Mathematical Methods of Physics III

Lecture Notes – Fall 2012

Claus Montonen – Esko Keski-Vakkuri
Jouko Mickelsson
(Work in progress)

Contents

1 Introduction **4**

2 Group Theory **5**
2.1 Group 5
2.2 Smallest Finite Groups 6
2.2.1 More about the permutation groups S_n 9
2.3 Continuous Groups 12
2.3.1 Examples of Lie groups 13
2.4 Groups Acting on a Set 16
2.4.1 Conjugacy classes and cosets 18
2.4.2 Normal subgroups and quotient groups 20

3 Representation Theory of Groups **22**
3.1 Complex Vector Spaces and Representations 22
3.2 Symmetry Transformations in Quantum Mechanics 26
3.3 Reducibility of Representations 27
3.4 Irreducible Representations 30
3.5 Characters 33

4 Representations of the symmetric group **40**
4.1 Conjugacy classes in S_n 40
4.2 A List of irreducible representations of S_n 40

5 Some tensor analysis and representation theory **44**
5.1 Tensor products of representations and the symmetric group 44
5.2 Creation and annihilation operator formalism 49

6 Semisimple Lie algebras **52**
6.1 Lie algebras 52
6.2 Ideals in Lie algebras 55
6.3 The structure of semisimple Lie algebras 57

7 Representations of semisimple Lie algebras **65**

8 Reduction of tensor products of representions **70**

9 Differentiable Manifolds **72**
9.1 Topological Spaces 72
9.1.1 Continuous Maps 74
9.2 Homotopy Groups 76
9.2.1 Paths and Loops 76

9.2.2 Homotopy 76
9.2.3 Properties of the Fundamental Group 78
9.2.4 Higher Homotopy Groups 78
9.3 Differentiable manifolds 79
9.3.1 The definition of a differentiable manifold 79
9.3.2 Differentiable maps 80
9.3.3 Manifold with a Boundary 83
9.3.4 Vector fields 83
9.3.5 The flow generated by a vector field 85
9.3.6 Dual Vector Space 87
9.3.7 1-forms (i.e. cotangent vectors) 89
9.3.8 Tensors 90
9.3.9 Tensor Fields 91
9.3.10 Differential Map and Pullback 92
9.3.11 Lie Derivative 93
9.3.12 Differential Forms 94
9.3.13 Exterior derivative 96
9.3.14 Lie Groups and Algebras 98
9.3.15 Structure Constants of the Lie Algebra 101
9.3.16 The adjoint representation of G 101
9.4 de Rham cohomology 102
9.5 Integration of differential forms 104

10 Riemannian Geometry (Metric Manifolds) 108
10.1 The Metric Tensor 109
10.2 The Induced Metric 109
10.3 An application to Maxwell's equations 110
10.4 Affine Connection 112
10.5 Parallel Transport and Geodesics 112
10.6 The Covariant Derivative of Tensor Fields 113
10.7 The Transformation Properties of Connection Coefficients 114
10.8 The Metric Connection 115
10.9 Curvature And Torsion 116
10.10 Geodesics of Levi-Civita Connections 118
10.11 Lie Derivative And the Covariant Derivative 119
10.12 Isometries 119
10.13 Killing Vector Fields 120
10.14 APPENDIX: Einstein's Field Equations 122
10.14.1 The Newtonian Limit 125
10.14.2 The Schwarzschild Metric 126

11 Principal bundles and Yang-Mills systems **128**
11.1 Cartan's structural equations 128
11.2 Principal bundles 129
11.3 Connection and curvature in a principal bundle 132
11.4 Parallel transport 136
11.5 Covariant differentiation in vector bundles 137
11.6 An example: The monopole line bundle 139
11.7 Yang-Mills equations 141

1 Introduction

The course Mathematical Methods of Physics III (MMP III) is third in the series of courses introducing mathematical concepts and tools which are often needed in physics. The first two courses MMP I-II focused on analysis, providing tools to analyze and solve the dynamics of physical systems. In MMP III the emphasis is on geometrical and topological concepts, needed for the understanding of the symmetry principles and topological structures of physics. In particular, we will learn group theory (the basic tool to understand symmetry in physics, especially useful in quantum mechanics, quantum field theory and beyond), topology (needed for many subtler effects in quantum mechanics and quantum field theory), and differential geometry (the language of general relativity and modern gauge field theories). There are also many more sophisticated areas of mathematics that are also often used in physics, notable omissions in this course are more advanced topics in fibre bundles and complex geometry.

Course material will be available on the course homepage.

Let me know of any typos and confusions that you find. The lecture notes are based on those prepared and used by Claus Montonen, and later expanded by Esko Keski-Vakkuri, who lectured the course before me. In practice, they often follow very closely (and often verbatim) the three recommended textbooks:

- H.F. Jones: Groups, Representations and Physics (IOP Publishing, 2nd edition, 1998)
- M. Nakahara: Geometry, Topology and Physics (IOP Publishing, 1990)
- H. Georgi: Lie Algebras in Particle Physics (Addison-Wesley, 1982)

I have added some material, both in the end of the group theory part (some complements for finite and compact groups, as well a chapter about representations of semi-simple Lie algebras, somewhat different from the earlier versions) and in the end of the differential geometry part, in particular about principal bundles and Yang-Mills theory.

You don't necessarily have to rush to buy the books, they can be found in the reference section of the library in Physicum.

2 Group Theory

2.1 Group

Definition. A **group** G is a set of elements $\{a, b, \ldots\}$ with a law of composition (multiplication) which assigns to each ordered pair $a, b \in G$ another element $ab \in G$. (Note: $ab \in G$ (closure) is often necessary to check in order for the multiplication to be well defined). The multiplication must satisfy the following conditions:

G1 (associative law): For all $a, b, c \in G$, $a(bc) = (ab)c$.

G2 (unit element): There is an element $e \in G$ such that for all $a \in G$ $ae = ea = a$.

G3 (existence of inverse): For all $a \in G$ there is an element $a^{-1} \in G$ such that $aa^{-1} = a^{-1}a = e$.

If G satisfies **G1**, it is called a **semigroup**; if it also satisfies **G2**, it is called a **monoid**. The number of elements in the set G is called the **order** of the group, denoted by $|G|$. If $|G| < \infty$, G is a **finite** group. If G is a discrete set, G is a **discrete** group. If G is a continuous set, G is a **continuous** group.

Comments

i) In general $ab \neq ba$, i.e. the multiplication is not commutative. If $ab = ba$ for all $a, b \in G$, the group is called **Abelian**.

ii) The inverse element is unique: suppose that both b, b' are inverse elements of a. Then $b' = b'e = b'(ab) = (b'a)b = eb = b$.

Examples

1. Z with "+" (addition) as a multiplication is a discrete Abelian group.

2. R with "+" as a multiplication is a continuous Abelian group, $e = 0$. $R \setminus \{0\}$ with "$\cdot$" (product) is also a continuous Abelian group, $e = 1$. We had to remove 0 in order to ensure that all elements have an inverse.

3. $Z_2 = \{0, 1\}$ with addition modulo 2 is a finite Abelian group with order 2. $e = 0,\ 1^{-1} = 1$.

Let us also consider the set of mappings (functions) from a set X to a set Y, $Map(X,Y) = \{f : X \to Y | f(x) \in Y \ for\ all\ x \in X,\ f(x)\ is\ uniquely\ determined\}$. There are special cases of functions:

i) $f : X \to Y$ is called an **injection** (or **one-to-one**) if $f(x) \neq f(x')\ \forall x \neq x'$.

ii) $f : X \to Y$ is called a **surjection** (or **onto**) if $\forall y \in Y\ \exists x \in X\ s.t.\ f(x) = y$.

iii) if f is both an injection and a surjection, it is called a **bijection**.

Now take the composition of maps as a multiplication: $fg = f \circ g$, $(f \circ g)(x) = f(g(x))$. Then $(Map(X,X), \circ)$ (the set of functions $f : X \to X$ with $\circ$ as the multiplication) is a semigroup. We had to choose $Y = X$ to be able to use the composition, as g maps to Y but f is defined in X. Further, $(Map(X,X), \circ)$ is in fact a monoid with the identity map $id :\ id(x) = x$ as the unit element. However, it is *not* a group, unless we restrict to bijections. The set of bijections $f : X \to X$ is called the set of **permutations** of X, we denote $Perm(X) = \{f \in Map(X,X) | f\ is\ a\ bijection\}$. Every $f \in Perm(X)$ has an inverse map, so $Perm(X)$ is a group. However, in general $f(g(x)) \neq g(f(x))$, so $Perm(X)$ is not an Abelian group. An important special case is when X has a finite number N of elements. This is called the **symmetric group** or the **permutation group**, and denoted by S_N. The order of S_N is $|S_N| = N!$ (exercise).

Definitions

i) We denote $g^2 = gg,\ g^3 = ggg = g^2 g,\ \ldots,\ g^n = \overbrace{g \cdots g}^{n}$ for products of the element $g \in G$.

ii) The **order of the element** n of $g \in G$ is the smallest number n such that $g^n = e$.

2.2 Smallest Finite Groups

Let us find all the groups of order n for $n = 1, \ldots, 4$. First we need a handy definition. A **homomorphism** in general is a mapping from one set X to another set Y preserving some structure. Further, if f is a bijection, it is called an **isomorphism**. We will see several examples of such structure-preserving mappings. The first one is the one that preserves the multiplication structure of groups.

Definition. A mapping $f : G \to H$ between groups G and H is called a **group homomorphism** if for all $g_1, g_2 \in G$, $f(g_1 g_2) = f(g_1) f(g_2)$. Further, if f is also a bijection, it is called a **group isomorphism**. If there exists a group isomorphism between groups G and H, we say that the groups are **isomorphic**, and denote $G \cong H$. Isomorphic groups have an identical structure, so they can be identified – there is only one abstract group of that structure.

Example. Take $G = \mathbb{R}_+$ with "$\cdot$" and $H = \mathbb{R}$ with "$+$" as a multiplication. Define the mapping $f : G \to H, f(x) = \ln x$. Now f is a group homomorphism, because $f(xy) = \ln(xy) = \ln x + \ln y = f(x) + f(y)$. In fact, f is also a group isomorphism, because it is a bijection: $f^{-1}(x) = e^x$.

Now let us move ahead to groups of order n.

Order $n = 1$. This is the trivial group $G = \{e\}$, $e^2 = e$.

Order $n = 2$. Now $G = \{e, a\},\ a \neq e$. The multiplications are $e^2 = e,\ ea = ae = a$. For a^2, let's first try $a^2 = a$. But then $a = ae = a(aa^{-1}) = a^2a^{-1} = aa^{-1} = e$, a contradiction. So the only possibility is $a^2 = e$. We can summarize this in the **multiplication table** or **Cayley table**:

	e	a
e	e	a
a	a	e

This group is called Z_2. You have already seen another realization of it: the set $\{0, 1\}$ with addition modulo 2 as the multiplication. Yet another realization of the group is $\{1, -1\}$ with product as the multiplication. This illustrates what was said before: for a given abstract group, there can be many ways to describe it. Consider one more realization: the permutation group $S_2 = Perm(\{1, 2\})$. Its elements are

$$e = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ 1 & 2 \end{pmatrix} \equiv \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

$$a = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ 2 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} ,$$

the arrows indicate how the numbers are permuted, we usually use the notation in the right hand side without the arrows. For products of permutations, the order in which they are performed is "right to left": we first perform the permutation on the far right, then continue with the next one to the left, and so one. This convention is inherited from that with composite mappings: (fg)(x)=f(g(x)). We can now easily show that S_2 is isomorphic with Z_2. Take e.g. $\{1, -1\}$ with the product as the realization of Z_2. Then we define the mapping $i : Z_2 \to S_2 :\ i(1) = e,\ i(-1) = a$. It is easy to see that i is a group homomorphism, and it is obviously a bijection. Hence it is an isomorphism, and $Z_2 \cong S_2$. There is only one abstract group of order 2.

Order $n = 3$**.** Consider now the set $G = \{e, a, b\}$. It turns out that there is again only one possible group of order 3. We can try to determine it by completing its multiplication table:

	e	a	b
e	e	a	b
a	a	?	?
b	b	?	?

First, guess $ab = b$. But then $a = a(bb^{-1}) = (ab)b^{-1} = bb^{-1} = e$, a contradiction. Try then $ab = a$. But now $b = (a^{-1}a)b = a^{-1}(ab) = a^{-1}a = e$, again contradiction. So $ab = e$. Similarly, $ba = e$. Then, guess $a^2 = a$. Now $a = aaa^{-1} = aa^{-1} = e$, doesn't work. How about $a^2 = e$? Now $b = a^2b = a(ab) = ae = a$, doesn't work. So $a^2 = b$. Similarly, can show $b^2 = a$. Now we have worked out the complete multiplication table:

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Our group is actually called Z_3. We can simplify the notation and call $b = a^2$, so $Z_3 = \{e, a, a^2\}$. Z_3 and Z_2 are special cases of *cyclic groups* $Z_n = \{e, a, a^2, \ldots, a^{n-1}\}$. They have a single "generating element" a with order n: $a^n = e$. The multiplication rules are $a^p a^q = a^{p+q(mod\ n)}$, $(a^p)^{-1} = a^{n-p}$. Sometimes in the literature cyclic groups are denoted by C_n. One possible realization of them is by complex numbers, $Z_n = \{e^{\frac{2\pi ik}{n}} | k = 0, 1, \ldots\}$ with product as a multiplication. This also shows their geometric interpretation: Z_n is the symmetry group of rotations of a regular directed polygon with n sides (see H.F.Jones). You can easily convince yourself that $Z_n = \{0, 1, \ldots, n-1\}$ with addition modulo n is another realization.

Order $n = 4$**.** So far the groups have been uniquely determined, but we'll see that from order 4 onwards we'll have more possibilities. Let's start with a definition.

Definition. A **direct product** $G_1 \times G_2$ of two groups is the set of all pairs (g_1, g_2) where $g_1 \in G_1$ and $g_2 \in G_2$, with the multiplication $(g_1, g_2) \cdot (g_1', g_2') = (g_1 g_1', g_2 g_2')$. The unit element is (e_1, e_2) where e_i is the unit element of G_i $(i = 1, 2)$. It is easy to see that $G_1 \times G_2$ is a group, and its order is $|G_1 \times G_2| = |G_1||G_2|$.

Now we can immediately find at least one group of order 4: the direct product $Z_2 \times Z_2$. Denote $Z_2 = \{e, f\}$ with $f^2 = e$, and introduce a shorter notation for the pairs: $E = (e,e),\ A = (e,f),\ B = (f,e),\ C = (f,f)$. We can easily find the multiplication table,

	E	A	B	C
E	E	A	B	C
A	A	E	C	B
B	B	C	E	A
C	C	B	A	E

The group $Z_2 \times Z_2$ is sometimes also called "Vierergruppe" and denoted by V_4.

There is another group of order 4, namely the cyclic group $Z_4 = \{e, a, a^2, a^3\}$. It is not isomorphic with $Z_2 \times Z_2$. (You can easily check that it has a different multiplication table.) It can be shown (exercise) that there are no other groups of order 4, just the above two.

Order $n \geq 5$. As can be expected, there are more possible non-isomorphic groups of higher finite order. We will not attempt to categorize them much further, but will mention some interesting facts and examples.

Definition. If H is a subset of the group G such that

i) $\forall\ h_1, h_2 \in H:\ h_1 h_2 \in H$

ii) $\forall\ h \in H:\ h^{-1} \in H$,

then H is called a **subgroup** of G. Note as a result of **i)** and **ii)**, every subgroup must include the unit element e of G.

Trivial examples of subgroups are $\{e\}$ and G itself. Other subgroups H are called **proper subgroups** of G. For those, $|H| \leq |G| - 1$.

Example. Take $G = Z_3$. Are there any proper subgroups? The only possibilities could be $H = \{e, a\}$ or $H = \{e, a^2\}$. Note that in order for H to be a group of order 2, it should be isomorphic with Z_2. But since $a^2 \neq e$ (because $a^3 = e$) and $(a^2)^2 = a^3 a = a \neq e$, neither is. So Z_3 has no proper subgroups.

2.2.1 More about the permutation groups S_n

It is worth spending some more time on the permutation groups, because on one hand they have a special status in the theory of finite groups (for a reason that I will explain later) and on the other hand they often appear in physics.

Let $X = \{1, 2, \ldots, n\}$. Denote a bijection of X by $p : X \to X,\ i \mapsto p(i) \equiv p_i$. We will now generalize our notation for the elements of S_n, you already saw it for S_2. We denote a $P \in S_n \equiv Perm(X)$ by

$$P = \begin{pmatrix} 1 & 2 & \cdots & n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix} .$$

Recall that the multiplication rule for permutations was the composite operation, with the "right to left" rule. In general, the multiplication is not commutative:

$$PQ = \begin{pmatrix} 1 & 2 & \cdots & n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix} \begin{pmatrix} 1 & 2 & \cdots & n \\ q_1 & q_2 & \cdots & q_n \end{pmatrix} \neq QP \ .$$

So, in general, S_n is not an abelian group. (Except S_2.) For example, in S_3,

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \tag{1}$$

but

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

which is not the same.

The identity element is

$$E = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$$

and the inverse of P is

$$P^{-1} = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ 1 & 2 & \cdots & n \end{pmatrix} \ .$$

An alternative and very useful way of writing permutations is the **cycle notation**. In this notation we follow the permutations of one label, say 1, until we get back to where we started (in this case back to 1), giving one **cycle**. Then we start again from a label which was not already included in the previously found cycle, and find another cycle, and so on until all the labels have been accounted for. The original permutation has then been decomposed into a certain number of *disjoint* cycles. This is best illustrated by an example. For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

of S_4 decomposes into the disjoint cycles $1 \to 2 \to 4 \to 1$ and $3 \to 3$. Reordering the columns we can write it as

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix} = \left(\begin{array}{ccc|c} 1 & 2 & 4 & 3 \\ 2 & 4 & 1 & 3 \end{array}\right) = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \end{pmatrix} \ .$$

In a cycle the bottom row is superfluous: all the information about the cycle (like $1 \to 2 \to 4 \to 1$) is already included in the order of the labels in the top row. So we can shorten the notation by simply omitting the bottom row. The above example is then written as

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix} = (124)(3) \ .$$

As a further abbreviation of the notation, we omit the 1-cycles (like (3) above), it being understood that any labels not appearing explicitly just transform into themselves. With the new shortened cycle notation, (1) reads

$$(23)(132) = (12) \tag{3}$$

and (2) reads as

$$(132)(23) = (13)\ . \tag{4}$$

In general, any permutation can always be written as the product of disjoint cycles. What's more, the cycles *commute* since they operate on different indices, hence the cycles can be written in any order in the product. In listing the individual permutations of S_n it is convenient to group them by cycle structure, i.e. by the number and length of cycles. For illustration, we list the first permutation groups S_n:

$n = 2$**:** $S_2 = \{E, (12)\}$.

$n = 3$**:** $S_3 = \{E, (12), (13), (23), (123), (132)\}$.

$n = 4$**:** $S_4 = \{E, (12), (13), (14), (23), (24), (34), (12)(34), (13)(24), (14)(23),$
$(123), (132), (124), (142), (134), (143), (234), (243),$
$(1234), (1243), (1324), (1342), (1423), (1432)\}$.

You can see that the notation makes it quite easy and systematic to write down all the elements in a concise fashion.

The simplest non-trivial permutations are the 2-cycles, which interchange two labels. In fact, *any permutation* can be built up from products of 2-cycles. First, an r-cycle can be written as the product of $r-1$ overlapping 2-cycles:

$$(n_1 n_2 \ldots n_r) = (n_1 n_2)(n_2 n_3) \cdots (n_{r-1} n_r)\ .$$

Then, since any permutation is a product of cycles, it can be written as a product of 2-cycles. This allows us to classify permutations as "even" and "odd". First, a 2-cycle which involves just one interchange of labels is counted as odd. Then, a product of 2-cycles is even (odd), if there are an even (odd) number 2-cycles. Thus, an r-cycle is **even** (**odd**), if r is odd (even). (Since it is a product of $r-1$ 2-cycles.) Finally, a generic product of cycles is even if it contains an even number of odd cycles, otherwise it is odd. In particular, the identity E is even. This allows us to find an interesting subgroup of S_n, the **alternating group** A_n which consists of the *even* permutations of S_n. The order of A_n is $|A_n| = \frac{1}{2} \cdot |S_n|$. Hence A_n is a proper subgroup of S_n. Note that the odd permutations do not form a subgroup, since any subgroup must contain the identity E which is even.

To keep up a promise, we now mention the reason why permutation groups have a special status among finite groups. This is because of the following theorem (we state it without proof).

Theorem 2.1 (Cayley's Theorem) *Every finite group of order n is isomorphic to a subgroup of S_n.*

Thus, because of Cayley's theorem, in principle we know everything about finite groups if we know everything about permutation groups and their subgroups.

