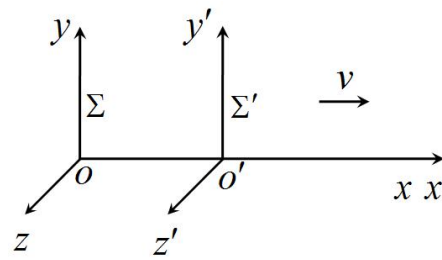
The error  $error_{we}$  is smaller than  $error_{ger}$ , indicating that our calculated proton-to-electron rest mass ratio is more accurate than the experimentally measured ratio.

## 12.Reasons Why the Lorentz Transformation is More Accurate

Extensive experiments have proven that the Galilean transformation is only suitable for objects moving at low speeds. For objects moving at high speeds, the Lorentz transformation must be used, as the Galilean transformation is merely an approximation of the Lorentz transformation at low speeds. This section elaborates on the reasons why the Lorentz transformation is more accurate.

As shown in Figure 21.11.9, consider two coordinate systems,  $\Sigma$  and  $\Sigma'$ . The coordinates of a point  $p$  in spacetime are represented as  $(x, y, z, t)$  in the  $\Sigma$  system and  $(x', y', z', t')$  in the  $\Sigma'$  system. The directions of the three spatial coordinate axes are identical for both systems. Assume that initially, their origins  $O$  and  $O'$  coincide, at which time their coordinates are both  $(0,0,0,0)$ . Furthermore, let the  $\Sigma'$  system move with a constant velocity  $v$  along the positive  $x$ -axis of the  $\Sigma$  system. Under these assumptions, the Galilean transformation is

$$\begin{cases} x = x' + vt, \\ y = y', \\ z = z', \\ t = t'. \end{cases} \quad (21.11.22)$$



**Figure 21.11.9** The  $\Sigma'$  system moves to the right with velocity  $v$ .

And the Lorentz transformation is

$$\begin{cases} x' = \gamma(x - vt), \\ y' = y, \\ z' = z, \\ t' = \gamma\left(t - \frac{v}{c^2}x\right); \end{cases} \quad (21.11.23)$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (21.11.24)$$

Equation (21.11.23) can be rewritten as

$$\begin{cases} x' = \gamma\left(x - \frac{v}{c}ct\right), \\ y' = y, \\ z' = z, \\ ct' = \gamma\left(ct - \frac{v}{c}x\right). \end{cases} \quad (21.11.25)$$

Omitting the  $z$  and  $z'$  axes, write equation (21.11.25) in the following matrix form:

$$\begin{pmatrix} x' \\ y' \\ ct' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & -\gamma \frac{v}{c} \\ 0 & 1 & 0 \\ -\gamma \frac{v}{c} & 0 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \\ ct \end{pmatrix}. \quad (21.11.26)$$

Let  $v = c \tanh \tau$ , then since  $-1 \leq \tanh \tau \leq 1$ , we have  $-c \leq v \leq c$ . Substitute  $v = c \tanh \tau$  into equation (21.11.24) and perform a transformation on  $\gamma$ :

$$\begin{aligned} \gamma &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - \frac{c^2 \tanh^2 \tau}{c^2}}} = \frac{1}{\sqrt{1 - \tanh^2 \tau}} \\ &= \frac{1}{\sqrt{1 - \frac{\sinh^2 \tau}{\cosh^2 \tau}}} = \frac{1}{\sqrt{\frac{\cosh^2 \tau - \sinh^2 \tau}{\cosh^2 \tau}}} \\ &= \frac{1}{\sqrt{\frac{1}{\cosh^2 \tau}}} = \cosh \tau. \end{aligned}$$

Consequently, we obtain

$$-\gamma \frac{v}{c} = -\cosh \tau \frac{c \tanh \tau}{c} = -\cosh \tau \frac{\sinh \tau}{\cosh \tau} = -\sinh \tau.$$

Since the matrix

$$\begin{pmatrix} \gamma & 0 & -\gamma \frac{v}{c} \\ 0 & 1 & 0 \\ -\gamma \frac{v}{c} & 0 & \gamma \end{pmatrix} = \begin{pmatrix} \cosh \sigma & 0 & -\sinh \sigma \\ 0 & 1 & 0 \\ -\sinh \sigma & 0 & \cosh \sigma \end{pmatrix},$$

equation (21.11.26) can be transformed into

$$\begin{pmatrix} x' \\ y' \\ ct' \end{pmatrix} = \begin{pmatrix} \cosh \sigma & 0 & -\sinh \sigma \\ 0 & 1 & 0 \\ -\sinh \sigma & 0 & \cosh \sigma \end{pmatrix} \begin{pmatrix} x \\ y \\ ct \end{pmatrix}. \quad (21.11.27)$$

Let

$$A_3 = \begin{pmatrix} \cosh \tau & 0 & -\sinh \tau \\ 0 & 1 & 0 \\ -\sinh \tau & 0 & \cosh \tau \end{pmatrix}. \quad (21.11.28)$$

The matrix  $A_3$  is identical to the matrix  $A_3$  in equation (21.11.3). It is evident that the transformation matrix of the Lorentz transformation (21.11.26) is an element of the proper Lorentz group  $L_+^\uparrow$ . The transformation equation (21.11.27) keeps  $-c^2 t^2 + x^2 + y^2$  ( $t, x, y \in R$ ) invariant, i.e.,

$$-c^2 t^2 + x^2 + y^2 = -c^2 t'^2 + x'^2 + y'^2.$$

Matrix (21.11.28) is a real  $3 \times 3$  matrix. We can also expand it into the following real  $4 \times 4$  matrix:

$$A'_3 = \begin{pmatrix} \cosh \tau & 0 & 0 & -\sinh \tau \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh \tau & 0 & 0 & \cosh \tau \end{pmatrix}. \quad (21.11.29)$$

The matrix  $A'_3$  is precisely the transformation matrix for equation (21.11.25).

The matrix  $A_3$  actually represents a one-parameter subgroup of the three-dimensional proper Lorentz group  $L_+^\uparrow$  and can act as a Lie transformation group from the left on particles

such as photons, electrons, protons, neutrinos and their antiparticles, as well as composite particles formed from these. For example, let the  $SU(2)$  group be the Lie group of the electron. Then we obtain the mapping:

$$f : A_3 \times SU(2) \rightarrow SU(2). \quad (21.11.30)$$

Suppose the coordinates of a point-like electron are  $p = (x, y, ct) \in SU(2)$ . Then mapping (21.11.30) can be written as

$$\begin{pmatrix} x' \\ y' \\ ct' \end{pmatrix} = \begin{pmatrix} \cosh \sigma & 0 & -\sinh \sigma \\ 0 & 1 & 0 \\ -\sinh \sigma & 0 & \cosh \sigma \end{pmatrix} \begin{pmatrix} x \\ y \\ ct \end{pmatrix}, \quad (21.11.31)$$

where  $p' = (x', y', ct') \in SU(2)$  are the new coordinates after transforming the old coordinates  $p = (x, y, ct)$  of the point-like electron. Equation (21.11.31) is essentially equation (21.11.27). Of course, we can replace the transformation matrix in equation (21.11.31) with the matrix  $A'_3$ , yielding:

$$\begin{pmatrix} x' \\ y' \\ z' \\ ct' \end{pmatrix} = \begin{pmatrix} \cosh \tau & 0 & 0 & -\sinh \tau \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh \tau & 0 & 0 & \cosh \tau \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ ct \end{pmatrix}. \quad (21.11.32)$$

Now we finally understand why the Lorentz transformation is more accurate than the Galilean transformation. It turns out that dark particles, which are a three-dimensional proper Lorentz group  $L_+^\uparrow$ , permeate our universe everywhere. Within this group, there exists a one-parameter subgroup whose elements can be represented by the matrix  $A_3$ . This subgroup  $A_3$ , acting as a transformation group, operates on visible particles such as photons, electrons, protons, neutrinos and their antiparticles, as well as composite particles or matter formed from these. In other words, these visible particles or matter are all subject to the transformative action of dark particles. Extensive experimental evidence confirming the correctness of the Lorentz transformation simultaneously proves the existence of dark particles and validates the interpretation of dark particles as the three-dimensional proper Lorentz group  $L_+^\uparrow$ .