As for physics uses of finite groups, the classic example is their role in solid state physics, where they are used to classify general crystal structures (the so-called crystallographic point groups). They are also useful in classical mechanics, reducing the number of relevant degrees of freedom in systems of symmetry. We may later study an example, finding the vibrational normal modes of a water molecule. In addition to these canonical examples, they appear in different places and roles in all kinds of areas of modern physics.

2.3 Continuous Groups

Continuous groups have an uncountable infinity of elements. The **dimension** of a continuous group G, denoted $\dim G$, is the number of continuous real parameters (coordinates) which are needed to uniquely parameterize its elements. In the product $g'' = g'g$, the coordinates of g'' must be continuous functions of the coordinates of g and g'. (We will make this more precise later when we discuss topology. The above requirement means that the set of real parameters of the group must be a *manifold*, in this context called the *group manifold*.)

Examples.

1. The set of real numbers R with addition as the product is a continuous group; $\dim R = 1$. Simple generalization: $R^n = \{(r_1, \ldots, r_n) | r_i \in R,\ i = 1, \ldots, n\} = \overbrace{R \times \cdots \times R}^{n\ times}$, with product $(r_1, \ldots, r_n) \cdot (r'_1, \ldots, r'_n) = (r_1 + r'_1, \ldots, r_n + r'_n)$, $\dim R^n = n$.

2. The set of complex numbers C with addition as the product, $\dim C = 2$ (recall that we count the number of real parameters).

3. The set of $n \times n$ real matrices $M(n, R)$ with addition as the product, $\dim M(n, R) = n^2$. Note group isomorphism: $M(n, R) \cong R^{n^2}$.

4. $U(1) = \{z \in C | |z|^2 = 1\}$, with multiplication of complex numbers as the product. $\dim U(1) = 1$ since there's only one real parameter $\theta \in [0, 2\pi]$, $z = e^{i\theta}$. Note a difference with $U(1)$ and R: both have $\dim = 1$ but the group manifold of the former is the circle S^1 while the group manifold of the latter is the whole infinite x-axis. A generalization of $U(1)$ is $U(1)^n = \overbrace{U(1) \times \cdots \times U(1)}^{n\ times}$,

$(e^{i\theta_1}, \ldots, e^{i\theta_n}) \cdot (e^{i\theta'_1}, \ldots, e^{i\theta'_n}) = (e^{i(\theta_1+\theta'_1)}, \ldots, e^{i(\theta_n+\theta'_n)})$. The group manifold of $U(1)^n$ is an n-torus $\overbrace{S^1 \times \cdots \times S^1}^{n}$. Again, the n-torus is different from R^n: on the former it is possible to draw loops which cannot be smoothly contracted to a point, while this is not possible on R^n.

All of the above examples are actually examples of **Lie groups**. Their group manifolds must be *differentiable manifolds*, meaning that we can take smooth (partial) derivatives of the group elements with respect to the real parameters. We'll give a precise definition later – for now we'll just focus on listing further examples of them.

2.3.1 Examples of Lie groups

1. The group of general linear transformations $GL(n, R) = \{A \in M(n, R)|\ \det A \neq 0\}$, with matrix multiplication as the product; $\dim GL(n, R) = n^2$. While $GL(n, R)$, $M(n, R)$ have the same dimension, their group manifolds have a different structure. To parameterize the elements of $M(n, R)$, only one coordinate neighborhood is needed (R^{n^2} itself). The coordinates are the matrix entries a_{ij}:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} .$$

   In $GL(n, R)$, the condition $\det A \neq 0$ removes a hyperplane (a set of measure zero) from R^{n^2}, dividing it into two disconnected coordinate regions. In each region, the entries a_{ij} are again suitable coordinates.

2. A generalization of the above is $GL(n, C) = \{n \times n$ complex matrices with non $-$ zero determinant$\}$, with matrix multiplication as the product. This has $\dim GL(n, C) = 2n^2$. Note that $GL(n, R)$ is a (proper) subgroup of $GL(n, C)$. The following examples are subgroups of these two.

3. The group of special linear transformations $SL(n, R) = \{A \in GL(n, R)|\ \det A = 1\}$. It is a subgroup of $GL(n, R)$ since $\det(AB) = \det A \det B$. The dimension is $\dim SL(n, R) = n^2 - 1$.

4. The orthogonal group $O(n, R) = \{A \in GL(n, R)|\ A^T A = 1_n\}$, *i.e.* the group of orthogonal matrices. (1_n denotes the $n \times n$ unit matrix.) A^T is the **transpose** of the matrix A:

$$A^T = \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix} ,$$

   i.e. if $A = (a_{ij})$ then $A^T = (a_{ji})$, the rows and columns are interchanged. Let's prove that $O(n, R)$ is a subgroup of $GL(n, R)$:

a) $1_n^T = 1_n$ so the unit element $\in O(n,R)$

b) If A, B are orthogonal, then AB is also orthogonal: $(AB)^T(AB) = B^TA^TAB = B^TB = 1_n$.

c) Every $A \in O(n,R)$ has an inverse in $O(n,R)$: $(A^{-1})^T = (A^T)^{-1}$ so $(A^{-1})^TA^{-1} = (A^T)^{-1}A^{-1} = (AA^T)^{-1} = ((A^T)^TA^T)^{-1} = 1_n^{-1} = 1_n$.

Note that orthogonal matrices preserve the length of a vector. The length of a vector $\vec{v}$ is $\sqrt{v_1^2 + \cdots v_n^2} = \sqrt{\vec{v}^T\vec{v}}$. A vector $\vec{v}$ gets mapped to $A\vec{v}$, so its length gets mapped to $\sqrt{(A\vec{v})^T(A\vec{v})} = \sqrt{\vec{v}^TA^TA\vec{v}} = \sqrt{\vec{v}^T\vec{v}}$, the same. We can interpret the orthogonal group as the group of rotations in R^n.
What is the dimension of $O(n,R)$? $A \in GL(n,R)$ has n^2 independent parameters, but the orthogonality requirement $A^TA = 1_n$ imposes relations between the parameters. Let us count how many relations (equations) there are. The diagonal entries of A^TA must be equal to one, this gives n equations; the entries above the diagonal must vanish, this gives further $n(n-1)/2$ equations. The same condition is then automatically satisfied by the "below the diagonal" entries, because the condition $A^TA = 1_n$ is symmetric: $(A^TA)^T = A^TA = (1_n)^T = 1_n$. Thus there are only $n^2 - n - n(n-1)/2 = n(n-1)/2$ free parameters. So $\dim O(n,R) = n(n-1)/2$.
Another fact of interest is that $\det A = \pm 1$ for every $A \in O(n,R)$. Proof: $\det(A^TA) = \det(A^T)\det A = \det A \det A = (\det A)^2 = \det 1_n = 1 \Rightarrow \det A = \pm 1$. Thus the group $O(n,R)$ is divided into two parts: the matrices with $\det A = +1$ and the matrices with $\det A = -1$. The former part actually forms a subgroup of $O(n,R)$, called $SO(n,R)$ (you can figure out why this is true, and not true for the part with $\det A$=-1). So we have one more example:

5. The group of special orthogonal transformations $SO(n,R) = \{A \in O(n,R)|\ \det A = 1\}$. $\dim SO(n,R) = \dim O(n,R) = n(n-1)/2$.

6. The group of unitary matrices (transformations) $U(n) = \{A \in GL(n,C)|\ A^\dagger A = 1_n\}$, where $A^\dagger = (A^*)^T = (A^T)^*$: $(A^\dagger)_{ij} = (A_{ji})^*$. Note that $(AB)^\dagger = B^\dagger A^\dagger$. These preserve the length of complex vectors $\vec{z}$. The length is defined as $\sqrt{z_1^*z_1 + \cdots z_n^*z_n} = \sqrt{\vec{z}^\dagger\vec{z}}$. Under A this gets mapped to $\sqrt{(A\vec{z})^\dagger A\vec{z}} = \sqrt{\vec{z}^\dagger A^\dagger A\vec{z}} = \sqrt{\vec{z}^\dagger\vec{z}}$. The unitary matrices are rotations in C^n. We leave it as an exercise to show that $U(n)$ is a subgroup of $GL(n,C)$, and $\dim U(n) = n^2$. Note that $U(1) = \{a \in C|\ a^*a = 1\}$, its group manifold is the unit circle S^1 on the complex plane.

7. The special unitary group $SU(n) = \{A \in U(n)|\ \det A = 1\}$. This is the complex analogue of $SO(n,R)$, and is a subgroup of $U(n)$. Exercise: $\dim SU(n) = n^2 - 1$. $U(n)$ and $SU(n)$ groups are important in modern physics. You will probably first become familiar with $U(1)$, the group of phase transformations in quantum

mechanics, and with $SU(2)$, in the context of spin. Let's take a closer look at the latter. It's dimension is three. What does its group manifold look like? Let's first parameterize the $SU(2)$ matrices with complex numbers a, b, c, d:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \; , \; A^\dagger = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} \; .$$

Then

$$\begin{aligned} \det A &= ad - bc = 1 \\ A^\dagger A &= \begin{pmatrix} |a|^2 + |c|^2 & a^* b + c^* d \\ b^* a + d^* c & |b|^2 + |d|^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \; . \end{aligned}$$

Let's first assume $a \neq 0$. Then $b = -c^* d / a^*$. Substituting to the determinant condition gives $ad - bc = d(|a|^2 + |c|^2)/a^* = d/a^* = 1 \Rightarrow d = a^*$. Then $c = -b^*$. So

$$A = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \; .$$

Assume then $a = 0$. Now $|c|^2 = 1$, $c^* d = 0 \Rightarrow d = 0$. Then $|c|^2 = |b|^2 = 1$. Write $b = e^{i\beta}, c = e^{i\gamma}$. Then $\det A = -bc = e^{i(\beta+\gamma+\pi)} = 1 \rightarrow \gamma = -\beta + (2n+1)\pi$. Then $c = e^{i\gamma} = e^{-i\beta} e^{i(2n+1)\pi} = -e^{-i\beta} = -b^*$. Thus

$$A = \begin{pmatrix} 0 & b \\ -b^* & 0 \end{pmatrix} \; .$$

Let us trade the two complex parameters with four real parameters x_1, x_2, x_3, x_4: $a = x_1 + i x_2$, $b = x_3 + i x_4$. Then A becomes

$$A = \begin{pmatrix} x_1 + i x_2 & x_3 + i x_4 \\ -x_3 + i x_4 & x_1 - i x_2 \end{pmatrix} \; .$$

The determinant condition $\det A = 1$ then turns into the constraint

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$$

for the four real parameters. This defines an unit 3-sphere. More generally, we define an **n-sphere** $S^n = \{(x_1, \ldots, x_{n+1}) \in R^{n+1} | \; \sum_{i=1}^{n+1} x_i^2 = 1\}$. The group manifold of $SU(2)$ is a three-sphere S^3. (And the group manifold of $U(1)$ was a 1-sphere S^1. As a matter of fact, these are the only Lie groups with n-sphere group manifolds.) The n-sphere is an example of so-called pseudospheres. We'll meet other examples in an exercise.

8. As an aside, note that $O(n, R), SO(n, R), U(n), SU(n)$ were associated with rotations in R^n or C^n, keeping invariant the lengths of real or complex vectors. One can generalize from real and complex numbers to quaternions and

octonions, and look for generalizations of the rotation groups. This produces other examples of (compact) Lie groups, the $Sp(2n), G_2, F_4, E_6, E_7$ and E_8. The *symplectic* group $Sp(2n)$ plays an important role in classical mechanics, it is associated with canonical transformations in phase space. The other groups crop up in string theory.

2.4 Groups Acting on a Set

We already talked about the orthogonal groups as rotations, implying that the group acts on points in R^n. We should make this notion more precise. First, review the definition of a homomorphism from p. 6, then you are ready to understand the following

Definition. Let G be a group, and X a set. The **(left) action** of G on X is a homomorphism $L : G \to Perm(X)$, $G \ni g \mapsto L_g \in Perm(X)$. Thus, L satisfies $(L_{g_2} \circ L_{g_1})(x) = L_{g_2}(L_{g_1}(x)) = L_{g_2 g_1}(x)$, where $x \in X$. The last equality followed from the homomorphism property. We often simplify the notation and denote $gx \equiv L_g(x)$. Given such an action, we say that X is a **(left)** G**-space**. Respectively, the **right action** of G in X is a homomorphism $R : G \to Perm(X)$, $R_{g_2} \circ R_{g_1} = R_{g_1 g_2}$ (note order in the subscript!), $xg \equiv R_g(x)$. We then say that X is a **right** G**-space**.

Two (left) G-spaces X, X' can be identified, if there is a bijection $i : X \to X'$ such that $i(L_g(x)) = L'_g(i(x))$ where L, L' are (left) actions of G on X, X'. A mathematician would say this in the following way: the diagram

$$\begin{array}{ccc} X & \xrightarrow{i} & X' \\ L_g \downarrow & \searrow & \downarrow L'_g \\ X & \xrightarrow{i} & X' \end{array}$$

commutes, *i.e.* the map in the diagonal can be composed from the vertical and horizontal maps through either corner.

Definition. The **orbit** of a point $x \in X$ under the action of G is the set $O_x = \{L_g(x) |\ g \in G\}$. In other words, the orbit is the set of all points that can be reached from x by acting on it with elements of G. Let's put this in another way, by first introducing a useful concept.

Definition. An **equivalence relation** $\sim$ in a set X is a relation between points in a set which satisfies

i) $a \sim a$ (reflective) $\forall\ a \in X$

ii) $a \sim b \Rightarrow b \sim a$ (symmetric) $\forall\ a, b \in X$

iii) $a \sim b$ and $b \sim c \Rightarrow a \sim c$ (transitive) $\forall\ a, b, c \in X$

Given a set X and an equivalence relation $\sim$, we can partition X into mutually disjoint subsets called **equivalence classes**. An equivalence class $[a] = \{x \in X |\ x \sim a\}$, the set of all points which are equivalent to a under $\sim$. The element a (or any other element in its equivalence class) is called the **representative** of the class. Note that $[a]$ is not an empty set, since $a \sim a$. If $[a] \bigcap [b] \neq \emptyset$, there is an $x \in X$ s.t. $x \sim a$ and $x \sim b$. But then, by transitivity, $a \sim b$ and $[a] = [b]$. Thus, different equivalence classes must be mutually disjoint ($[a] \neq [b] \Rightarrow [a] \bigcap [b] = \emptyset$). The set of all equivalence classes is called the **quotient space** and denoted by $X/\sim$.

Example. Let n be a non-negative integer. Define an equivalence relation among integers $r, s \in Z$: $r \sim s$ if $r - s = 0\ (mod\ n)$. (Prove that this indeed is an equivalence relation.) The quotient space is $Z/\sim = \{[0], [1], [2], \ldots, [n-1]\}$. Define the addition of equivalence classes: $[a] + [b] = [a + b]$. Then $Z/\sim$ with addition as a multiplication is a finite Abelian group, isomorphic to the cyclic group: $Z/\sim \cong Z_n$. (Exercise: prove the details.)

Back to orbits then. A point belonging to the orbit of another point defines an equivalence relation: $y \sim x$ if $y \in O_x$. The equivalence class is the orbit itself: $[x] = O_x$. Since the set X is partitioned into mutually disjoint equivalence classes, it is partitioned into mutually disjoint orbits under the action of G. We denote the quotient space by X/G. It may happen that there is only one such orbit, then $O_x = X\ \forall x \in X$. In this case we say that the action of G on X is **transitive**, and X is a **homogenous space**.

Examples.

1. $G = Z_2 = \{1, -1\}$, $X = R$. Left actions: $L_1(x) = x$, $L_{-1}(x) = -x$. Orbits: $O_0 = \{0\}$, $O_x = \{x, -x\}$ ($\forall\ x \neq 0$). The action is not transitive.

2. $G = SO(2, R)$, $X = R^2$. Parameterize

$$SO(2, R) \ni g = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} ,$$

and write

$$R^2 \ni x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

Left action:

$$L_g(x) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\theta\ x_1 - \sin\theta\ x_2 \\ \sin\theta\ x_1 + \cos\theta\ x_2 \end{pmatrix}$$

(rotate vector x counterclockwise about the origin by angle θ). Orbits are circles with radius r about the origin: $O_0 = \{0\}, O_{x\neq 0} = \{x \in R^2 |\ x_1^2 + x_2^2 = r^2\}$, $r = \sqrt{x_1^2 + x_2^2}$. The action is not transitive. $R^2/SO(2,R) = \{r \in R|\ r \geq 0\}$.

3. $G = GL(n,R)$, $X = R^n$. Left action: $L_A(x) = x'$ where $x'_i = \sum_{j=1}^n A_{ij}x_j$. The orbit of the origin 0 is $O_0 = \{0\}$, other points have other orbits. So the action is not transitive.

2.4.1 Conjugacy classes and cosets

We can also let the group act on itself, *i.e.* take $X = G$. A simple way to define the left action of G on G is the *translation*, $L_g(g') = gg'$. Every group element belongs to the orbit of identity, since $L_g(e) = ge = g$. So $O_e = G$, the action is transitive. A more interesting way to define group action on itself is by *conjugation*.

Definition. Two elements g_1, g_2 of a group G are **conjugate** if there is an element $g \in G$ such that $g_1 = gg_2g^{-1}$. The element g is called the **conjugating element**.

We then take conjugation as the left action, $L_g(g') = gg'g^{-1}$. In general conjugation is not transitive. The orbits have a special name, they are called **conjugacy classes**.

It is also very interesting to consider the action of subgroups H of G on G. Define this time a *right* action of H on G by translation, $R_h(g) = gh$. If H is a proper subgroup, the action need not be transitive.

Definition. The orbits, or the equivalence classes

$$[g] = \{g' \in G|\ \exists h \in H\ s.t.\ g' = gh\} = \{gh|\ h \in H\}$$

are called **left cosets** of H, and usually they are denoted gH. The quotient space $G/H = \{gH|\ g \in G\}$ is the set of left cosets. (Similarly, we can define the left action $L_h(g) = hg$ and consider the right cosets Hg. Then the quotient space is denoted $H\backslash G$.)

Comments.

1. $ghH = gH$ for all $h \in H$.

2. If $g_1H = g_2H$, there is an $h \in H$ such that $g_2 = g_1h$ *i.e.* $g_1^{-1}g_2 \in H$.

3. There is a one-one correspondence between the elements of every coset and between the elements of H itself. The map $f_g : H \to gH$, $f_g(h) = gh$ is obviously a surjection; it is also an injection since $gh_1 = gh_2 \Rightarrow h_1 = h_2$. In particular, if H is finite, all the orders are the same: $|H| = |gH| = |g'H|$. This leads to the following theorem:

Theorem 2.2 (Lagrange's Theorem) *The order $|H|$ of any subgroup H of a finite group G must be a divisor of $|G|$: $|G| = n|H|$ where n is a positive integer.*

Proof. Under right action of H, G is partitioned into mutually disjoint orbits gH, each having the same order as H. Hence $|G| = n|H|$ for some n.

Corollary. If $p = |G|$ is a prime number, then $G \cong Z_p$.

Proof. Pick $g \in G$, $g \neq e$, denote the order of the element g by m. Then $H = \{e, g, \ldots g^{m-1}\} \cong Z_m$ is a subgroup of G. But according to Lagrange's theorem $|G| = nm$. For this to be prime, $n = 1$ or $m = 1$. But $g \neq e$, so $m > 1$ so $n = 1$ and $|G| = |H|$. But then it must be $H = G$.

Definition. Let group G act on a set X. The **little group** of $x \in X$ is the subgroup $G_x = \{g \in G | \, L_g(x) = x\}$ of G. It contains all elements of G which leave x invariant. It obviously contains the unit element e, you can easily show the other properties of a subgroup. The little group is also sometimes called the **isotropy group, stabilizer** or **stability group**.

Back to cosets. The set of cosets G/H is a G-space, if we define the left action $l_g : G/H \to G/H$, $l_g(g'H) = gg'H$. The action is transitive: if $g_1 H \neq g_2 H$, then $l_{g_1 g_2^{-1}}(g_2 H) = g_1 H$. The inverse is also true:

Theorem 2.3 *Let group G act transitively on a set X. Then there exists a subgroup H such that X can be identified with G/H. In other words, there exists a bijection $i : G/H \to X$ such that the diagram*

$$\begin{array}{ccc} G/H & \xrightarrow{i} & X \\ l_g \downarrow & \searrow & \downarrow L_g \\ G/H & \xrightarrow{i} & X \end{array}$$

commutes.

Proof. Choose a point $x \in X$, denote its isotropy group G_x by H. Define a map $i : G/H \to X$, $i(gH) = L_g(x)$. It is well defined: if $gH = g'H$, then $g = g'h$ with some $h \in H$ and $L_g(x) = L_{g'h}(x) = L_{g'}(L_h(x)) = L_{g'}(x)$. It is an injection: $i(gH) = i(g'H) \Rightarrow L_g(x) = L_{g'}(x) \Rightarrow x = L_{g^{-1}}(L'_g(x)) = L_{g^{-1}g'}(x) \Rightarrow g^{-1}g' \in H \Rightarrow g' = gh \Rightarrow gH = g'H$. It is also a surjection: G acts transitively so for all $x' \in X$ there exists g s.t. $x' = L_g(x) = i(gH)$. The diagram commutes: $(L_g \circ i)(g'H) = L_g(L_{g'}(x)) = L_{gg'}(x) = i(gg'H) = (i \circ l_g)(g'H)$.

Corollary. A consequence of the proof is that the orbit of a point $x \in X$, O_x, can be identified with G/G_x since G acts transitively on its orbits. Thus the orbits are determined by the subgroups of G, in other words the action of G on X is determined by the subgroup structure.

Example. $G = SO(3,R)$ acts on R^3, the orbits are the spheres $|x|^2 = x_1^2 + x_2^2 + x_3^2 = r^2$, *i.e.* S^2 when $r > 0$. Choose the point $x =$ north pole $= (0,0,r)$ on every orbit $r > 0$. Its little group is

$$G_x = \left\{ \begin{pmatrix} A_{2\times 2} & 0 \\ 0 & 1 \end{pmatrix} \mid A_{2\times 2} \in SO(2,R) \right\} \cong SO(2,R) \ .$$

By Theorem 2.3 and its Corollary, $SO(3,R)/SO(2,R) = S^2$.

2.4.2 Normal subgroups and quotient groups

Since the quotient space G/H is constructed out of a group and its subgroup, it is natural to ask if it can also be a group. The first guess for a multiplication law would be

$$(g_1 H)(g_2 H) = g_1 g_2 H \ .$$

This definition would be well defined if the right hand side is independent of the labeling of the cosets. For example $g_1 H = g_1 h H$, so we then need $g_1 g_2 H = g_1 h g_2 H$ *i.e.* find $h' \in H$ s.t. $g_1 g_2 h' = g_1 h g_2$. But this is not always true. We can circumvent the problem if H belongs to a particular class of subgroups, so called *normal* (also called *invariant, selfconjugate*) subgroups.

Definition. A **normal subgroup** H of G is one which satisfies $gHg^{-1} = \{ghg^{-1} \mid h \in H\} = H$ for all $g \in G$.

Another way to say this is that H is a normal subgroup, if for all $g \in G, h \in H$ there exists a $h' \in H$ such that $gh = h'g$.

Consider again the problem in defining a product for cosets. If H is a normal subgroup, then $g_1 h g_2 = g_1(hg_2) = g_1(g_2 h') = g_1 g_2 h'$ is possible. One can show that the above multiplication satisfies associativity, existence of identity (it is eH) and existence of inverse $(gH)^{-1} = g^{-1}H$. Hence G/H is a group if H is a normal subgroup. When G/H is a group, it is called a **quotient group**.

Comments:

1. If H is a normal subgroup, its left and right cosets are the same: $gH = Hg$.
2. If G is Abelian, all of its subgroups are normal.
3. $|G/H| = |G|/|H|$ (follows from Lagrange's theorem).

Example. Consider the cyclic group $C_{2n} = \{e, a, \dots, a^{2n-1}\}, n \in \mathbb{Z}$. Take $H = \{e, a^2, a^4, \dots, a^{2(n-1)}\}$. You can easily see that H is a subgroup of C_{2n}. Because cyclic groups are Abelian, H is normal. The two cosets are $H = a^2H = \dots = a^{2(n-1)}H$ and $aH = \{a, a^3, a^5, \dots, a^{2n-1}\} = a^3H = \dots = a^{2n-1}H$. Because $(aH)H = aH$, $HH = H$ and $(aH)(aH) = a^2H = H$, the quotient group $C_{2n}/H \cong C_2$.

Example. Consider $G = SU(2)$, $H = \{1_2, -1_2\} \cong Z_2$. $A1_2 = 1_2A$ for all $A \in SU(2)$, hence H is a normal subgroup. One can show that the quotient group $G/H = SU(2)/Z_2$ is isomorphic with $SO(3, R)$. This is an important result for quantum mechanics, we will analyze it more in a future problem set.

This is also an example of a *center*. A **center** of a group G is the set of all elements of $g' \in G$ which commute with every element $g \in G$. In other words, it is the set $\{g' \in G|\ g'g = gg'\ \forall g \in G\}$. You can show that a center is a normal subgroup, so the quotient of a group and its center is a group. The center of $SU(2)$ is $\{1_2, -1_2\}$.

We finish by showing another way of finding normal subgroups and quotient groups. Let the map $\mu : G_1 \to G_2$ be a group homomorphism. Its **image** is the set

$$Im\mu = \{g_2 \in G_2|\ \exists g_1 \in G_1\ s.t.\ g_2 = \mu(g_1)\}$$

and its **kernel** is the set

$$Ker\mu = \{g_1 \in G_1|\ \mu(g_1) = e_2\}\ .$$

In other words, the kernel is the set of all elements of G_1 which map to the unit element of G_2. You can show that $Im\mu$ is a subgroup of G_2, $Ker\mu$ a subgroup of G_1. Further, $Ker\mu$ is a normal subgroup: if $k \in Ker\mu$ then $\mu(gkg^{-1}) = \mu(g)e_2\mu(g^{-1}) = \mu(gg^{-1}) = \mu(e_1) = e_2$ *i.e.* $gkg^{-1} \in Ker\mu$. Hence $G_1/Ker\mu$ is a quotient group. In fact, it also isomorphic with $Im\mu$!

Theorem 2.4 $G_1/Ker\mu \cong Im\mu$.

Proof. Denote $K \equiv Ker\mu$. Define $i : G_1/K \to Im\mu$, $i(gK) = \mu(g)$. If $gK = g'K$ then there is a $k \in K$ s.t. $g = g'k$. Then $i(gK) = \mu(g) = \mu(g'k) = \mu(g')e_2 = i(g'K)$ so i is well defined. Injection: if $i(gK) = i(g'K)$ then $\mu(g) = \mu(g')$ so $e_2 = (\mu(g))^{-1}\mu(g') = \mu(g^{-1})\mu(g') = \mu(g^{-1}g')$ so $g^{-1}g' \in K$. Hence $\exists k \in K$ s.t. $g' = gk$ so $g'K = gK$. Surjection: i is a surjection by definition. Thus i is a bijection. Homomorphism: $i(gKg'K) = i(gg'K) = \mu(gg') = \mu(g)\mu(g') = i(gK)i(g'K)$. i is a homomorphism and a bijection, *i.e.* an isomorphism.

For example, our previous example $SU(2)/Z_2 \cong SO(3, R)$ can be shown this way, by constructing a surjective homomorphism $\mu : SU(2) \to SO(3, R)$ such that $Ker\mu = \{1_2, -1_2\}$.

3 Representation Theory of Groups

In the previous section we discussed the action of a group on a set. We also listed some examples of Lie groups, their elements being $n \times n$ matrices. For example, the elements of the orthogonal group $O(n, R)$ corresponded to rotations of vectors in R^n. Now we are going to continue along these lines and consider the action of a generic group on a (complex) vector space, so that we can represent the elements of the group by matrices. However, a vector space is more than just a set, so in defining the action of a group on it, we have to ensure that it respects the vector space structure.

3.1 Complex Vector Spaces and Representations

Definition. A complex **vector space** V is an Abelian group (we denote its multiplication by "+" and call it a sum), where an additional operation, **scalar multiplication** by a complex number $\mu \in C$ has been defined, such that the following conditions are satisfied:

i) $\mu(\vec{v}_1 + \vec{v}_2) = \mu\vec{v}_1 + \mu\vec{v}_2$

ii) $(\mu_1 + \mu_2)\vec{v} = \mu_1\vec{v} + \mu_2\vec{v}$

iii) $\mu_1(\mu_2\vec{v}) = (\mu_1\mu_2)\vec{v}$

iv) $1\ \vec{v} = \vec{v}$

v) $0\ \vec{v} = \vec{0}$ ($\vec{0}$ is the unit element of V)

We could have replaced complex numbers by real numbers, to define a real vector space, or in general replaced the set of scalars by something called a "field". Complex vector spaces are relevant for quantum mechanics. A comment on notations: we denote vectors with arrows: $\vec{v}$, but textbooks written in English often denote them in boldface: **v**. If it is clear from the context whether one means a vector or its component, one may also simply use the notation v for a vector.

Definition. Vectors $\vec{v}_1, \ldots, \vec{v}_n \in V$ are **linearly independent**, if $\sum_{i=1}^{n} \mu_i \vec{v}_i = \vec{0}$ only if the coefficients $\mu_1 = \mu_2 = \cdots = \mu_n = 0$. If there exist at most n linearly independent vectors, n is the **dimension** of V, we denote $\dim V = n$. If $\dim V = n$, a set $\{\vec{e}^1, \ldots, \vec{e}^n\}$ of linearly independent vectors is called a **basis** of the vector space. Given a basis, any vector $\vec{v}$ can be written in a form $\vec{v} = \sum_{i=1}^{n} v_i \vec{e}^i$, where the **components** v_i of the vector are found uniquely.

Definition. A map $L : V_1 \to V_2$ between two vector spaces V_1, V_2 is **linear**, if it satisfies

$$L(\mu_1 \vec{v}_1 + \mu_2 \vec{v}_2) = \mu_1 L(\vec{v}_1) + \mu_2 L(\vec{v}_2)$$

for all $\mu_1, \mu_2 \in C$ and $\vec{v}_1, \vec{v}_2 \in V$. A linear map is also called a **linear transformation**, or especially in physics context, a **(linear) operator**. If a linear map is also a bijection, it is called an **isomorphism**, then the vector spaces V_1 and V_2 are **isomorphic**, $V_1 \cong V_2$. It then follows that $\dim V_1 = \dim V_2$. Further, all n-dimensional vector spaces are isomorphic. An isomorphism from V to itself is called an **automorphism**. The set of automorphisms of V is denoted $\mathrm{Aut}(V)$. It is a group, with composition of mappings $L \circ L'$ as the law of multiplication. (Existence of inverse is guaranteed since automorphisms are bijections).

Definition. The **image** of a linear transformation is

$$\mathrm{im} L = f(V_1) = \{L(\vec{v}_1) |\; \vec{v}_1 \in V_1\} \subset V_2$$

and its **kernel** is the set of vectors of V_1 which map to the null vector $\vec{0}_2$ of V_2:

$$\ker L = \{\vec{v}_1 \in V_1 |\; L(\vec{v}_1) = \vec{0}_2\} \subset V_1 \;.$$

You can show that both the image and the kernel are vector spaces. I also quote a couple of theorems without proofs.

Theorem 3.1 $\dim V_1 = \dim \ker L + \dim \mathrm{im} L$.

Theorem 3.2 *A linear map $L : V \to V$ is an automorphism if and only if* $\ker L = \{\vec{0}\}$.

Note that a linear map is defined uniquely by its action on the basis vectors:

$$L(\vec{v}) = L(\sum_{i=1}^{n} v_i \vec{e}^{\,i}) = \sum_i v_i L(\vec{e}^{\,i})$$

then we expand the vectors $L(\vec{e}^{\,i})$ in the basis $\{\vec{e}^{\,j}\}$ and denote the components by L_{ji}:

$$L(\vec{e}^{\,i}) = \sum_j L_{ji} \vec{e}^{\,j}.$$

Now

$$L(\vec{v}) = \sum_i \sum_j v_i L_{ji} \vec{e}^{\,j} = \sum_j \left(\sum_i L_{ji} v_i \right) \vec{e}^{\,j} \;,$$

so the image vector $L(\vec{v})$ has the components $L(\vec{v})_j = \sum_i L_{ji} v_i$. Let $\dim V_1 = \dim V_2 = n$. The above can be written in the familiar matrix language:

$$\begin{pmatrix} L(\vec{v})_1 \\ L(\vec{v})_2 \\ \vdots \\ L(\vec{v})_n \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & \cdots & L_{1n} \\ L_{21} & L_{22} & \cdots & L_{nn} \\ \vdots & & \ddots & \vdots \\ L_{n1} & \cdots & & L_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} .$$

We will often shorten the notation for linear maps and write $L\vec{v}$ instead of $L(\vec{v})$, and $L_1 L_2 \vec{v}$ instead of $L_1(L_2(\vec{v}))$. From the above it should also be clear that the group of automorphisms of V is isomorphic with the group of invertible $n \times n$ complex matrices:

$$\mathrm{Aut}(V) = \{L : V \to V |\ L \text{ is an automorphism}\} \cong GL(n, C) \ .$$

(The multiplication laws are composition of maps and matrix multiplication.)

Now we have the tools to give a definition of a representation of a group. The idea is that we define the action of a group G on a vector space V. If V were just a set, we would associate with every group element $g \in G$ a permutation $L_g \in Perm(V)$. However, we have to preserve the vector space structure of V. So we define the action just as before, but replace the group $Perm(V)$ of permutations of V by the group $\mathrm{Aut}(V)$ of automorphisms of V.

Definition. A (linear) representation of a group G in a vector space V is a homomorphism $D : G \to \mathrm{Aut}(V)$, $G \ni g \mapsto D(g) \in \mathrm{Aut}(V)$. The **dimension of the representation** is the dimension of the vector space $\dim V$.

Note:

1. D is a homomorphism: $D(g_1 g_2) = D(g_1) D(g_2)$.
2. $D(g^{-1}) = (D(g))^{-1}$.

Example. Let $G = C_4 = \{e, c, c^2, c^3\}$ and $V = \mathbb{R}^2$. One possible representation of G is $D : G \to \mathrm{Aut}(V)$,

$$D(c) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, D(e) = D(c^4) = (D(c))^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1}.$$

Note that the matrix $D(c)$ corresponds to a 90° rotation in the $\mathbb{R}^2$ plane.

We say that a representation D is **faithful** if $\mathrm{Ker}D = \{e\}$. Then $g_1 \neq g_2 \Rightarrow D(g_1) \neq D(g_2)$. Whatever the $\mathrm{Ker}D$ is, D is always a faithful representation of the quotient group $G/\mathrm{Ker}D$.

A mathematician would next like to classify all possible representations of a group. Then the first question is when two representations are the same (equivalent).

Definition. Let D_1, D_2 be representations of a group G in vector spaces V_1, V_2. An **intertwining operator** is a linear map $A : V_1 \to V_2$ such that the diagram

$$\begin{array}{ccc} V_1 & \stackrel{A}{\to} & V_2 \\ D_1(g)\downarrow & \searrow & \downarrow D_2(g) \\ V_1 & \stackrel{A}{\to} & V_2 \end{array}$$

commutes, *i.e.* $D_2(g)A = AD_1(g)$ for all $g \in G$. If A is an isomorphism (we then need $\dim V_1 = \dim V_2$), the representations D_1 and D_2 are **equivalent**. In other words, there then exists a **similarity transformation** $D_2(g) = AD_1(g)A^{-1}$ for all $g \in G$.

Example. Let $\dim V_1 = n$, $V_2 = C^n$. Thus any n-dimensional representation is equivalent with a representation of G by invertible complex matrices, the homomorphism $D_2 : G \to GL(n, C)$.

Definition. A **scalar product** in a vector space V is a map $V \times V \to C$, $(\vec{v}_1, \vec{v}_2) \mapsto \langle \vec{v}_1 | \vec{v}_2 \rangle \in C$ which satisfies the following properties:

i) $\langle \vec{v} | \mu_1 \vec{v}_1 + \mu_2 \vec{v}_2 \rangle = \mu_1 \langle \vec{v} | \vec{v}_1 \rangle + \mu_2 \langle \vec{v} | \vec{v}_2 \rangle$

ii) $\langle \vec{v} | \vec{w} \rangle = \langle \vec{w} | \vec{v} \rangle^*$

iii) $\langle \vec{v} | \vec{v} \rangle \geq 0$ and $\langle \vec{v} | \vec{v} \rangle = 0 \leftrightarrow \vec{v} = \vec{0}$.

Given a scalar product, it is possible to normalize (*e.g.* by the Gram-Schmidt method) the basis vectors such that $\langle \vec{e}^i | \vec{e}^j \rangle = \delta^{ij}$. Such an **orthonormal** basis is usually the most convenient on to use. The **adjoint** $A^\dagger$ of an operator (linear map) $A : V \to V$ is the one which satisfies $\langle \vec{v} | A^\dagger \vec{w} \rangle = \langle A\vec{v} | \vec{w} \rangle$ for all $\vec{v}, \vec{w} \in V$.

Definition. An operator (linear map) $U : V \to V$ is **unitary** if $\langle \vec{v} | \vec{w} \rangle = \langle U\vec{v} | U\vec{w} \rangle$ for all $\vec{v}, \vec{w} \in V$. Equivalently, a unitary operator must satisfy $U^\dagger U = id_V = 1$. It follows that the corresponding $n \times n$ matrix must be unitary, *i.e.* an element of $U(n)$. Unitary operators form a subgroup $\mathrm{Unit}(V)$ of $\mathrm{Aut}(V) \cong GL(n, C)$.

Definition. An **unitary representation** of a group G is a homomorphism $D : G \to \mathrm{Unit}(V)$.

Definition. If U_1, U_2 are unitary representations of G in V_1, V_2, and there exists an intertwining isomorphic operator $A : V_1 \to V_2$ which preserves the scalar product, $\langle A\vec{v} | A\vec{w} \rangle_{V_2} = \langle \vec{v} | \vec{w} \rangle_{V_1}$ for all $\vec{v}, \vec{w} \in V_1$, the represenations are **unitarily equivalent**.

Example. Every n-dimensional unitary representation is unitarily equivalent with a representation by unitary matrices, a homomorphism $G \to U(n)$.

As always after defining a fundamental concept, we would like to classify all possibilities. The basic problem in group representation theory is to classify all unitary representations of a group, up to unitary equivalence.

3.2 Symmetry Transformations in Quantum Mechanics

We have been aiming at unitary representations in complex vector spaces because of their applications in Quantum Mechanics (QM). Recall that the set of all possible states of a quantum mechanical system is the Hilbert space $\mathcal{H}$, a complex vector space with a scalar product. State vectors are usually denoted by $|\psi\rangle$ as opposed to our previous notation $\vec{v}$, and the scalar product of two vectors $|\psi\rangle, |\chi\rangle$ is denoted $\langle\psi|\chi\rangle$. Note that usually the Hilbert space is an infinite dimensional vector space, whereas in our discussion of representation theory we've been focusing on finite dimensional vector spaces. Let's not be concerned about the possible subtleties which ensue, in fact in many cases finite dimensional representations will still be relevant, as you will see.

According to QM, the time evolution of a state is controlled by the Schrodinger equation,

$$\imath\hbar\frac{d}{dt}|\psi\rangle = H|\psi\rangle$$

where H is the Hamilton operator, the time evolution operator of the system. Suppose that the system possesses a symmetry, with the symmetry operations forming a group G. In order to describe the symmetry, we need to specify how it acts on the state vectors of the system – we need to find its representation in the vector space of the states, the Hilbert space. The norm of a state vector, its scalar product with itself $\langle\psi|\psi\rangle$ is associated with a probability density and normalized to one, similarly the scalar product $\langle\psi|\chi\rangle$ of two states is associated with the probability (density) of measurements. Thus the representations of the symmetry group G must preserve the scalar product. In other words, the representations must be unitary. Moreover, in a closed system probability is preserved under the time evolution. Thus, unitarity of the representations must also be preserved under the time evolution.

We can summarize the above in a more formal way: if $g \mapsto U_g$ is a faithful unitary representation of a group G in the Hilbert space of a quantum mechanical system, such that for all $g \in G$

$$U_g H U_g^{-1} = H \tag{5}$$

where H is the Hamilton operator of the system, the group G is a **symmetry group** of the system.

The condition (5) arises as follows. Suppose a state vector $|\psi\rangle$ is a solution of the Schrodinger equation. In performing a symmetry operation on the system, the state vector is mapped to a new vector $U_g|\psi\rangle$. But if the system is symmetric, the new state $U_g|\psi\rangle$ must also be a solution of the Schrodinger equation: $i\hbar(d/dt)U_g|\psi\rangle = HU_g|\psi\rangle$). But then it must be $i\hbar(d/dt)|\psi\rangle = i\hbar(d/dt)U_g^{-1}U_g|\psi\rangle = U_g^{-1}HU_g|\psi\rangle = H|\psi\rangle \Rightarrow U_g^{-1}HU_g = H$.

Consider in particular the energy eigenstates $|\phi_n\rangle$ at energy level E_n:

$$H|\phi_n\rangle = E_n|\phi_n\rangle \ .$$

An energy level may be degenerate, say with k linearly independent energy eigenstates $\{|\phi_{n1}, \ldots, |\phi_{nk}\rangle\}$. They span a k-dimensional vector space $\mathcal{H}_n$, a subspace of the full Hilbert space. If the system has a symmetry group,

$$HU_g|\phi_n\rangle = U_gH|\phi_n\rangle = E_nU_g|\phi_n\rangle$$

so all states $U_g|\phi_n\rangle$ are eigenstates at the same energy level E_n. Thus the representation U_g maps the eigenspace $\mathcal{H}_n$ to itself; in other words the representation U_g is a k-dimensional representation of G acting in $\mathcal{H}_n$. By an inverse argument, suppose that the system has a symmetry group G. Its representations then determine the possible degeneracies of the energy levels of the system.