## References

### Mathematics Section:

- [1] Department of Mathematics and Mechanics, Peking University. *Advanced Algebra* [M]. Beijing: Higher Education Press, 1978.
- [2] Bai Zhengguo, Shen Yibing, Shui Naixiang, Guo Xiaoying. *Preliminary Riemannian Geometry (Revised Edition)* [M]. Beijing: Higher Education Press, 2004.
- [3] Chern S. S., Chen Weihuan. *Lectures on Differential Geometry* [M]. Beijing: Peking University Press, 1983.
- [4] Chen Weihuan. *Introduction to Differentiable Manifolds* [M]. Beijing: Higher Education Press, 2001.
- [5] Chen Weihuan, Li Xingxiao. *Introduction to Riemannian Geometry (Volumes I & II)* [M]. Beijing: Peking University Press, 2002.
- [6] Hou Boyuan, Hou Boyu. *Differential Geometry for Physicists (Second Edition)* [M]. Beijing: Science Press, 2004.
- [7] Liang Canbin, Zhou Bin. *Differential Geometry and General Relativity: An Introduction (Volumes I, II & III)* [M]. Beijing: Science Press, 2006.
- [8] Song Xiangnuan. The Physical Meaning of Complex Variable Function Integrals [J]. Tianjin: Journal of Tianjin University of Commerce, 1996, (3): 41–42.
- [9] Su Buqing, Liu Dingyuan. *Elementary Differential Geometry* [M]. Shanghai: Shanghai Science and Technology Press, 1985.
- [10] Sun Zhiming. *Tensors in Physics* [M]. Beijing: Beijing Normal University Press, 1985.
- [11] [UK] B. F. Schutz, translated by Feng Chengtian, Li Shunqi. *Geometrical Methods of Mathematical Physics* [M]. Shanghai: Shanghai Scientific and Technological Literature Publishing House, 1986.
- [12] Tao Ruibao. *Group Theory in Physics (Volume I)* [M]. Shanghai: Shanghai Science and Technology Press, 1986.
- [13] Wang Gaoxiong, Zhou Zhiming, Zhu Siming, Wang Shousong. *Ordinary Differential Equations* [M]. Beijing: Higher Education Press, 1983.
- [14] Shoshichi Kobayashi, Katsumi Nomizu, translated by Xie Kongbin, Chen Yuzhuo, Xie Yunpeng. *Foundations of Differential Geometry* [M]. Beijing: Science Press, 2010.
- [15] Xu Senlin, Hu Zisheng, Xue Chunhua. *Differential Topology* [M]. Beijing: Tsinghua University Press, 2008.
- [16] I. M. Singer, J. A. Thorpe, translated by Gan Danyan. *Lecture Notes on Elementary Topology and Geometry* [M]. Shanghai: Shanghai Science and Technology Press, 1985.
- [17] You Chengye. *Lectures on Basic Topology* [M]. Beijing: Peking University Press, 1997.
- [18] John M. Lee. *Introduction to Smooth Manifolds, Second Edition* [M]. Beijing: World Publishing Corporation, 2015.
- [19] Loring W. Tu. *An Introduction to Manifolds, Second Edition* [M]. Beijing: World Publishing Corporation, 2015.
- [20] Peter Petersen. *Riemannian Geometry, Second Edition* [M]. Beijing: Science Press, 2007.
- [21] V. I. Arnold, translated by Shen Jiaqi, Zhou Baoxi, Lu Tinghe. *Ordinary Differential Equations* [M]. Beijing: Science Press, 2010.

### Physics Section:

- [22] Chen Ximou, Shu Yousheng. Other Approaches to Establishing Maxwell's Equations [J]. University Physics, 1984, (2): 7–12.
- [23] Guo Shuohong. *Electrodynamics* [M]. Beijing: Higher Education Press, 1978.
- [24] Huang Chaoguang. *Lectures on General Relativity* [M]. Beijing: Science Press, 2023.
- [25] Liang Shaorong, Liu Changnian, Sheng Zhenghua. *Electromagnetism* [M]. Beijing: Higher Education Press, 1988.
- [26] Luo Changxun. *Introduction to Quantum Field Theory* [M]. Shaanxi: Shaanxi Normal University Press, 1986.
- [27] Liu Liao, Zhao Zheng. *General Relativity* [M]. Beijing: Higher Education Press, 2004.
- [28] [US] H. Goldstein. *Classical Mechanics (Second Edition)* [M]. Beijing: Science Press, 1986.

- [29] [US] W. Rindler, translated by Jiang Shan. *Essential Relativity* [M]. Anhui: Anhui Science and Technology Press, 1986.
- [30] Ni Guangjiong, Li Hongfang. *Modern Physics* [M]. Shanghai: Shanghai Science and Technology Press, 1979.
- [31] Wu Dayou. *Quantum Mechanics, Part A* [M]. Beijing: Science Press, 1984.
- [32] Yang Binglin, translated by Liu Guoli, Wang Wenyu, Wang Fei. *Dark Matter and Related Cosmology* [M]. Beijing: Science Press, 2019.
- [33] Zeng Jinyan. *Quantum Mechanics (Volumes I & II)* [M]. Beijing: Science Press, 2014.
- [34] Zhang Yici. *A Concise Course in Quantum Mechanics* [M]. Beijing: Higher Education Press, 1979.
- [35] Zhang Li, Ge Molin. *Frontier Issues in Quantum Mechanics (Third Edition)* [M]. Beijing: Tsinghua University Press, 2022.
- [36] Zhao Zheng, Liu Wenbiao. *Foundations of General Relativity* [M]. Beijing: Tsinghua University Press, 2010.