3.3 Reducibility of Representations

It turns out that some representations are more fundamental than others. A generic representation can be decomposed into so-called irreducible representations. That is our next topic. Again, we start with some definitions.

Definition. A subset W of a vector space V is called a **subspace** if it includes all possible linear combinations of its elements: if $\vec{v}, \vec{w} \in W$ then $\lambda\vec{v} + \mu\vec{w} \in W$ for all $\lambda, \mu \in C$.

Let D be a representation of a group G in vector space V. The representation space V is also called a **G-module**. (This terminology is used in Jones.) Let W be a subspace of V. We say that W is a **submodule** if it is closed under the action of the group G: $\vec{w} \in W \Rightarrow D(g)\vec{w} \in W$ for all $g \in G$. Then, the restriction of $D(g)$ in W is an automorphism $D(g)_W : W \to W$.

Definition. A representation $D : G \to \mathrm{Aut}(V)$ is **irreducible**, if the only submodules are $\{\vec{0}\}$ and V. Otherwise the representation is **reducible**.

Example. Choose a basis $\{\vec{e}^i\}$ in V, let $\dim V = n$. Suppose that all the matrices $D(g)_{ij} = \langle\vec{e}^i|D(g)ve^j\rangle$ turn out to have the form

$$D(g) = \begin{pmatrix} M(g) & S(g) \\ 0 & T(g) \end{pmatrix} \tag{6}$$

where $M(g)$ is a $n_1 \times n_1$ matrix, $T(g)$ is a $n_2 \times n_2$ matrix, $n_1 + n_2 = n$, and $S(g)$ is a $n_1 \times n_2$ matrix. Then the representation is reducible, since

$$W = \left\{ \begin{pmatrix} \vec{v} \\ \vec{0} \end{pmatrix} \mid \vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_{n_1} \end{pmatrix} \right\} \tag{7}$$

is a submodule:

$$D(g) \begin{pmatrix} \vec{v} \\ \vec{0} \end{pmatrix} = \begin{pmatrix} M(g)\vec{v} + S(g)\vec{0} \\ T(g)\vec{0} \end{pmatrix} = \begin{pmatrix} M(g)\vec{v} \\ \vec{0} \end{pmatrix} \in W. \tag{8}$$

If in addition $S(g) = 0$ for all $g \in G$, the representation is obviously built up by combining two representations $M(g)$ and $T(g)$. It is then an example of a *completely reducible* representation. We'll give a formal definition shortly.

Definition. A **direct sum** $V_1 \oplus V_2$ of two vector spaces V_1 and V_2 consists of all pairs (v_1, v_2) with $v_1 \in V_1, v_2 \in V_2$, with the addition of vectors and scalar multiplication defined as

$$\begin{aligned} (v_1, v_2) + (v_1', v_2') &= (v_1 + v_1', v_2 + v_2') \\ \lambda(v_1, v_2) = (\lambda v_1, \lambda v_2) \end{aligned}$$

It is simple to show that $\dim(V_1 \oplus V_2) = \dim V_1 + \dim V_2$. If a scalar product has been defined in V_1 and V_2, one can define a scalar product in $V_1 \oplus V_2$ by

$$\langle (v_1, v_2) | (v_1', v_2') \rangle = \langle v_1, v_1' \rangle + \langle v_2 | v_2' \rangle \ .$$

Suppose D_1, D_2 are representations of G in V_1, V_2, one can then define a **direct sum** representation $D_1 \oplus D_2$ in $V_1 \oplus V_2$:

$$(D_1 \oplus D_2)(g)(v_1, v_2) = (D_1(g)v_1, D_2(g)v_2) \ .$$

In this case it is useful to adopt the notation

$$V_1 = \left\{ \begin{pmatrix} \vec{v}_1 \\ \vec{0} \end{pmatrix} \right\} \ ; \ V_2 = \left\{ \begin{pmatrix} \vec{0} \\ \vec{v}_2 \end{pmatrix} \right\}$$

so that

$$V_1 \oplus V_2 = \left\{ \begin{pmatrix} \vec{v}_1 \\ \vec{v}_2 \end{pmatrix} \right\} = \{ (\vec{v}_1, \vec{v}_2) \} \ .$$

Now the matrices of the direct sum representation are of the block diagonal form

$$(D_1 \oplus D_2)(g) = \begin{pmatrix} D_1(g) & 0 \\ 0 & D_2(g) \end{pmatrix} \ .$$

Definition. A representation D in vector space V is **completely reducible** if for every submodule $W \subset V$ there exists a *complementary submodule* W' such that $V = W \oplus W'$ and $D \cong D_W \oplus D_{W'}$.

Comments.

1. According to the definition, we need to show that D is equivalent with the direct sum representation $D_W \oplus D_{W'}$. For the matrices of the representation, this means that there must be a similarity transformation which maps all the matrices $D(g)$ into a block diagonal form:

$$AD(g)A^{-1} = \begin{pmatrix} D_W(g) & 0 \\ 0 & D_{W'}(g) \end{pmatrix} .$$

2. Strictly speaking, according to the definition also an irreducible representation is completely reducible, as $W = V, W' = \{0\}$ or vice versa satisfy the requirements. We will exclude this case, and from now on by completely reducible representations we mean those which are not irreducible.

The goal in the **reduction** of a representation is to decompose it into irreducible pieces, such that

$$D \cong D_1 \oplus D_2 \oplus D_3 \oplus \cdots$$

(then $\dim D = \sum_i \dim D_i$). This is possible if D is completely reducible. So, given a representation, how do we know if it is completely reducible or not? Interesting representations from quantum mechanics point of view turn out to be completely reducible:

Theorem 3.3 *Unitary representations are completely reducible.*

Proof. Since we are talking about unitary representations, it is implied that the representation space V has a scalar product. Let W be a submodule. We define its *orthogonal complement* $W_\perp = \{\vec{v} \in V | \ \langle \vec{v} | \vec{w} \rangle = 0 \ \forall \vec{w} \in W\}$. I leave it as an excercise to show that $V \cong W \oplus W_\perp$. We then only need to show that $W_\perp$ is also a submodule (closed under the action of G). Let $\vec{v} \in W_\perp$, and denote the unitary representation by U. For all $\vec{w} \in W$ and $g \in G$ $\langle U(g)\vec{v} | \vec{w} \rangle = \langle U(g)\vec{v} | U(g)U^{-1}(g)\vec{w} \rangle = \langle \vec{v} | U^\dagger(g)U(g)U^{-1}(g)\vec{w} \rangle \stackrel{a}{=} \langle \vec{v} | U^{-1}(g)\vec{w} \rangle = \langle \vec{v} | U(g^{-1})\vec{w} \rangle \stackrel{b}{=} \langle \vec{v} | \vec{w}' \rangle \stackrel{c}{=} 0$, where the step a follows since U is unitary, step b since W is a G-module, and the step c is true since $\vec{v} \in W_\perp$. Thus $U(g)\vec{v} \in W_\perp$ so $W_\perp$ is closed under the action of G.

If G is a finite group, we can say more.

Theorem 3.4 *Let D be a finite dimensional representation of a finite group G, in vector space V. Then there exists a scalar product in V such that D is unitary.*

Proof. We can always define a scalar product in a finite dimensional vector space, *e.g.* by choosing a basis and defining $\langle\vec{v}|\vec{w}\rangle = \sum_{i=1}^{n} v_i^* w_i$ where v_i, w_i are the components of the vectors. Given a scalar product, we then define a "group averaged" scalar product $\langle\langle\vec{v}|\vec{w}\rangle\rangle = \frac{1}{|G|}\sum_{g'\in G}\langle D(g')\vec{v}|D(g')\vec{w}\rangle$. It is straightforward to show that $\langle\langle|\rangle\rangle$ satisfies the requirements of a scalar product. Further,

$$\begin{aligned}
\langle\langle D(g)\vec{v}|D(g)\vec{w}\rangle\rangle &= \frac{1}{|G|}\sum_{g'\in G}\langle D(g')D(g)\vec{v}|D(g')D(g)\vec{w}\rangle \\
&= \frac{1}{|G|}\sum_{g'\in G}\langle D(g'g)\vec{v}|D(g'g)\vec{w}\rangle \\
&= \frac{1}{|G|}\sum_{g''\in G}\langle D(g'')\vec{v}|D(g'')\vec{w}\rangle = \langle\langle\vec{v}|\vec{w}\rangle\rangle \ .
\end{aligned}$$

In other words, D is unitary with respect to the scalar product $\langle\langle|\rangle\rangle$.

Since we have previously shown that unitary representations are completely reducible, we have shown the following fact, called Maschke's theorem.

Theorem 3.5 (Maschke's Theorem) *Every finite dimensional representation of a finite group is completely reducible.*

3.4 Irreducible Representations

Now that we have shown that many representations of interest are completely reducible, and can be decomposed into a direct sum of irreducible representations, the next task is to classify the latter. We will first develop ways to identify inequivalent irreducible representations. Before doing so, we must discuss some general theorems.

Theorem 3.6 (Schur's Lemma) *Let D_1 and D_2 be two irreducible representations of a group G. Every intertwining operator between them is either a null map or an isomorphism; in the latter case the representations are equivalent, $D_1 \cong D_2$.*

Proof. Let A be an intertwining operator between the representations, *i.e.* the diagram

$$\begin{array}{ccc}
V_1 & \stackrel{A}{\to} & V_2 \\
D_1(g)\downarrow & \searrow & \downarrow D_2(g) \\
V_1 & \stackrel{A}{\to} & V_2
\end{array}$$

commutes: $D_2(g)A = AD_1(g)$ for all $g \in G$. Let's first examine if A can be an injection. Note first that if $\mathrm{Ker}A \equiv \{\vec{v}\in V_1|\ A\vec{v} = \vec{0}_2\} = \{\vec{0}_1\}$, then A is an injection since if $A\vec{v} = A\vec{w}$ then $A(\vec{v}-\vec{w}) = 0 \Rightarrow \vec{v}-\vec{w} \in \mathrm{Ker}A = \{\vec{0}_1\} \Rightarrow \vec{v} = \vec{w}$. So what is $\mathrm{Ker}A$? Recall that $\mathrm{Ker}A$ is a subspace of V_1. Is it also a submodule, *i.e.* closed under the action of G? Let $\vec{v} \in \mathrm{Ker}A$. Then $AD_1(g)\vec{v} = D_2(g)A\vec{v} = \vec{0}_2$, hence $D_1(g)\vec{v} \in \text{Ker}A$ *i.e.* $\text{Ker}A$ is a submodule. But since D_1 is an irreducible representation, either $\text{Ker}A = V_1$ or $\text{Ker}A = \{\vec{0}_1\}$. In the former case all vectors of V_1 map to the null vector of V_2, so A is a null map $A = 0$. In the latter case, A is an injection. We then use a similar reasoning to examine if A is also a surjection. Let $\vec{v}_2 \in \text{Im}A \equiv \{\vec{v} \in V_2 |\ \exists \vec{v}_1 \in V_1 \text{ s.t. } \vec{v} = A\vec{v}_1\}$. Then we can write $\vec{v}_2 = A\vec{v}_1$. Then $D_2(g)\vec{v}_2 = D_2(g)A\vec{v}_1 = A(D_1(g)\vec{v}_1)$ so also $D_2(g)\vec{v}_2 \in \text{Im}A$. Thus, $\text{Im}A$ is a submodule of V_2. But since D_2 is irreducible, either $\text{Im}A = \{\vec{0}_2\}$ *i.e.* $A = 0$, or $\text{Im}A = V_2$ *i.e.* A is a surjection. To summarize, either $A = 0$ or A is a bijection *i.e.* an isomorphism (since it is also a linear operator).

Corollary. If D is an irreducible representation of a group G in (complex) vector space V, then the only operator which commutes with all $D(g)$ is a multiple of the identity operator.

Proof. If $\forall g \in G\ AD(g) = D(g)A$, then for all $\mu \in C$ also $(A - \mu 1)D(g) = D(g)(A - \mu 1)$. According to Schur's lemma, either $(A - \mu 1)^{-1}$ exists for all $\mu \in C$or $(A - \mu 1) = 0$. However, it is always possible to find at least one $\mu \in C$ such that $(A - \mu 1)$ is not invertible. In the finite dimensional case this is follows from the fundamental theorem of algebra, which guarantees that the polynomial equation $\det(A - \mu 1) = 0$ has solutions for μ. (The infinite dimensional case is more delicate, but turns out to be true as well). So it must be $A = \mu 1$.

We will next discuss a sequence of theorems, starting from the rather abstract *fundamental orthogonality theorem* and then moving towards its more intuitive and user-friendly forms.

Theorem 3.7 (Fundamental Orthogonality Theorem) *Let D^1 and D^2 be two unitary irreducible representations of a group G in vector spaces V_1 and V_2. Fix basis $\{v_1^k, \ldots, v_{m_k}^k\}$ (with $k = 1, 2$) in the vector spaces V_k. Write a linear map $T : V_2 \to V_1$ as a matrix (T_{ij}) in these basis. Then*

$$\sum_{g \in G} D^1_{ij}(g) D^2_{kl}(g^{-1}) = \frac{n}{m_1} T_{il}(T^{-1})_{kj} \cdot \delta$$

where $n = |G|$, the representations $D^k(g)$ are written as matrices in the corresponding basis, and

$$\delta = \begin{cases} 0 \text{ when } D^1, D^2 \text{ are nonequivalent} \\ 1 \text{ when } D^2(g) = T^{-1} D^1(g) T \text{ for all } g \in G. \end{cases} .$$

Proof. 1) Assume that D^1, D^2 are nonequivalent. Let M be an arbitrary $m_1 \times m_2$ matrix and set

$$F = \sum_g D^1(g) M D^2(g^{-1}).$$

Then $D^1F = FD^2$ by a direct calculation and so $F = 0$ by Schur's Lemma. Choosing now $M_{ab} = \delta_{aj}\delta_{bk}$ and taking the matrix element $F_{il} = 0$ gives the claim in the nonequivalent case.

2) Let then $D^2(g) = T^{-1}D^1(g)T$ for all $g \in G$ for a linear isomorphism $T : V_2 \to V_1$. In this case $m_1 = m_2$. Let us define the matrices M, F as in the case (1). Now

$$(FT^{-1})D^1 = FD^2T^{-1} = D^1(FT^{-1})$$

and so again by Schur's Lemma $FT^{-1} = \lambda \cdot \mathbf{1}$. We write $\lambda = \lambda_{jk}$ since the constant λ depends on the choice of indices in the definition of M. Now we have

$$\lambda_{jk}T_{il} = \sum_g D^1_{ij}(g)D^2_{kl}(g^{-1}).$$

Multiplying this equation with $(T^{-1})_{li}$ and summing over i, l we get

$$\begin{aligned}\lambda_{jk}m_1 = & \quad \textstyle\sum_g \sum_{i,l}(T^{-1})_{li}D^1_{ij}(g)D^2_{kl}(g^{-1}) \\ = & \quad \textstyle\sum_g \sum_l (D^2(g)T^{-1})_{lj}D^2_{kl}(g^-)) \\ = & \quad \textstyle\sum_g (D^2(g^{-1})D^2(g)T^{-1})_{kj} = n \cdot (T^{-1})_{kj}.\end{aligned}$$

In the unitary case $D^2_{kl}(g^{-1}) = \overline{D^2}_{lk}(g)$ and the left hand side can be interpreted as a scalar product of two vectors, then the right hand side is an orthogonality relation for them. Namely, consider a given representation (labeled by α), and the ijth elements of its representation matrices. They form a $|G|$-component vector $(D^{(\alpha)}_{ij}(g_1), D^{(\alpha)}_{ij}(g_2), \ldots, D^{(\alpha)}_{ij}(g_{|G|}))$ where g_i are all the elements of the group G. So we have a collection of vectors, labeled by α, i, j. Then the case (1) is an orthogonality relation for the vectors, with respect to the scalar product $\langle \vec{v}|\vec{v}'\rangle = \sum_{i=1}^{|G|} v_i^* v_i'$. However, in a $|G|$ dimensional vector space there can be at most $|G|$ mutually orthogonal vectors. The index pair ij has $(\dim D^{(\alpha)})^2$ possible values, so the upper bound on the total number of the above vectors is

$$\sum_\alpha (\dim D^{(\alpha)})^2 \le |G| \ ,$$

where the sum is taken over all possible unitary inequivalent representations (labeled by α). In fact, the sum turns out to be equal to the order $|G|$. This theorem is due to Burnside:

$$\sum_\alpha (\dim D^{(\alpha)})^2 = |G|.$$

We shall prove it later.

Burnside's theorem helps to rule out possibilities for irreducible representations. Consider *e.g.* $G = S_3$, $|S_3| = 6$. The possible dimensions of inequivalent irreducible representations are 2,1,1 or 1,1,1,1,1,1. It turns out that S_3 has only three inequivalent irreducible representations (show it). So the irreps have dimensions 2,1,1.

3.5 Characters

Characters are a convenient way to classify inequivalent irreducible representations.

To start with, let $\{\vec{e}^1, \ldots, \vec{e}^n\}$ be an orthonormal basis in a n-dimensional vector space V with respect to scalar product $\langle | \rangle$.

Definition. A **trace** of a linear operator A is

$$\text{tr } A \equiv \sum_{i=1}^{n} \langle \vec{e}^i | A \vec{e}^i \rangle \ .$$

Note. Trace is well defined, since it is independent of a choice of basis. Let $\{\vec{e}'^1, \ldots, \vec{e}'^n\}$ be another basis. Then $\text{tr } A = \sum_i \langle \vec{e}^i | A\vec{e}^i \rangle = \sum_{ij} \langle \vec{e}^i | \vec{e}'^j \rangle \langle \vec{e}'^j | A \vec{e}^i \rangle = \sum_{ij} \langle A^\dagger \vec{e}'^j | \vec{e}^i \rangle \langle \vec{e}^i | \vec{e}'^j \rangle = \sum_{ij} \langle A^\dagger \vec{e}'^j | e'^j \rangle = \sum_j \langle \vec{e}'^j | A \vec{e}'^j \rangle$. Recall also that associated with the operator A is a $n \times n$ matrix with components $A_{ij} = \langle \vec{e}^i | A \vec{e}^j \rangle$. Thus tr A is equal to the trace of the matrix.

Now, let $D^{(\alpha)}(g)$ be an unitary representation of a finite group G in V.

Definition. The **character** of the representation $D^{(\alpha)}$ is the map

$$\chi^{(\alpha)} : G \to C, \ \chi^{(\alpha)}(g) = \text{tr } D^{(\alpha)}(g) \ .$$

Note. Equivalent representations have the same characters: $\text{tr } (AD^{(\alpha)}A^{-1}) = \text{tr } (A^{-1}AD^{(\alpha)}) = \text{tr } D^{(\alpha)}$, where we used cyclicity of the trace: $\text{tr } ABC = \text{tr } CAB = \text{tr } BCA$ etc.

Recall that conjugation $L_g(g_0) = gg_0g^{-1}$ is one way to define how G acts on itself, the orbits $\{gg_0g^{-1} | \ g \in G\}$ were called conjugacy classes. Since $\text{tr } D(gg_0g^{-1}) = \text{tr } (D(g)D(g_0)D^{-1}(g)) = \text{tr } D(g_0)$, group elements related by conjugation have the same character (again, use cyclicity of trace). So characters can be interpreted as mappings

$$\chi^{(\alpha)} : \{\text{conjugacy classes of } G\} \to C$$

Note also that the character of the unit element is the same as the dimension of the representation: $\chi^{(\alpha)}(e) = \text{tr } D^{(\alpha)}(e) = \text{tr } id_V = \dim V = \dim D^{(\alpha)}$.

Recall then the fundamental orthogonality theorem, in its basis-dependent form, Theorem 3.7. Now we are going to set $i = j, k = l$ in 3.7 and sum over i and k. The left hand side becomes

$$\sum_{g \in G} \sum_i \overline{D^{(\alpha)}}_{ii}(g) \sum_k D^{(\beta)}_{kk}(g) = \sum_{g \in G} \overline{\chi^{(\alpha)}}(g) \chi^{(\beta)}(g) \ .$$

The right hand side becomes

$$\frac{|G|}{\dim D^{(\alpha)}} \ \delta_{\alpha\beta} \sum_{ik} \delta_{ik}\delta_{ik} = \frac{|G|}{\dim D^{(\alpha)}} \ \delta_{\alpha\beta} \sum_i \delta_{ii} = |G| \ \delta_{\alpha\beta} \ .$$

We have derived an **orthogonality theorem for characters**:

$$\sum_{g\in G} \overline{\chi^{(\alpha)}}(g)\chi^{(\beta)}(g) = |G|\ \delta_{\alpha\beta}\ . \tag{9}$$

It can be used to analyze the reduction of a representation. In the reduction of a representation D, it may happen that an irreducible representation $D^{(\alpha)}$ appears multiple times in the the direct sum:

$$D = D^{(1)} \oplus D^{(1)} \oplus D^{(1)} \oplus D^{(2)} \oplus D^{(3)} \oplus \cdots$$

Then we shorten the notation and multiply each irreducible representation by an integer n_α to account for how many times $D^{(\alpha)}$ appears:

$$D = 3D^{(1)} \oplus D^{(2)} \oplus D^{(3)} \oplus \cdots = \bigoplus_\alpha n_\alpha D^{(\alpha)}\ .$$

n_α is called the **multiplicity** of the representation $D^{(\alpha)}$ in the decomposition. Since tr is a linear operation, obviously the characters of the representation satisfy

$$\chi = \sum_\alpha n_\alpha \chi^{(\alpha)}$$

with the same coefficients n_α. If we know the character χ of the reducible representation D, and all the characters $\chi^{(\alpha)}$ of the irreducible representations, we can calculate the multiplicities of each irreducible representation in the decomposition by using the orthogonality theorem of characters:

$$n_\alpha = \frac{1}{|G|}\sum_g \overline{\chi^{(\alpha)}}(g)\chi(g)\ .$$

Then, once we know all the multiplities, we know what is the decomposition of the representation D. In practise, characters of finite groups can be looked up from *character tables*. You can find them *e.g.* in *Atoms and Molecules*, by M. Weissbluth, pages 115-125. For more explanation of construction of character tables, see Jones, section 4.4. You will work out some character tables in a problem set.

Again, the orthogonality of characters can be interpreted as an orthogonality relation for vectors, with useful consequences. Let $C_1, C_2, \ldots, C_k$ be the conjugacy classes of G, denote the number of elements of C_i by $|C_i|$. Then (9) implies

$$\sum_{\{C_i\}} |C_i|\overline{\chi^{(\alpha)}}(C_i)\chi^{(\beta)}(C_i) = |G|\ \delta_{\alpha\beta}\ . \tag{10}$$

Consider then the vectors $\vec{v}_\alpha = (\sqrt{|C_1|}\chi^{(\alpha)}(C_1), \ldots, \sqrt{|C_k|}\chi^{(\alpha)}(C_k))$. The number of such vectors is the same as the number of irreducible representations. On the other hand, (10) tells that the vectors are mutually orthogonal, so the can be no more of them than the dimension of the vector space k, the number of conjugacy classes.

Theorem 3.8 *The number of nonequivalent unitary irreducible representations of a finite group is less or equal to the number of its conjugacy classes.*

If the group is Abelian, the conjugacy class of each element contains only the element itself: $gg_0g^{-1} = g_0gg^{-1} = g_0$. So the number of conjugacy classes is the same as the order of the group $|G|$, this is then also the upper bound of the number of unitary irreducible representations.

Theorem 3.9 *All unitary irreducible representations of an Abelian group are one dimensional.*

Proof. In a representation D of an abelian group $D(g)$ commutes with any other element $D(g')$, so by Schur's Lemma $D(g) = \lambda_g \cdot \mathbf{1}$ for some constant $\lambda_g \in \mathbb{C}$. But now any 1-dimensional subspace of the representation space is invariant, so the whole representation must be 1-dimensional by the irreducibility.

We introduce now an important algebraic tool in the representation theory of finite groups, namely *the group algebra.* Let G be any finite group and consider *formal linear combinations* $\sum_{g\in G} a_g \cdot g$ where $a_g \in \mathbb{C}$. Note that here a sum $g_1 + g_2$ does NOT mean a group multiplication even when G is abelian. Formal linear combinations can be added as $\sum_g a_g \cdot g + \sum_g b_g \cdot g = \sum_g (a_g + b_g) \cdot g$ and multiplied naturally by complex numbers. In addtion, we define a multiplication as

$$(\sum_g a_g \cdot g) \cdot (\sum_{g'} b_{g'} \cdot g') = \sum_{g,g'} a_g b_{g'} \cdot gg' = \sum_h (\sum_g a_g b_{g^{-1}h}) \cdot h$$

which by a direct computation is associative. We denote by $A(G)$ the group algebra of G. As a vector space, its dimension is equal to $|G|$.

Given a representation D of G we can extend it to a representation of the algebra $A(G)$ by setting

$$D(\sum_g a_g \cdot g) = \sum_g a_g D(g)$$

where the sum on the right-hand-side is simply interpreted as a sum of matrices. In this way representations of G and representations of $A(G)$ are in 1-1 correspondence. *The regular representation* of G is defined as the representation in $A(G)$ (viewed as a vector space!) defined by

$$D(g)x = D(g) \sum_h a_h \cdot h = gx = \sum_h a_h \cdot gh = \sum_h a_{g^{-1}h} \cdot h.$$

The dimension of D is $|G|$. This representation is faithful: For example, $D(g)e = ge = g = D(g')e$ if and only if $g = g'$. In general, it is reducible,

$$D = \oplus_i q_i D^{(i)}$$

where the $D^{(i)}$'s are nonequivalent irreducible representations and $q_i = 0, 1, 2, \ldots$. By the orthogonality of characters we know that

$$q_i = \frac{1}{|G|} \sum_g \overline{\chi(g)} \chi^{(i)}(g)$$

where χ is the character of D and $\chi^{(i)}$ is the character of $D^{(i)}$. Since $\chi(g) = 0$ when $g \neq e$ and $\chi(e) = |G|$ we find

$$q_i = \frac{1}{|G|} |G| \chi^{(i)}(e) = \operatorname{tr} D^{(i)}(e)$$

and thus q_i is equal to $m_i = \dim D^{(i)}$.

Theorem 3.10 *(Burnside's Theorem) Let m_i be the dimensions of the nonequivalent irreducible representations of a finite group G. Then*

$$\sum_i m_i^2 = |G|.$$

Proof. Using the decomposition

$$\chi = \sum q_i \chi^{(i)}$$

for the character of the regular representation we get

$$|G| = \frac{1}{|G|} \sum_g \overline{\chi(g)} \chi(g) = \frac{1}{|G|} \sum_g \sum_{i,j} q_i q_j \overline{\chi^{(i)}(g)} \chi^{(j)}(g) = \sum_{i,j} q_i q_j \delta_{i,j} = \sum_i q_i^2 = \sum_i m_i^2$$

according to the orthogonality relations of the characters.

Denote by $G_i \subset G$ the different conjugacy classes in G, with $i = 1, 2, \ldots, r$. By abuse of notation, we also denote $G_i = \sum_{g \in G_i} g \in A(G)$ the corresponding element in the group algebra. Since $aG_iG_ja^{-1} = aG_ia^{-1}aG_ia^{-1} = G_iG_j$ the product (in $A(G)$!) of two conjugacy classes can be written as

$$G_i G_j = \sum_k h_{ij}^k G_k$$

where the h_{ij}^k's are nonnegative integers. We define $G_i^{-1} = \sum_{g \in G_i} g^{-1}$ and so G_i^{-1} is an element in $A(G)$ corresponding to another conjugacy class $G_{i'}$. Denote by n_i the number of elements in the conjugacy class G_i; clearly $n_i = n_{i'}$. Denote by G_1 the conjugacy class containing only the neutral element e. Now $G_iG_{i'}$ contains the element e exactly n_i times, so $h_{ii'}^1 = n_i$. On the other hand, when $j \neq i'$ then G_iG_j does not contain the element e and so $h_{ij}^1 = 0$ for $j \neq i'$. For an irreducible representation $D^{(i)}$ we set

$$T_j^{(i)} = \sum_{g \in G_j} D^{(i)}(g) = D^{(i)}(G_j).$$

By the representation property $T_j^{(i)} D^{(i)}(g) = D^{(i)}(g) T_j^{(i)}$ for all $g \in G$ and thus by irreducibility

$$T_j^{(i)} = \lambda_j \cdot \mathbf{1}$$

for some complex number λ_j. Since $\chi^{i)}(g) = \chi^{(i)}(h)$ when g, h are in the same conjugacy class, we have

$$\operatorname{tr} T_j^{(i)} = m_i \cdot \lambda_j = \operatorname{tr} \sum_{g \in G_i} D^{(i)}(g) = n_j \chi_j^{(i)}$$

where m_i is the dimension of $D^{(i)}$ and $\chi_j^{(i)} = \chi^{(i)}(g)$ for any $g \in G_j$. Thus

$$\lambda_j = \frac{n_j}{m_i} \chi_j^{(i)} = n_j \frac{\chi_j^{(i)}}{\chi_1^{(i)}}.$$

Using

$$\begin{aligned} T_j^{(i)} T_k^{(i)} \quad &= \textstyle\sum_{a \in G_j, b \in G_k} D^{(i)}(a) D^{(i)}(b) = \sum D^{(i)}(ab) = \\ &= D^{(i)}(G_j G_k) = \textstyle\sum_{s=1}^r h_{jk}^s D_s^{(i)}(G_s) = \sum_{s=1}^r h_{jk}^s T_s^{(i)} \end{aligned}$$

we get $\lambda_j \lambda_k = \sum_s h_{jk}^s \lambda_s$, that is,

$$n_j n_k \chi_j^{(i)} \chi_k^{(i)} = \sum_{s=1}^r h_{jk}^s \chi_s^{(i)} n_s \chi_1^{(i)}$$

and so

$$n_j n_k \sum_{i=1}^p \chi_j^{(i)} \chi_k^{(i)} = \sum_{s=1}^r h_{jk}^s \sum_{i=1}^p \chi_s^{(i)} \chi_1^{(i)} n_s$$

where $D^{(i)}$ are all nonequivalent irreducible representations of G, with $i = 1, 2, \ldots, p$. Using the formula

$$\chi = \sum_{i=1}^p m_i \chi^{(i)},$$

where m_i is the dimension of $D^{(i)}$, and the fact that $\chi(e) = |G|$ and $\chi(g) = 0$ for $g \notin G_1$ we have $\sum_i m_i \chi_j^{(i)} = \chi_j$, which is equal to $|G|$ for $j = 1$ and zero for $j \neq 1$. This implies

$$n_k n_l \sum_i \chi_k^{(i)} \chi_l^{(i)} = \sum_{s=1}^r \sum_{i=1}^p h_{kl}^s n_s m_i \chi_s^{(i)} = \sum_s h_{kl}^s n_s |G| \delta_{s,1} = h_{kl}^1 n_1 |G| = n_k |G| \delta_{kl'},$$

where $n_i = |G_i|$ and $\chi_j^{(i)} = \chi^{(i)}(g)$ for $g \in G_j$. Thus we have proven

Theorem 3.11 *The characters of the inequivalent irreducible representations of a finite group G satisfy the orthogonality relations*

$$\sum_i \chi_k^{(i)} \chi_l^{(i)} = \frac{[G|}{n_l} \cdot \delta_{kl'},$$

where the index j' corresponds to the conjugacy class G_j^{-1}.

Theorem 3.12 *The number of irreducible inequivalent representations of a finite group G is equal to the number of conjugacy classes in G.*

Proof. Define the vectors $\chi_i = (\chi_i^{(1)}, \dots, \chi_i^{(p)}) \in \mathbb{C}^p$. From the previous Theorem follows that the vectors $\chi_1, \dots, \chi_r$ are linearly independent, so $r \leq p$. On the other hand, we already know that the number r of inequivalent representations is less or equal to the number p of conjugacy classes, Theorem 3.8.

Theorem 3.13 *The matrix elements $\{D_{jk}^{(i)}(g)\}$ of inequivalent irreducible (unitary) representations of a finite group G form a complete set of functions on G. If $f: G \to \mathbb{C}$ is any function on G then we can write*

$$f(g) = \sum_{i,j;\alpha} a(i,j;\alpha) D_{ij}^{(\alpha)}(g)$$

where

$$a(i,j;\alpha) = \frac{dim\, D^{(\alpha)}}{|G|} \sum_g f(g) \overline{D^{(\alpha)}}_{ij}(g).$$

Proof. From Theorem 3.7 follows that the matrix elements of the different irreducible representation form an orthogonal system. On the other hand, on G there are exactly $|G|$ linearly independent functions; but we know that the number of matrix elements $= \sum m_i^2 = |G|$.

This theorem can be extended to much wider class of groups than the finite groups, namely the compact groups. The class of compact groups includes many important groups in the applications like the groups $U(n), SU(n), SO(n), Sp(2n,\mathbb{C}) \cap U(2n), \dots\dots$ We state the Theorem without proofs.

As a tool we need *the Haar measure* on a compact group G. We need a volume measure on G in order to define integration of functions, which is in turn necessary since for infinite groups we cannot use the usual summation over group elements, like in the previous theorems for finite groups; instead, we have to use integration over the group,

$$\sum_{g \in G} f(g) \mapsto \int_G f(g) d\mu_G.$$

There is a general theorem which guarantees the existence of a left (or right) invariant measure for all *locally compact groups.* A locally compact group is a topological group (the multiplication and taking the inverse are continuous operations) such the the neutral element has a compact neighborhood. Nonlocally compact groups turn up usually only in infinite-dimensional situations. All the matrix groups which we have met before are locally compact. A measure is left invariant if

$$\int_G f(g) d\mu_G = \int_G f(g_0 g) d\mu_G$$

for all $g_0 \in G$ and for all integrable functions f. One defines a right invariant measure in a similar way.

Example 1 In the case of the matrix group $GL(n, \mathbb{R})$ the integration is simply defined as

$$\int_G f(g) d\mu_G = \int \frac{f(g)}{|\det(g)|^n} \prod_{i,j} dg_{ij},$$

as the usual (Lebesgue, or in the case of a continuous function, Riemann) integration over the matrix entries $g_{ij} \in \mathbb{R}$.

Example 2 The group $U(1) = S^1$ is compact and the measure is the standard (normalized) measure $\frac{d\phi}{2\pi}$ on the circle.

Example 3 The group $SU(2)$ can be identified as the 3-dimensional unit sphere S^3 and the Haar measure is the (normalized) volume element, evaluated using the standard rotation invariant metric coming from $S^3 \subset \mathbb{R}^4$.

In the case of a compact group there is a measure which is at the same time left and right invariant, and it is uniquely defined up to positive scaling factor $d\mu_G \mapsto \lambda \cdot d\mu_G$. We can then normalize the measure such that the volume of G is equal to one.

Theorem 3.14 *The matrix elements of the different irreducible unitary representations of a compact group G satisfy*

$$\int_G \overline{D^{(\alpha)}}_{ij}(g) D^{(\beta)}_{kl}(g) d\mu_G = \frac{1}{dim\, D^{(\alpha)}} \delta_{\alpha,\beta} \delta_{i,k} \delta_{j,l}$$

where $d\mu_G$ is the (normalized) Haar measure on G. In addition, any square-integrable complex function f on G can be uniquely written as a linear combination of the matrix elements, where the coefficients in the expansion of f are given as

$$a(i,j;\alpha) = dim\, D^{(\alpha)} \cdot \int_G f(g) \overline{D^{(\alpha)}}_{ij}(g) d\mu_G.$$

As in the case of a finite group we can define the regular representation of a compact group as the representation in the Hilbert space $H = L^2(G, d\mu)$ of complex valued square-integrable functions given by $[D(g_0)f](g) = f(g_0^{-1}g)$. By the above theorem, the regular representation is a direct sum of the irreps $D^{(\alpha)}$, each appearing with the multiplicity = dim $D^{(\alpha)}$ in the decomposition.

Note that in the case of $G = U(1) = S^1$ the above theorem gives just the Fourier decomposition of a square-integrable function! All the representations are 1-dimensional and are given as $g \mapsto D(g) = g^n = e^{2\pi i n\phi}$ for $n \in \mathbb{Z}$. The generalization to other (Lie) groups leads to a branch of mathematics called *harmonic analysis* on Lie groups.

4 Representations of the symmetric group

4.1 Conjugacy classes in S_n

Recall the notation for elements in the symmetric group S_n :

$$f = \begin{pmatrix} 1 & 2 & \dots & n \\ f_1 & f_2 & \dots & f_n \end{pmatrix}$$

denotes the map $k \mapsto f_k$ in the set $\{1\,2\,3\dots n\}$. The p-cycles in S_n are denoted by $(i_1\, i_2 \dots i_p)$ are permutations where $i_1 \mapsto i_2$, $i_2 \mapsto i_3$ .... and $i_p \mapsto i_1$. We have shown that any permutation is a product of disjoint cycles. Furthermore, a cycle $(i_1 \dots i_p)$ is a product of transpositions,

$$(i_1 \dots i_p) = (i_1 i_2)(i_2 i_3) \dots (i_{p-1} i_p).$$

If $f \in S_n$ is an arbitrary permutation then

$$f(i_1 i_2 \cdots i_p) f^{-1} = (f_{i_1} f_{i_2} \cdots f_{i_p})$$

as can be seen directly from the definitions.

We denote by $[n_1, n_2, \dots, n_p]$ the set of elements in S_n which can be written as products of disjoint cycles of lengths $n_1, n_2, \dots, n_p$.

Theorem 4.1 *The conjugacy classes in S_n are subsets of the type $[n_1, n_2, \dots, n_p]$ where we can choose $n_1 \geq n_2 \geq \dots \geq n_p$ with $n_1 + n_2 + \dots + n_p = n$.*

Proof. We have seen above that the sets $[n_1, n_2, \dots, n_p]$ are conjugacy classes. On the other hand, given an element g in a conjugacy class, we can write it as a a product of cycles in some $[n_1, n_2, \dots, n_p]$. . Any other element in the same conjugacy class is then of the form fgf^{-1} for $f \in S_n$ and these are all included in the same $[n_1, n_2, \dots, n_p]$.

It follows now from the Theorem 3.12 that the number of nonequivalent irreducible representations of S_n is equal to the number of partitions $n = n_1 + n_2 + \dots + n_p$ to positive integers with $n_1 \geq n_2 \geq \dots \geq n_p$. In particular, for $n = 2$ we have two irreps and for $n = 3$ the number is 3.

4.2 A List of irreducible representations of S_n

We give the description of all nonequivalent irreducible representations of S_n without proof. For proofs, see for example the monograph D.E. Robinson: *Representation Theory of the Symmetric Group.*

Let $n_1 \geq n_2 \geq \dots \geq n_p$ be a partition of n to positive integers. We first form the *Young pattern* with p rows as given in the Figure 1, with row lengths n_i. The

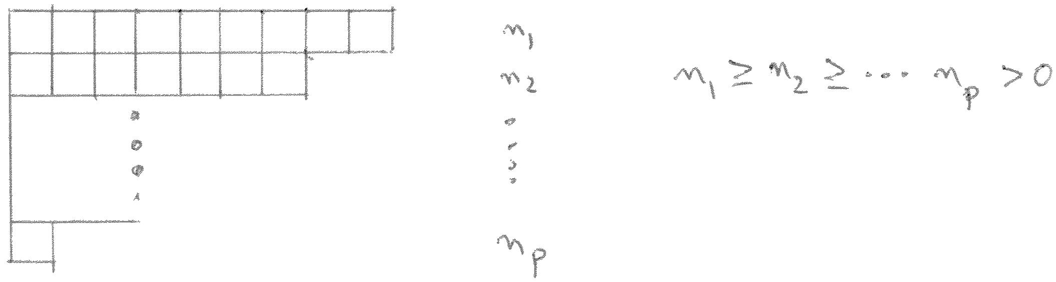
Young tableau is then formed by filling the rows with integers $12 \cdots n$. According to the following rules:

(1) In each row the numbers appear in increasing order from left to right

(2) In each column the numbers appear in increasing order from top to bottom

(3) Each number appears exactly once.

See the Figure (2).



Young pattern, Figure (1)

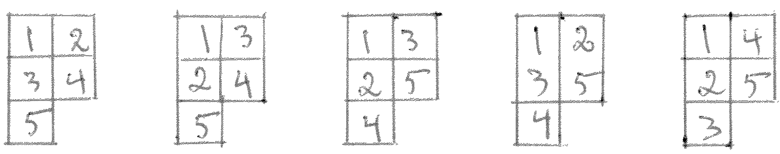


Figure (2): Different Young tableaux for $n = 5$, $n_1 = n_2 = 2$, $n_3 = 1$

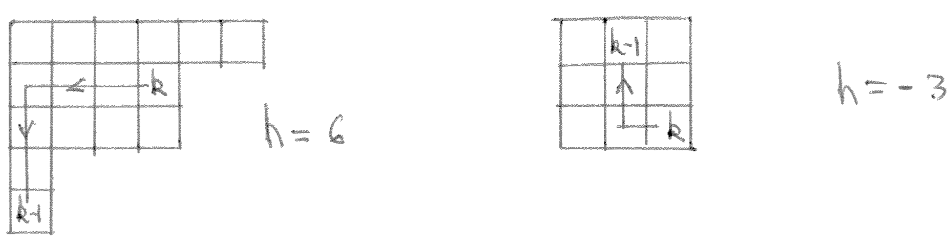


Figure (3): Examples of hooks

Theorem 4.2 *The number of different Young tableau corresponding to a partition* $[n_1, n_2, \ldots, n_p]$ *is equal to*

$$\frac{n!}{l_1! l_2! \cdots l_p!} \prod_{i<k} (l_i - l_k)$$

where $l_i = n_i + p - i$ *with* $i = 1, 2, \ldots, p$.

Label the rows by $1, 2, 3, \ldots$ from to to bottom. Denote by s_i the label of that row which contains the integer i. We denote this set by $|s_1, s_2, \ldots, s_m>$, *the Yamanouchi symbol* of the Young tableau. By the ordering principle of the integers in the rows of a Young tableau, the Yamanouchi symbol completely characterizes a Young tableau.

Next we define a complex vector space $V = V[n_1, n_2, \ldots, n_p]$ corresponding to a given Young pattern. The basis is labelled by the Young tableaux corresponding to the Young pattern. We can as well label the basis vectors in V by the Yamanouchi symbols $|s_1, s_2, \ldots, s_n>$.

A representation of S_n in V is now defined as follows. A transposition $(k-1, k)$ acts in V as a linear operator $D(k-1, k)$ such that

$$D(k-1,k)|s_1, s_2, \ldots, s_n> =$$
$$\begin{cases} +|s_1, s_2, \ldots, s_n> & \text{if } k-1, k \text{ appear in the same row} \\ -|s_1, s_2, \ldots, s_n> & \text{if } k-1, k \text{ are located in the same column} \\ \frac{1}{h}|s_1, s_2, \ldots, s_n> + \sqrt{1-h^{-2}}|s_1, \ldots, s_k, s_{k-1}, \ldots, s_n> & \text{otherwise ,} \end{cases}$$

where the *hook* h is the distance between the numbers $k, k-1$ in the given Young tableau; that is, it is $\pm$ the number of steps in the horizontal direction + the number of steps in the vertical direction needed to reach k from the location of $k-1$; we choose the positive sign if if $s_k < s_{k-1}$ and the negative sign for $s_k > s_{k-1}$. See the Figure 3.

Since every element in S_n can be written as a product of transpositions we get an action for every permutation as a linear operator in V.

Theorem 4.3 *The above construction defines an irreducible representation of* S_n *in* $V = V[n_1, n_2, \ldots, n_p]$ *for every partition of* n *as a sum of decreasing sequence of positive integers. Furthermore, these representations are nonequivalent and they form a complete set of nonequivalent irreducible representations of* S_n.

We can define an inner product in the vector space V be declaring that the basis vectors $|s_1, s_2, \ldots, s_n>$ corresponding to the different Yamanouchi symbols form an orthonormal basis. This representation is unitary (Exercise: Check this from the defining relations!)

Given any group G and a subgroup H, a representation D of G can be restricted to the subgroup giving naturally a representation of H. In a typical case, even when

D is irreducible, its restriction $D|_H$ is not irreducible. In the case of the symmetric group we have

Theorem 4.4 *When an irreducible representation of S_n corresponding to a Young pattern $[n_1, n_2, \dots, n_p]$ is restricted to the subgroup $S_{n-1} \subset S_n$ (those elements which leave n fixed) it is a direct sum of irreducible representations of S_{n-1} corresponding to Young tableaux $[m_1, m_2, \dots, m_q]$ such that there exists $1 \le i \le p$ with $m_j = n_j$ for $j \neq i$ and $m_i = n_i - 1$. Each of these representations occur with multiplicity $= 1$ in the decomposition.*

5 Some tensor analysis and representation theory

5.1 Tensor products of representations and the symmetric group

We shall explain some constructions of representations of classical Lie groups without proofs. By classical groups one means the matrix groups $GL(n, \mathbb{R}), GL(n, \mathbb{C})$ and their compact subgroups $SO(n), U(n), SU(n)$ and the symplectic group $Sp(2n)$ (both the real and the complex form). There is a close relation between representations and the corresponding Lie algebras (see the Exercises to the Section 2). The Lie algebra of a matrix group G consists of matrices X such that $e^{tX} \in G$ for all $t \in \mathbb{R}$. So given a representation D of a G in a vector space V we can define the matrices

$$d(X) = \frac{d}{dt} D(e^{tX})|_{t=0}$$

and then

$$[d(X), d(Y)] = d[[X, Y]).$$

This relation follows from the equation

$$D(e^{tX})D(e^{sY})D(e^{-tX})D(e^{-sY}) = D(e^{tX}e^{sY}e^{-tX}e^{-sY})$$

after a differentiation with respect to both t, s at $t = s = 0$. Starting from a representation $d(X)$ of a Lie algebra one could try to form a representation of the corresponding group using the exponential map $g = e^X \mapsto D(g) = e^{d(X)}$. However, there might be a problem with this construction since the exponential map is 1-1 only in some neighborhood of the point $X = 0$. We state without a proof:

Theorem 5.1 *Let G be a simply connected Lie group (i.e., a Lie group where any continuous loop can be continuously deformed to a point) and $\mathfrak{g}$ its Lie algebra. Then any finite-dimensional representation of $\mathfrak{g}$ can be exponentiated to a representation of the group G.*

Tensor analysis provides some very simple constructions of representations. It is somewhat harder to see that we get *all* irreducible representations this way. The reader is recommended to look at the classical text H. Weyl: *Classical Groups and their Invariants and Representations.*

A useful tool in the tensor analysis comes from physics: The use of the algebra of bosonic or fermionic creation and annihilation operators. We shall briefly discuss this method, through examples, in the end of the section. The linear groups $SU(n)$ and $SO(n)$ appear in physics often as symmetries of many particle systems. This could be for example a nucleus exhibiting various kinds of particle interchange and combined rotational symmetries. If the symmetry is exact, that is, the group commutes with the hamiltonian, then one can classify eigenvectors of the hamiltonian belonging to the same eigenvalue using the representation theory of the symmetry group G. Even in the case when the symmetry is only approximate it might still be of advantage to classify the physical states according to representations of G ('supermultiplets').

Let V, W be a pair of finite-dimensional vector spaces (over real or complex numbers). The tensor product $V \otimes W$ is then a vector space of dimension nm, where $n = \dim V$ and $m = \dim W$. All vector spaces of the same dimension are isomorphic, so the construction of $V \otimes W$ is not critical. It suffices to say that given a basis $\{v_1, \dots, v_n\}$ in V and a basis $\{w_1, \dots, w_m\}$ in W then a basis in $V \otimes W$ is given by the symbols $v_i \otimes w_j$. By linearity of the tensor product, if $v = \sum a_i v_i \in V$ and $w = \sum_j b_j w_j$ then the product $v \otimes w \in V \otimes W$ is defined as $v \otimes w = \sum_{i,j} a_i b_j v_i \otimes w_j$.

For those readers who are more familiar with linear algebra, the tensor product can be defined in a basis independent way as the vector space $\mathrm{Hom}(V^*, W)$ of linear maps from *the dual vector space* V^* to W. Given linear maps $A : V \to V$ and $B : W \to W$ one can define a linear map $A \otimes B : V \otimes W \to V \otimes W$ by

$$(A \otimes B)(v \otimes w) = A(v) \otimes B(w).$$

To see how the symmetry operates on many particle systems let us assume first that G is represented in a vector space V ('single particle space') with basis vectors $v_1 \dots v_n$. A 2-particle system is then described using the tensor product space $V \otimes V$ carrying the tensor product representation of G. Tensors can be split two antisymmetric and symmetric tensors. Writing a general element of $V \otimes V$ as $t = \sum t_{ij} v_i \otimes v_j$ we can split

$$t = a + s, \qquad a_{ij} = \frac{1}{2}(t_{ij} - t_{ji}), s_{ij} = \frac{1}{2}(t_{ij} + t_{ij}),$$

where s is symmetric and a is antisymmetric in the indices.

Writing a group element $g \in G$ as a matrix g_{ij} acting on the coordinates in the v_i basis we observe that in the tensor product representation the G action is $t'_{ij} = g_{ia} g_{jb} t_{ab}$ (sum over repeated indices) and therefore by linearity

$$a'_{ij} = g_{ia} g_{jb} a_{ab}, \qquad s'_{ij} = g_{ia} g_{jb} s_{ab},$$

i.e. the antisymmetric and symmetric parts transform separately. We have therefore two subrepresentations, one in the space of antisymmetric tensors and one in the space of symmetric tensors.

In general, the antisymmetric and symmetric parts can be further reduced to irreducible components. There are some exceptions, most notably the case when $G = SU(n)$ or $G = GL(n)$ acting in V through the defining representation. In these cases one can prove that the representations A and S are already irreducible.

One can go on and consider 3-, 4-,...n-particle systems. For example, in quantum mechanics a system of indistinguishable half-integer spin particles (fermions, e.g. electrons) obeys the Pauli exclusion principle: no two particles should be in the same state. Mathematically, this means that the system is described by elements in the completely antisymmetric tensor product space $\Lambda^k V$. Here k is the number of particles. The number of particles cannot exceed the number of one-particle levels n for combinatorial reasons; there are no completely antisymmetric tensors of rank $k > n$. For $k \leq n$ the number of independent antisymmetric tensors is

$$N(k, n) = \frac{n!}{k!(n-k)!}$$

This is the number of ways how one can select k *different* numbers from the sequence $1, 2, \ldots, n$. Each such selection defines a basis vector in $\Lambda^k V$ by

$$(i_1, \ldots, i_k) \mapsto \sum_\sigma \epsilon(\sigma) v_{i_1} \otimes \cdots \otimes v_{i_k}$$

where the sum is over all permutations of k letters and $\epsilon(\sigma) = \pm 1$ depending whether the permutation is a product of even or odd number of transpositions. It is clear that any antisymmetric tensor can be written uniquely as a linear combination of these elementary tensors.

In the case of integral spin particles (bosons) there is no Pauli exclusion principle; instead, the multiparticle wave function should be completely symmetric with respect to the interchange of arguments (Bose statistics). That is, the k particle states should be elements in the completely symmetrized tensor product $S^k V$. A complete basis in $S^k V$ is obtained by symmetrizing the vectors $v_{i_1} \otimes \cdots \otimes v_{i_k}$ with $i_1 \leq i_2 \leq \cdots \leq i_k$. Now $i_1 < i_2 + 1 < i_3 + 2 \cdots < i_k + k - 1$ are different positive integers in the set $1, 2, \ldots, n + k - 1$ and therefore the dimension

$$dim(S^k V) = \frac{(n + k - 1)!}{k!(n-1)!}.$$

In situations where not all of the particles are indistinguishable one has to deal with tensors of *mixed symmetry* type. For example, we could consider third rank tensors obtained from arbitrary tensors by an application of the mixed symmetry operator

$$R = (1 - (13))(1 + (12)),$$

where (ij) means the transposition of the i:th and of the j:th index; thus

$$(Rt)_{i_1 i_2 i_3} = t_{i_1 i_2 i_3} + t_{i_2 i_1 i_3} - t_{i_3 i_2 i_1} - t_{i_2 i_3 i_1}.$$

Note that the order of permutations is important. We denote tensors Rt symbolically by the Young diagram

$$\begin{array}{|c|c|}\hline i_1 & i_2 \\ \hline i_3 & \\ \cline{1-1}\end{array}$$

The completely symmetric tensors are denoted by $\boxed{i_1}\boxed{i_2}\ldots\boxed{i_k}$ and the completely antisymmetric ones by

$$\begin{array}{|c|}\hline i_1 \\ \hline i_2 \\ \hline \end{array}\\ \cdot \\ \cdot \\ \begin{array}{|c|}\hline i_k \\ \hline \end{array}$$

As another example of tensors of mixed symmetry type consider the Young diagram

$$\begin{array}{|c|c|}\hline i_1 & i_2 \\ \hline i_3 & i_4 \\ \hline \end{array}$$

The corresponding *Young symmetrizer* is $R = QP$ where

$$P = (1 + (12))(1 + (34)) \text{ and } Q = (1 - (13))(1 - (24)).$$

The general principle is the following: To each row in the Young diagram one associates a symmmetrizer in the corresponding tensor indices. Then one forms the product of all row symmetrizers; here the order is unimportant because the different rows do not mix. To each column one associates an antisymmerizer in the indices included in the column. Finally one multiplies by the product of antisymmetrizers from the left. So in the case of the above diagram one has

$$\begin{aligned}(Rt)_{i_1 i_2 i_3 i_4} &= t_{i_1 i_2 i_3 i_4} - t_{i_3 i_2 i_1 i_4} - t_{i_1 i_4 i_3 i_2} + t_{i_3 i_4 i_1 i_2} \\ &+ t_{i_2 i_1 i_3 i_4} - t_{i_2 i_3 i_1 i_4} - t_{i_4 i_1 i_3 i_2} + t_{i_4 i_3 i_1 i_2} \\ &+ t_{i_1 i_2 i_4 i_3} - t_{i_3 i_2 i_4 i_1} - t_{i_1 i_4 i_2 i_3} + t_{i_3 i_4 i_2 i_1} \\ &+ t_{i_2 i_1 i_4 i_3} - t_{i_2 i_3 i_4 i_1} - t_{i_4 i_1 i_2 i_3} + t_{i_4 i_3 i_2 i_1}\end{aligned}$$

All the permutation operators R commute with the linear group transformations $g \in G$. For this reason a tensor of the type Rt is transformed into a similar tensor Rt'. Thus the space RV^k of tensors of type R carries a representation of the group G. In fact, one can show that in the case of $G = SU(n)$ or $GL(n)$ in the defining representation this is irreducible. Not so in the case of $SO(n)$. The reason is simple: For the orthogonal group there are geometric invariants formed by the partial traces

$t_{jji_1i_2\ldots}$ of the tensors. For example, all the tensors for which this partial trace vanishes form an invariant subspace (the orthogonal transformations preserve the real euclidean inner product).

The operators R are *idempotents* modulo a normalization factor. This means that $R^2 = n_R \cdot R$ for some integer n_R. Exercise: Prove this in the case of the 3-box Young diagram above. The idempotent property means that (the normalized) symmetrization operators R act as projectors in the space of all tensors, projecting to the various irreducible representations of $SU(n)$ (or $GL(n)$).

Let $g \mapsto D^{(i)}(g)$ be representations of a matrix group G in vector spaces V_i, $i = 1, 2$. When $g = g(t) = e^{tX}$ for some $X \in \mathfrak{g}$ then in the tensor product we have

$$\frac{d}{dt}\Big|_{t=0}(D^{(1)} \otimes D^{(2)})(e^{tX})(v_1 \otimes v_2) =$$
$$\frac{d}{dt}\Big|_{t=0}D^{(1)}(g)v_1 \otimes D^{(2)}(g)v_2 = d^{(1)}(X)v_1 \otimes v_2 + v_1 \otimes d^{(2)}(X)v_2$$

so

$$d(X) = d^{(1)}(X) \otimes \mathbf{1} + \mathbf{1} \otimes d^{(2)}(X)$$

where d denotes the representation of $\mathfrak{g}$ in $V_1 \otimes V_2$ corresponding to the group representation $D = D^{(1)} \otimes D^{(2)}$.

Example $G = SU(3)$, defining representation in $V = \mathbb{C}^3$. Consider representations of $SU(3)$ also as representations of its Lie algebra $\mathfrak{su}(3)$, and also as its *complexification* A_2 (by taking complex linear combinations of the elements in the Lie algebra). The Young diagram $\begin{array}{|c|c|}\hline i_1 & i_2 \\ \hline i_3 \\ \cline{1-1}\end{array}$ gives the *adjoint representation.* The adjoint representation of any Lie algebra $\mathfrak{g}$ is defined as the natural representation in the vector space $\mathfrak{g}$, as $ad_x(y) = [x, y]$ with $x \in \mathfrak{g}$ and also $y \in \mathfrak{g}$, but with y considered as an element in the representation space. To see this consider the tensor $u = R(e_1 \otimes e_1 \otimes e_2)$, where e_i is the standard basis in $\mathbb{C}^3$. The eigenvalues of diagonal matrices for a tensor product Lie algebra representation add up, so u is an eigenvector of h_1 (here $h_i = e_{ii} - \frac{1}{3} \cdot \mathbf{1}$) with eigenvalue $\frac{2}{3} + \frac{2}{3} - \frac{1}{3} = 1$ and the eigenvalue for h_2 is $-\frac{1}{3} - \frac{1}{3} + \frac{2}{3} = 0$ giving the *highest weight* (more about weights later) $(1, 0)$ of the adjoint representation of $\mathfrak{su}(3)$. Furthermore, u is annihilated by e_{12} and e_{23}. For example,

$$e_{12}(e_1 \otimes e_1 \otimes e_2) = e_1 \otimes e_1 \otimes e_1$$

which is mapped to zero by R because of the antisymmetrization Q. Thus $e_{12}u = 0$. Similarly,

$$e_{23}(e_1 \otimes e_1 \otimes e_2) = 0$$

(since $e_{23}e_1 = 0 = e_{23}e_2$) and therefore also $e_{23}u = 0$. It follows that u is a highest weight vector. Finally, one checks that $R(e_1 \otimes e_1 \otimes e_2) \neq 0$. This agrees with the action of the Lie algebra $\mathfrak{su}(3)$, in the adjoint representation, on the vector $v = e_{13}$. This is in agreement of the classification of irreducible representations of Lie algebras like $\mathfrak{su}(n)$ in terms of highest weights and highest weight vectors to be discussed in more detail later.

5.2 Creation and annihilation operator formalism

In the case of completely symmetric wave functions (bosons) there is a simple formalism to describe the many particle states. To each bases vector v_i for one-particle states one associates a **creation operator** a_i^* with the commutation relations

$$[a_i^*, a_j^*] = 0.$$

A **vacuum** (zero particle state) is denoted by $|0>$. Multiparticle states are then obtained as polynomials

$$|k_1, k_2, \dots, k_n >= (a_1^*)^{k_1} \dots (a_n^*)^{k_n} |0>$$

acting on the vacuum; here the k_i's are arbitrary nonnegative integers. The bosonic structure of the indistinguishable particles is encoded in the commutation relations: the order of factors is unimportant and therefore the states $|k_1 \dots k_n>$ can be put to correspond vectors in the completely symmetric tensor product $S^k V$, where $k = k_1 + \dots + k_n$,

$$|k_1 \dots k_n > \mapsto S(v_1 \otimes \dots v_1 \otimes v_2 \otimes \dots v_2 \otimes \dots \otimes v_n \otimes \dots \otimes v_n)$$

where S is the complete symmetrization opeator (sum over all permutations of k factors), the number of v_1's is k_1, ..., the number of v_n's is k_n.

To describe the inner product in the Hilbert space of multiparticle states (called the *bosonic Fock space* $\mathcal{F}$) it is convenient to introduce also the **annihilation operators** a_i with the commutation relations

$$[a_i, a_j] = 0, \text{ but } [a_i, a_j^*] = \delta_{ij}.$$

The inner product is now fixed uniquely by the requirement that 1) the annihilation operator a_i is the adjoint of a_i^*, 2) the vacuum is annihilated by all annihilation operators, $a_i|0>= 0$, and 3) the normalization $<0|0>= 1$. For example,

$$\begin{aligned} <1,1|1,1> &= <0|(a_1^* a_2^*)^* (a_1^* a_2^*)|0> = <0|a_2 a_1 a_1^* a_2^*|0> \\ &= <0|a_2 [a_1, a_1^*] a_2^*|0> = <0|a_2 a_2^*|0> = <0|[a_2, a_2^*]|0> = <0|0> = 1 \end{aligned}$$

We define the operators

$$e_{ij} = a_i^* a_j.$$

It is easy to check the commutation relations

$$[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta_{il} e_{kj}.$$

We have thus constructed the Lie algebra of the general linear group $GL(n, \mathbb{C})$ acting in the bosonic Fock space. This representation is reducible. Define the *particle number operator*

$$N = \sum_i a_i^* a_i.$$

This commutes with all the operators e_{ij} and it follows that the different eigenspaces of N are invariant under the Lie algebra $\mathbf{gl}(n)$. This corresponds to the fact that the Fock space consists of completely symmetric tensors of arbitrary rank; the symmetric tensors of fixed rank form an irreducible representation space. Let $|m> = (a_1^*)^m |0>$. This vector is of rank m and is annihilated by all e_{ij} with $i < j$. It is also an eigenvector of all elements e_{ii} (the *Cartan subalgebra,* to be discussed later). The property $e_{ij}|m> = 0$ for $i < j$ means that $|m>$ is a *highest weight vector* corresponding to the weight $\lambda(e_{ii}) = m \cdot \delta_{1i}$.

As already noted before, the group $GL(n)$ acts irreducibly in the space of completely symmetric tensors; therefore a complete set of vectors in the subspace $\mathcal{F}_m = \{\psi \in \mathcal{F} | N\psi = m\psi\}$ is obtained by acting with the operators e_{ij} on the highest weight vector $\psi_m \in \mathcal{F}_m$. We can write

$$\mathcal{F} = \mathcal{F}_0 \oplus \mathcal{F}_1 \oplus \mathcal{F}_2 \ldots$$

and each $\mathcal{F}_m$ carries an irreducible representation of $GL(n)$.

In order to construct more general representations using the Fock space methods one has to increase the number of independent bosonic oscillator modes. We can prove that all finite-dimensional highest weight representations of $GL(n)$ or $SU(n)$ can be constructed using a set a_{ij}, a_{ij}^* of creation and annihilation operators with $1 \leq i, j \leq n$, commutation relations

$$[a_{ij}, a_{kl}^*] = \delta_{ik}\delta_{jl},$$

all other commutators being zero. The Lie algebra is constructed as

$$e_{ij} = \sum_k a_{ik}^* a_{jk}.$$

For each sequence $m = (m_1, m_2, \ldots, m_n)$ of nonnegative integers we construct the vector

$$\psi(m) = \prod_k (\det(a_{ij}^*)_{i,j \leq k})^{m_k} |0> .$$

Using the antisymmetry of a determinant as a function of the row vectors we first observe that $e_{ij}\psi(m) = 0$ for all $i < j$. The vector $\psi(m)$ is also an eigenvector of each e_{ii}; e_{ii} acts like a number operator for the oscillator modes with first index equal to i. The determinants are homogenenous functions of order 1 in each of the rows and columns and it follows that the action of e_{ii} on $\psi(m)$ is just a multiplication by the total degree $m_n + m_{n-1} + \cdots + m_i$. Thus we get for the components $\lambda_i = \lambda(e_{ii})$ of the highest weight, $\lambda_i = m_i + m_{i+1} \cdots + m_n$. In particular

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$$

and all the components are integers. Conversely, for each such a sequence λ there is a unique set of nonnegative integers m with the above relation to λ.

Remark In the next Section we shall discuss the representation theory of simple Lie algebras like A_ℓ, the complexification of $\mathfrak{su}(\ell+1)$. (Here $\ell = n-1$.)The representations are labelled by highest weights; here the weight is given by the set of eigenvalues λ_i of the diagonal elements in $\mathfrak{su}(\ell+1)$. There is a natural inner product in the set of weights and a bracket $<\lambda,\mu>= 2(\lambda,\mu)/(\mu,\mu)$. One can then show that $<\lambda,\alpha>= 2\frac{(\lambda,\alpha)}{(\alpha,\alpha)} = 0, 1, 2, \ldots$ for each so-called simple root $\alpha = \alpha_{i,i+1}$ of A_ℓ are essentially the conditions on the components λ_i derived above. All the finite-dimensional representations of A_{n-1} are generated by the different highest weight vectors $\psi(m)$ in the bosonic Fock space for n^2 independent oscillators. In the Young diagram notation, the representation λ corresponds to the diagram with row lengths $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$, read from top to bottom.

The completely antisymmetric representations (only one column in the Young diagram) are best constructed using the **fermionic oscillators** b_i^*, b_i, $i = 1, 2, \ldots, n$. The defining relations are described by *anticommutators* $[A,B]_+ = AB + BA$ instead of commutators,

$$[b_i^*, b_j]_+ = \delta_{ij}$$

and all other anticommutators are zero. The Lie algebra of $GL(n)$ is now constructed as

$$e_{ij} = b_i^* b_j.$$

The commutation relations can be checked using the identity

$$[AB, CD] = A[B,C]_+D - [A,C]_+BD + CA[B,D]_+ - C[A,D]_+B.$$

The fermionic Fock space consists of all creation operator polynomials acting on the vacuum $|0>$. As in the bosonic case the vacuum is defined by the relations $b_i|0>= 0$. The vacuum is again normalized, $<0|0>= 1$ and b_i^* is supposed to be the adjoint of b_i. These requirements fix the inner product uniquely.

The bosonic Fock space was infinite-dimensional. In the fermionic case the dimension is finite. The reason is that, because of the anticommutation relations, all the powers $(b_i^*)^k$ vanish identically for $k > 1$. The only nonzero vectors in the Fock space are of the type

$$b_{i_1}^* b_{i_2}^* \ldots b_{i_k}^* |0>,$$

where all the indices i_μ are distinct. By the anticommutation relations we can assume that $i_1 > i_2 \cdots > i_k$ (a change in the ordering corresponds just a multiplicative factor

± 1.) Thus the number of independent vectors of length k is $\binom{n}{k}$, which is equal to the number of independent components of a fully antisymmetric tensor of rank k in dimension n. We can again introduce a number operator $N = \sum_k b_k^* b_k$. The eigenvalue of N is now the rank of the antisymmetric tensor, or in other words, the number of boxes in the one-column Young diagram.

6 Semisimple Lie algebras

6.1 Lie algebras

Let $\mathbb{F}$ be the field of real or complex numbers. A *Lie algebra* is a vector space $\mathfrak{g}$ over $\mathbb{F}$ with a *Lie product* (or *commutator*) $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ such that

1. $x \mapsto [x, y]$ is linear for any $y \in \mathfrak{g}$,
2. $[x, y] = -[y, x]$,
3. $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$.

The last condition is called the *Jacobi identity* . From (1) and (2) it follows that also $y \mapsto [x, y]$ is linear for any $x \in \mathfrak{g}$. In this chapter we shall consider only finite-dimensional Lie algebras. In any vector space $\mathfrak{g}$ one can always define a *trivial Lie product* $[x, y] \equiv 0$. A Lie algebra with this commutator is *Abelian.* The space $\mathbf{gl}(n, \mathbb{R})$ of all real $n \times n$ matrices is naturally a Lie algebra with respect to the matrix commutator $[X, Y] = XY - YX$, and correspondingly the complex algebra $\mathbf{gl}(n, \mathbb{C})$.

Some other nontrivial examples follow:

Example 1 Let $\mathbf{o}(n)$ denote the space of all real antisymmetric $n \times n$ matrices. The commutator of a pair of matrices is defined by

$$[x, y] = xy - yx$$

(ordinary matrix multiplication in xy). Since $(xy)^t = y^t x^t$, where x^t denotes the transpose of the matrix x, the commutator of two antisymmetric matrices is again antisymmetric. The commutator clearly satisfies (1) and (2); (3) is checked by a simple computation. The dimension of the real vector space $\mathbf{o}(n)$ is $\frac{1}{2}n(n-1)$.

The matrix Lie algebras, like $\mathbf{o}(n)$ above, are closely related to groups of matrices. Let $O(n)$ denote the group of all orthogonal $n \times n$ matrices A, $A^t A = 1$. Then the Lie algebra $\mathbf{o}(n)$ consists precisely of those matrices x for which $A(s) = \exp sx \in O(n)$ for all $s \in \mathbb{R}$. Namely, taking the derivative of $A(s)^t A(s) = 1$ at $s = 0$ one gets $x^t + x = 0$. So $A(s) \in O(n)$ implies $x \in \mathbf{o}(n)$. On the other hand if $x \in \mathbf{o}(n)$ then $(\exp sx)^t = \exp sx^t = \exp(-sx) = (\exp sx)^{-1}$, so $A(s) \in O(n)$.

Example 2 The real vector space $\mathbf{u}(n)$ consisting of *anti- Hermitian* $n \times n$ matrices $x, x^* = -x$, where $x^* = \overline{x}^t$ and the bar means complex conjugation, is a Lie algebra with respect to the matrix commutator. Its dimension is n^2. Denoting by $U(n)$ the group of *unitary* matrices A, $A^*A = 1$, one can prove as in the case of orthogonal matrices that $\exp sx \in U(n) \, \forall s \in \mathbb{R}$ iff $x \in \mathbf{u}(n)$.

Example 3 The traceless anti-Hermitian $n \times n$ matrices form a Lie algebra to be denoted by $\mathbf{su}(n)$ and it corresponds to the group $SU(n) = \{A \in U(n) \mid \det A = 1\}$. The dimension of $\mathbf{su}(n)$ is $n^2 - 1$.

Example 4 Let J be the antisymmetric $2n \times 2n$ matrix

$$\begin{pmatrix} 0 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & -1 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Since $\det J = (-1)^{n+1} \neq 0$ the form $\langle x, y \rangle = x^t J y$ is nondegenerate (the vectors x, y are written as column matrices). Define $\mathbf{sp}(2n, \mathbb{R})$ to consist of all real $2n \times 2n$ matrices x such that $x^t J + Jx = 0$. This is a Lie algebra and one can associate to $\mathbf{sp}(2n, \mathbb{R})$ the group $Sp(2n, \mathbb{R})$ consisting of real matrices A such that $A^t J A = J$, or equivalently such that A preserves the form $\langle u, v \rangle = u^t J v$, $\langle Au, Av \rangle = \langle u, v \rangle$ for all $u, v \in \mathbb{R}^{2n}$. $Sp(2n, \mathbb{R})$ is the *symplectic group* defined by J.

One can analogously define the complex orthogonal Lie algebra $\mathbf{o}(n, \mathbb{C})$ and the complex symplectic Lie algebra $\mathbf{sp}(2n, \mathbb{C})$.

We have also the Lie algebra $\mathbf{sl}(n, \mathbb{C})$ of complex traceless $n \times n$ matrices and correspondingly the real Lie algebra $\mathbf{sl}(n, \mathbb{R})$.

Let $\{X_1, X_2, \dots, X_n\}$ be a vector space basis of a Lie algebra $\mathfrak{g}$. We define the *structure constants* c_{ij}^k by

$$[X_i, X_j] = c_{ij}^k X_k$$

(sum over the repeated index k; we shall use the same summation convention also later). From the defining properties (1) and (2) follows that the commutator $[X, Y]$ for arbitrary $X, Y \in \mathfrak{g}$ is determined by the structure constants. The Jacobi identity can be written as

$$c_{ij}^l c_{lk}^m + c_{jk}^l c_{li}^m + c_{ki}^l c_{lj}^m = 0$$

$\forall i, j, k, m$. By the antisymmetry of the Lie product we have $c_{ij}^k = -c_{ji}^k$.

Example Let $\mathfrak{g}$ be a two dimensional Lie algebra with a basis $\{X_1, X_2\}$. If $\mathfrak{g}$ is not commutative we can define a nonzero element

$$e_1 = [X_1, X_2] = \alpha X_1 + \beta X_2.$$

Choose a pair of numbers γ, δ such that $\alpha\delta - \beta\gamma = 1$ and set

$$e_2 = \gamma X_1 + \delta X_2.$$

Then $[e_1, e_2] = e_1$. Thus we have found the general structure of a noncommutative two dimensional Lie algebra.

Let $\mathfrak{g}$ and $\mathfrak{g}'$ be Lie algebras. A linear map $\phi : \mathfrak{g} \to \mathfrak{g}'$ is a *homomorphism* if

$$\phi([x, y]) = [\phi(x), \phi(y)]$$

$\forall x, y \in \mathfrak{g}$. An invertible homomorphism is an *isomorphism* . The inverse of an isomorphism is also an isomorphism. An isomorphism of $\mathfrak{g}$ into itself is an *automorphism* of the Lie algebra $\mathfrak{g}$.

A linear subspace $\mathbf{k} \subset \mathfrak{g}$ is a *subalgebra* of $\mathfrak{g}$ if $[x, y] \in \mathbf{k}\, \forall x, y \in \mathbf{k}$. A subalgebra is a Lie algebra in its own right.

A subspace $\mathbf{k} \subset \mathfrak{g}$ is an *ideal* if $[x, y] \in \mathbf{k}\, \forall x \in \mathfrak{g}$ and $y \in \mathbf{k}$. In particular, an ideal is always a subalgebra. If $\mathbf{k} \subset \mathfrak{g}$ is an ideal then the quotient space $\mathfrak{g}/\mathbf{k}$ is naturally a Lie algebra: The commutator of the cosets $x + \mathbf{k}$ and $y + \mathbf{k}$ is by definition the coset $[x, y] + \mathbf{k}$. If $x' + \mathbf{k} = x + \mathbf{k}$ and $y' + \mathbf{k} = y + \mathbf{k}$ (i.e., $x' - x \in \mathbf{k}$ and $y' - y \in \mathbf{k}$) then $[x', y'] = [x + (x' - x), y + (y' - y)] \equiv [x, y] \mod \mathbf{k}$ by the ideal property of $\mathbf{k}$; thus $[x', y']$ represents the same element in $\mathfrak{g}/\mathbf{k}$ as $[x, y]$ and so the commutator is well-defined in $\mathfrak{g}/\mathbf{k}$.

Theorem 6.1 *Let $\phi : \mathfrak{g} \to \mathfrak{g}'$ be a homomorphism which is onto* (i.e., $\mathfrak{g}' = \mathrm{im}\phi$). Then the Lie algebras $\mathfrak{g}'$ and $\mathfrak{g}/\ker\phi$ are isomorphic.

Proof. Define $\psi : \mathfrak{g}/\ker\phi \to \mathfrak{g}'$ by $\psi(x + \ker\phi) = \phi(x)$. Obviously ϕ is one-to-one and it is a homomorphism by $\psi([x + \ker\phi, y + \ker\phi]) = \psi([x, y] + \ker\phi) = \phi([x, y]) = [\psi(x + \ker\phi), \psi(y + \ker\phi)]$.

A linear map $\delta : \mathcal{A} \to \mathcal{A}$ in an algebra is a *derivation* if

$$\delta(a * b) = \delta(a) * b + a * \delta(b)$$

for all $a, b \in \mathcal{A}$.

Let $\mathrm{Der}(\mathcal{A})$ be the set of all derivations of $\mathcal{A}$. Then $\mathrm{Der}(\mathcal{A})$ is a Lie subalgebra of the Lie algebra of all endomorphisms of $\mathcal{A}$.

In the special case when $\mathcal{A} = \mathfrak{g}$ is a Lie algebra we can define a derivation ad_X of $\mathfrak{g}$ for any $X \in \mathfrak{g}$ by

$$\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}, \ \mathrm{ad}_X(Y) = [X, Y].$$

This defines a homomorphism ad: $\mathfrak{g} \to \mathrm{Der}(\mathfrak{g})$; this is called *the adjoint representation* of $\mathfrak{g}$. The derivations ad_X are called *inner derivations,* the rest are *outer derivations.*

6.2 Ideals in Lie algebras

A left (right) ideal in an algebra $\mathcal{A}$ is a linear subspace $I \subset \mathcal{A}$ such that $x * y \in I$ ($y * x \in I$) for all $x \in \mathcal{A}$ and $y \in I$. An (two sided) ideal is both left and right ideal.

If $\mathcal{A}$ is a Lie algebra, there is no difference between left and right ideals since $x * y = [x, y] = -[y, x]$.

The *center* of a Lie algebra $\mathfrak{g}$ is the subspace $Z(\mathfrak{g}) = \{x \in \mathfrak{g} | [x, y] = 0 \forall y \in \mathfrak{g}\}$. Clearly the center is an ideal. Another ideal is the subspace $[\mathfrak{g}, \mathfrak{g}]$ consisting of all linear combinations of commutators in the Lie algebra.

Lemma The vector space sum of two ideals in $\mathfrak{g}$ is again an ideal in $\mathfrak{g}$. The commutator $[I, J]$ of a pair of ideals is also an ideal.

Proof. The first claim follows directly from the definition. The second is a simple consequence of the Jacobi identity.

A Lie algebra $\mathfrak{g}$ is *simple* if its only ideals are the trivial ideals 0 and $\mathfrak{g}$ itself and if $\mathfrak{g}$ is not the commutative one dimensional Lie algebra. If $\mathfrak{g}$ is simple then $\mathfrak{g} = [\mathfrak{g}, \mathfrak{g}]$ and $Z(\mathfrak{g}) = 0$.

The basic example. Let $\mathfrak{g} = \mathbf{sl}(2, \mathbb{C})$. We choose a bases $h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Then

$$[h, x] = 2x, \ \ [h, y] = -2y, \ \ [x, y] = h.$$

Let $I \subset \mathfrak{g}$ be a nonzero ideal. We choose $0 \neq z = ax + by + ch \in I$. Then

$$[x, z] = bh - 2cx \text{ and } [x, bh - 2cx] = -2bx.$$

Thus $bx \in I$ and $[y, [y, z]] = -2ay \in I$.

1) If $a \neq 0$ then $y \in I$ and so $[x, y] = h \in I$ and $-\frac{1}{2}[x, h] = x \in I$ and so $I = \mathfrak{g}$. Likewise the case $b \neq 0$.

2) If $a = b = 0$ then $c \neq 0$ and $z = ch \in I$, so $h \in I$, $y = \frac{1}{2}[y, h] \in I$. and $x = -\frac{1}{2}[x, h] \in I$. It follows that $I = \mathfrak{g}$.

Thus $\mathbf{sl}(2, \mathbb{C})$ is simple. Actually, the above proof holds for $\mathbf{sl}(2, \mathbb{F})$ when $\mathbb{F}$ is an arbitrary field of characteristic not equal to 2.

Theorem 6.2 *1. Let $\phi : \mathfrak{g} \to \mathfrak{g}'$ be a Lie algebra homomorphism and $I \subset \mathfrak{g}$ an ideal such that $I \subset ker\phi$. Then there exists a unique homorphism $\psi : \mathfrak{g}/I \to \mathfrak{g}'$ such that $\phi = \psi \circ \pi$, where $\pi : \mathfrak{g} \to \mathfrak{g}/I$ is the canonical homomorphism.*

2. If $I, J \subset \mathfrak{g}$ is a pair of ideals with $I \subset J$ then J/I is an ideal in $\mathfrak{g}/I$ and $(\mathfrak{g}/I)/(J/I) \simeq \mathfrak{g}/J$.

3. If $I, J \subset \mathfrak{g}$ is any pair of ideals then $(I + J)/J \simeq I/(I \cap J)$.

Proof. (1) Define the map $\psi : \mathfrak{g}/I \to \mathfrak{g}'$ by $\psi(x+I) = \phi(x)$. It is easy to see that this is a homomorphism which satisfies the requirement. If ψ' is another such a homomorphism, then $(\psi' - \psi) \circ \pi = 0$ and so $\psi' - \psi = 0$ since π is onto.

(2) The first statement follows directly from definitions. For the second, define a map : $(\mathfrak{g}/I)/(J/I) \to \mathfrak{g}/I$ by $f((x+I)+J/I) = x+J$. This map is the required isomorphism.

(3) Define $f : I/(I \cap J) \to (I+J)/J$ by $f(x + I \cap J) = x + J$ and check that this is an isomorphism.

A *representation* of a Lie algebra $\mathfrak{g}$ in a vector space V is a Lie algebra homomorphism $\phi : \mathfrak{g} \to \mathrm{End}(V)$. As an example, any Lie algebra has the natural adjoint representation in the vector space $V = \mathfrak{g}$, $\mathrm{ad}_x(y) = [x,y]$.

A representation is *irreducible* if the representation space V does not have any invariant subspaces except of course 0 and V; a subspace $W \subset V$ is invariant if $\phi(x)v \in W$ for all $x \in \mathfrak{g}$ and $v \in W$.

If $\mathfrak{g}$ is a simple Lie algebra then the adjoint representation is necessarily irreducible. Conversely, if $\mathfrak{g}$ is noncommutative and the adjoint representation is irreducible then $\mathfrak{g}$ is simple.

If $\mathfrak{g}$ is simple then $Z(\mathfrak{g}) = 0$ and it follows that the kernel of the adjoint representation ad: $\mathfrak{g} \to \mathrm{End}(\mathfrak{g})$ is zero. Thus $\mathfrak{g}$ is isomorphic to a subalgebra of $\mathrm{End}(\mathfrak{g})$. Choosing a basis in $\mathfrak{g}$ we see that any simple Lie algebra is isomorphic to a Lie algebra of matrices.

Let $\delta \in \mathrm{Der}(\mathfrak{g})$, $\mathfrak{g}$ any finite-dimensional Lie algebra. Since δ is a linear operator in a finite-dimensional vector space we may form the exponential

$$e^{\delta} = 1 + \delta + \frac{1}{2!}\delta^2 + \frac{1}{3!}\delta^3 + \dots$$

to define a linear operator $\exp(\delta) : \mathfrak{g} \to \mathfrak{g}$.

Theorem 6.3 *The map* $\exp(\delta)$ *is an automorphism of* $\mathfrak{g}$.

Proof. First, $\exp(\delta)$ is a linear isomorphism since it has the inverse $\exp(-\delta)$. But

$$\begin{aligned}
\exp(\delta)[x,y] &= \sum_n \frac{1}{n!}\delta^n[x,y] \\
&= \sum_n \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} [\delta^k(x), \delta^{n-k}(y)] \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} [\frac{1}{k!}\delta^k(x), \frac{1}{i!}\delta^i(y)] = [e^{\delta}(x), e^{\delta}(y)]
\end{aligned}$$

and so $\exp(\delta)$ is a Lie algebra homomorphism. Here $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ are the binomial coefficients.

The automorphisms of the type $\exp(\delta)$ when $\delta = \mathrm{ad}_x$ are called *inner automorphisms.* They generate a group (upon multiplication), to be denoted by $\mathrm{Int}(\mathfrak{g})$; this is a subgroup of the group $\mathrm{Aut}(\mathfrak{g})$ of all automorphims of $\mathfrak{g}$.

Theorem 6.4 *The group Int*$(\mathfrak{g})$ *is a normal subgroup of Aut*$(\mathfrak{g})$*.*

Proof. Let $\phi \in \mathrm{Aut}(\mathfrak{g})$ and $x, y \in \mathfrak{g}$. Then

$$\phi \circ \mathrm{ad}_x \circ \phi^{-1}(y) = \phi([x, \phi^{-1}(y)]) = [\phi(x), y] = \mathrm{ad}_{\phi(x)}(y)$$

and thus $\phi \circ \mathrm{ad}_x \circ \phi^{-1} = \mathrm{ad}_{\phi(x)}$ which proves the statement.

6.3 The structure of semisimple Lie algebras

A Lie algebra L is semisimple if the canonical symmetric bilinear form, the **Killing form** , defined by

$$(X, Y) = \mathrm{tr}\ ad_X ad_Y,$$

is nondegenerate; we denote $ad_X : L \to L$ is the linear map $ad_X(Y) = [X, Y]$, the adjoint representation of L. Nondegenerate means that $(X, Y) = 0$ for all Y if and only if $X = 0$. In the case of the Lie algebra of $SU(2)$ or of $SO(3)$ a basis is given by vectors L_k such that $[L_1, L_2] = L_3$, and cyclic permutations of this relation (the angular momentum algebra in physics!) and it is easy to compute that

$$(L_j, L_k) = -2\delta_{jk}$$

which clearly shows that the Killing form is nondegenerate. From the definition it follows that the Killing form is invariant in the sense that

$$([X, Y], Z) = -(Y, [X, Z])$$

for all elements X, Y, Z. Now an exponential of an antisymmetric matrix is an orthogonal matrix (exercise!) and therefore the transformations $\exp(ad_X) : L \to L$ are orthogonal with respect to the Killing form, that is, the Killing form is preserved under these transformations.

One can show that a semisimple Lie algebra is a direct sum of simple Lie algebras. The angular momentum Lie algebra is simple. Other simple Lie algebras are the Lie algebras associated to the rotation groups $SO(n)$ when $n \neq 2, 4$, the Lie algebras of the special unitary groups $SU(n)$, and the Lie algebras of the symplectic groups $Sp(2n)$. In fact, this list is almost exhaustive. Any complex simple Lie algebra is either isomorphic to a complexification of one of the Lie algebras in this list or it is one of the so-called **exceptional** Lie algebras G_2, F_4, E_6, E_7, E_8. These latter Lie algebras are associated also to some groups, but they cannot be described so easily as the other *classical* Lie algebras.

We list now some basic properties of the semisimple Lie algebras. It is useful to think about some specific example, like the complexified Lie algebra A_ℓ of the group $SU(\ell+1)$. In a semisimple Lie algebra L there is a **Cartan subalgebra** which is a maximal abelian subalgebra $\mathfrak{h}$ such that the maps $ad_h : L \to L$ can be simultaneously diagonalized for all $h \in \mathfrak{h}$. In the case of A_ℓ the Cartan subalgebra can be chosen as the algebra of diagonal traceless matrices. A basis for $\mathfrak{h}$ is given by

$$h_n = e_{nn} - \frac{1}{\ell+1} \sum e_{ii},$$

where e_{ij} is the $(\ell+1)\times(\ell+1)$ matrix such that the matrix element at the position (ij) is 1 and all other matrix elements are equal to 0. The commutation relations for the e_{ij}'s are

$$[e_{ij}, e_{mn}] = \delta_{jm} e_{in} - \delta_{in} e_{mj}.$$

It follows that

$$ad_{h_n} e_{ij} = (\delta_{ni} - \delta_{nj}) e_{ij}.$$

From these relations one can then check that

$$(h_i, h_j) = 2(\ell+1)\delta_{ij} - 2.$$

In general, we call the eigenvalues of the Cartan subalgebra in the adjoint action **roots** for the pair $(L, \mathfrak{h})$. The Lie algebra L decomposes to eigenspaces which are called the root subspaces,

$$L = \mathfrak{h} \oplus L_\alpha \oplus L_\beta \oplus \dots$$

The labels $\alpha, \beta, \dots$ of the root subspace are some linear functions from $\mathfrak{h}$ to $\mathbb{C}$; the root subspace L_α is characterized by the property

$$[h, x] = \alpha(h) x \text{ for } x \in L_\alpha \text{ and } h \in \mathfrak{h} \ .$$

Thus in the above example $(A_\ell, \mathfrak{h})$ the one-dimensional root subspaces are spanned by the vectors e_{ij} with $i \neq j$. The zero root subspace is the Cartan subalgebra itself.

One can prove that the root subspaces (for nonzero roots) are all one-dimensional for any semisimple Lie algebra and the subspaces corresponding to different roots are orthogonal with respect to the Killing form.

In the set Φ of nonzero roots one can select a set Δ of **simple roots.** All other roots are either linear combinations of roots in Δ with positive integral coefficients (these are the positive roots Φ^+) or linear combinations with negative integral coefficients (these are the negative roots Φ^-). We have $\Phi = \Phi^+ \cup \Phi^-$ and $\Delta \subset \Phi^+$. In the case of A_ℓ we can take Δ as the set of roots corresponding to the root subspaces $e_{n,n+1}$ for $n = 1, 2, \dots, \ell$. Let us denote these roots by $\alpha_1, \dots, \alpha_\ell$. The positive roots are generated as follows. First, $e_{13} = [e_{12}, e_{23}]$ which implies, using the Jacobi identity,

$$[h, e_{13}] = [[h, e_{12}], e_{23}] + [e_{12}, [h, e_{23}]] = (\alpha_1(h) + \alpha_2(h)) e_{13}.$$

So $\alpha_{13} = \alpha_1 + \alpha_2$. By induction we obtain $\alpha_{ij} = \alpha_i + \cdots + \alpha_{j-1}$ for $i < j$. So the set of positive roots contains all α_{ij} for $i < j$. Similarly, the set of negative roots consists of the roots α_{ij} with $j < i$.

One can introduce a scalar product in the real vector space $\mathbb{R}^\ell$ spanned by the root vectors. The simple roots form a basis of this vector space. First one proves that the restriction of the Killing form to the Cartan subalgebra $\mathfrak{h}$ is nondegenerate. It follows that there is a natural linear isomorphism from the dual vector space $\mathfrak{h}^*$ to $\mathfrak{h}$, $\lambda \mapsto h_\lambda$,

$$(h, h_\lambda) = \lambda(h) \text{ for all } h \in \mathfrak{h}.$$

We can define an inner product in $\mathfrak{h}$ by setting $(\lambda, \mu) = (h_\lambda, h_\mu)$.

Example Consider the algebra A_ℓ (notation as before). For each root $\alpha_{ij} \in \mathfrak{h}^*$ we construct the vector $h_{ij} = h_{\alpha_{ij}}$. We can write

$$h_{ij} = \sum_{k=1}^{\ell+1} a_k h_k \, , \ a_k \in \mathbb{C}, \text{ with } \sum_k a_k = 0.$$

Using $(h_i, h_j) = 2(\ell+1)\delta_{ij} - 2$ we obtain

$$\begin{aligned} (h_{ij}, h_k) &= \alpha_{ij}(h_k) = \delta_{ik} - \delta_{jk} \\ &= (\Sigma a_n h_n, h_k) = \sum_{n=1}^{l} 2a_n[(\ell+1)\delta_{nk} - 1] \\ &= 2(\ell+1)a_k - 2\sum_{n=1}^{\ell+1} a_n = 2(\ell+1)a_k. \end{aligned}$$

We have a linear system of equations for the unknowns a_k. The solution is easily found to be $a_i = 1/2(\ell+1)$, $a_j = -1/2(\ell+1)$ and $a_k = 0$ for $k \neq i, j$. Thus

$$h_{ij} = \frac{1}{2(\ell+1)}(h_i - h_j).$$

From this we can compute the inner products

$$\begin{aligned} (\alpha_{ij}, \alpha_{mn}) &= \frac{1}{4(\ell+1)^2}(h_i - h_j, h_m - h_n) \\ &= \frac{1}{2(\ell+1)}(\delta_{im} + \delta_{jn} - \delta_{jm} - \delta_{in}). \end{aligned}$$

Usually it is sufficient to know the root space structure of a semisimple Lie algebra in terms of the inner products of the roots and an explicit knowledge of a matrix realization of the algebra in question is not needed.

The **rank** of a semisimple Lie algebra is the dimension of its Cartan subalgebra.

Theorem 6.5 *Let $\mathfrak{g}$ be a semisimple Lie algebra of rank l, $\mathfrak{h} \subset \mathfrak{g}$ a Cartan subalgebra, and Δ a system of simple roots for $(\mathfrak{g}, \mathfrak{h})$. Then Δ forms a basis of $\mathfrak{h}^*$. Let E denote the real vector space spanned by Δ. Then the dual $(\cdot, \cdot) : \mathfrak{h}^* \times \mathfrak{h}^* \to \mathbb{C}$ of the Killing form is a positive definite inner product in E.*

An important tool in the study of semisimple Lie algebras is *the Weyl group* of the root system. The Weyl group W of a root system Φ is generated by the reflections σ_α with $\alpha \in \Phi$. The reflection is defined as the linear map in $\mathbb{R}^\ell$

$$\sigma_\alpha(\beta) = \beta - < \beta, \alpha > \alpha$$

for $\beta \in \mathbb{R}^\ell$; we have inrtoduced the notation

$$< \alpha, \beta >= 2\frac{(\alpha, \beta)}{(\beta, \beta)}$$

for vectors such that $\beta \neq 0$. Note that $\sigma_\alpha(\alpha) = -\alpha$ and $\sigma_\alpha(\beta) = \beta$ if the vector β is orthogonal to α. Thus σ_α is indeed a reflection in the plane orthogonal to α. The basic property of the reflection is that it preserves inner products and thus also the brackets $< \cdot, \cdot >$. Also one can prove

Theorem 6.6 *The reflections σ_α and therefore also any element of the Weyl group maps the system Φ onto itself. Furthermore, if $\Delta, \Delta' \subset \Phi$ are systems of simple roots then there exists $\sigma \in W$ such that $\sigma(\Delta) = \Delta'$.*

Example In the case of A_ℓ a simple root α_i is orthogonal with respect to α_j except when $j = i$ or $j = i \pm 1$ and in that latter case $< \alpha_j, \alpha_i >= -1$ and so the *the fundamental reflections* $\sigma_i = \sigma_{\alpha_i}$ act on the basis vectors α_j as $\sigma_i(\alpha_j) = \alpha_j$ when $j \neq i, i \pm 1$, $\sigma_i(\alpha_i) = -\alpha_i$ and $\sigma_i(\alpha_{i\pm1}) = \alpha_{i\pm1} + \alpha_i$.

We can now check the following properties of the root system Φ of A_ℓ. Since the dimension of $\mathfrak{h}$ and thus of $\mathfrak{h}^*$ is ℓ, we can view Φ as a subset of vectors in $E = \mathbb{R}^\ell$.

(1) The system Φ spans E and $0 \notin \Phi$

(2) If $\alpha \in \Phi$ then $k\alpha \in \Phi$ if and only if $k = \pm 1$

(3) for any $\alpha \in \Phi$ also $\sigma_\alpha(\Phi) \subset \Phi$

(4) the numbers $< \alpha, \beta >$ are integers for $\alpha, \beta \in \Phi$

Actually, one can take these properties as *axioms* for root systems. Namely, one can prove that the root system of any semisimple Lie algebra satisfies the conditions (1) - (4) above.

Let $\Delta = \{\alpha_1, \ldots, \alpha_\ell\}$ be a system of simple roots. Denote

$$M_{ij} =< \alpha_i, \alpha_j >= 2 \cdot \frac{(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)}.$$

The numbers M_{ij} form a $\ell \times \ell$ integral matrix, called the *Cartan matrix* of the root system. In the 2-dimensional cases we have the matrices

$$A_1 \times A_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}; A_2 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}; B_2 = \begin{pmatrix} 2 & -2 \\ -1 & 2 \end{pmatrix}; G_2 = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}.$$

When Δ' is another basis then $\sigma(\Delta) = \Delta'$ for some $\sigma \in W$. The brackets $< \alpha, \beta >$ are invariant under the Weyl group. It follows that the Cartan matrix does not depend on the choice of Δ, modulo reordering of the basis.

Theorem 6.7 *Let (E, Φ) and (E', Φ') be a pair of root systems with $\Delta \subset \Phi$ and $\Delta' \subset \Phi'$ systems of simple roots. If the Cartan matrices M and M' are equal (with some choice of ordering of basis) then the root systems are isomorphic.*

Proof. Set $\Delta = \{\alpha_1, \dots, \alpha_\ell\}$ and $\Delta' = \{\alpha'_1, \dots, \alpha'_\ell\}$. We can define a linear isomorphism $\phi : E \to E'$ by $\phi(\alpha_i) = \alpha'_i$ since the simple roots form a basis. Then for any $\alpha, \beta \in \Delta$,

$$\begin{aligned} \sigma_{\phi(\alpha)}(\phi(\beta)) &= \phi(\beta) - < \phi(\beta), \phi(\alpha) > \phi(\alpha) \\ &= \phi(\beta) - < \beta, \alpha > \phi(\alpha) = \phi(\beta - < \beta, \alpha > \alpha) = \phi(\sigma_\alpha(\beta)). \end{aligned}$$

The second equality follows from the asssumption that the Cartan matrices are equal. Since Δ is a basis, we obtain $\sigma_{\phi(\alpha)} \circ \phi = \phi \circ \sigma_\alpha$, that is, $\phi \circ \sigma_\alpha \circ \phi^{-1} = \sigma_{\phi(\alpha)}$ for all $\alpha \in \Delta$. Since the simple reflections generate the Weyl group, we reduce that the map $\sigma \to \phi \circ \sigma \circ \phi^{-1}$ from W to W' is *an isomorphims of Weyl groups.*

Let next $\beta \in \Phi$ and choose $\sigma \in W$ such that $\sigma(\beta) \in \Delta$. Then

$$\phi(\beta) = (\phi \circ \sigma^{-1} \circ \phi^{-1})\phi(\sigma(\beta)) \in \Phi'$$

and so $\phi(\Phi) \subset \Phi'$. In the same way one shows that $\phi^{-1}(\Phi') \subset \Phi$ and thus $\phi(\Phi) = \Phi'$. If γ is another element of Φ then, by the linearity of $< \cdot, \cdot >$ in the first argument and by the equality of Cartan matrices,

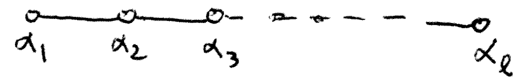
$$\begin{aligned} < \gamma, \beta > &= < \sigma(\gamma), \sigma(\beta) > = < \phi \circ \sigma(\gamma), \phi \circ \sigma(\beta) > \\ &= < (\phi \circ \sigma^{-1} \circ \phi^{-1})(\phi \circ \sigma(\gamma)), (\phi \circ \sigma^{-1} \circ \phi^{-1})(\phi \circ \sigma(\beta)) > = < \phi(\gamma), \phi(\beta) > . \end{aligned}$$

We have used the fact that the Weyl groups W, W' preserve the brackets. We have shown that ϕ is an isomorphism of the root systems.

If $\alpha \neq \beta$ is a pair of positive roots then $< \alpha, \beta > < \beta, \alpha >$ is one of the integers $0, 1, 2, 3$. We determine *the Coxeter graph* of the root system Φ from its Cartan matrix. The graph consists of ℓ nodes corresponding to the number of simple roots and lines connecting the nodes. The number of lines connecting the nodes α_i, α_j (for $i \neq j$) is equal to $< \alpha_i, \alpha_j > < \alpha_j, \alpha_i >$.

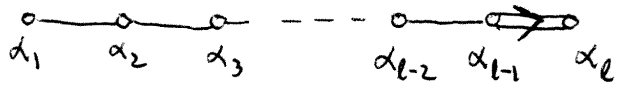
Appendix B: Dynkin diagrams

A_ℓ



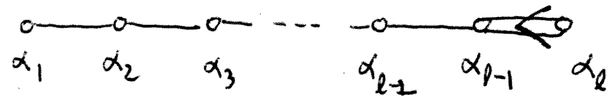
$\ell \geq 1$

B_ℓ



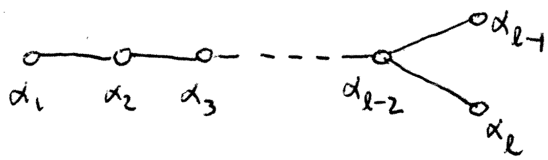
$\ell \geq 2$

C_ℓ



$\ell \geq 3$

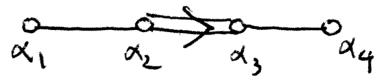
D_ℓ



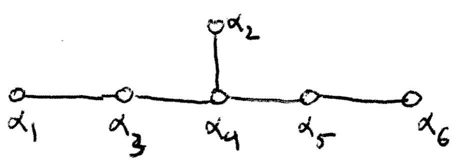
$\ell \geq 4$

G_2

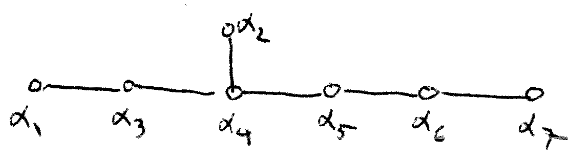
F_4



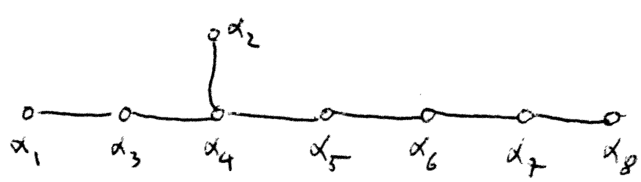
E_6



E_7



E_8



In the case when all simple roots have equal lengths the *Dynkin diagram* is equal to the Coxeter graph. In the case when a pair α_i, α_j of simple roots have unequal lengths we set an arrow to point towards the shorter root. On the enclosed sheet B we list all the Dynkin diagrams of simple Lie algebras.

The Dynkin diagram determines completely the Cartan matrix and therefore also the root system of a semisimple Lie algebra. In the case when the simple root lengths are equal, we have $<\alpha_i, \alpha_j> = -(<\alpha_i, \alpha_j><\alpha_j, \alpha_i>)^{1/2}$, for $i \neq j$. This gives all the matrix elements of the Cartan matrix. Suppose then that $(\alpha_i, \alpha_i) \neq (\alpha_j, \alpha_j)$ but we know that α_i is shorter, for example. Then from the table of of root lengths and angles (EXERCISE!) we see that $<\alpha_i, \alpha_j><\alpha_j, \alpha_i>$ is either 2 or 3. In the former case $<\alpha_i, \alpha_j> = -1$ and $<\alpha_j, \alpha_i> = -2$. In the latter case $<\alpha_i, \alpha_j> = -1$ and $<\alpha_j, \alpha_i> = -3$.

For example, from the Dynkin diagram of F_4 we can read its Cartan matrix

$$F_4 : \qquad \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -2 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

A root system Φ is *irreducible* when its Dynkin diagram is connected. Let $\Delta = \Delta_1 \cup \Delta_2 \cdots \cup \Delta_t$ be a decomposition of the simple roots corresponding to the connected components of the Dynkin diagram. Then $\Delta_i \perp \Delta_j$ for $i \neq j$ and let E_i be the subspace of E spanned by the roots Δ_i, $E = E_1 \oplus \cdots \oplus E_t$. Denote Φ_i the subset of roots which are linear combinations of the roots Δ_i.

Now the Weyl group W maps Φ_i onto itself: To see this it is sufficient to show that $\sigma_\alpha(\Phi_i) \subset \Phi_i$ for any simple root α. If $\alpha \notin \Delta_i$ then $\sigma_\alpha(\beta) = \beta - <\beta, \alpha> \alpha = \beta$ for any $\beta \in \Phi_i$. But if $\alpha \in \Delta_i$ then $\sigma_\alpha(\beta) = \beta - <\beta, \alpha> \alpha \in \Phi_i$ by the definition of Φ_i.

If $\beta \in \Phi$ is an arbitrary root we may choose $\sigma \in W$ such that $\sigma(\beta) \in \Delta$. But then $\sigma(\beta)$ belongs to some Δ_i and by the observation above $\beta \in \Phi_i$. Thus we have

$$\Phi = \Phi_1 \cup \Phi_2 \cdots \cup \Phi_t.$$

We have proven:

Theorem 6.8 *Any root system $\Phi \subset E$ is a union of irreducible root systems $\Phi_i \subset E_i$ with $E = E_1 \oplus \cdots \oplus E_t$, as an orthogonal direct sum.*

Now we list all irreducible root systems. We denote the standard basis vectors in $\mathbb{R}^\ell$ by $\mathbf{e}_1, \ldots, \mathbf{e}_\ell$.

Theorem 6.9 *Let E be the subspace of the euclidean space $\mathbb{R}^{\ell+1}$ with $\ell \geq 1$ consisting of vectors α such that $(\alpha, \sum \mathbf{e}_i) = 0$. Let L be the integral lattice in E and set $\Phi = \{\alpha \in L | (\alpha, \alpha) = 2\}$. Then (E, Φ) is an irreducible root system and its Dynkin diagram is the Dynkin diagram of the Lie algebra A_ℓ.*

Proof. Clearly

$$\Phi = \{\mathbf{e}_i - \mathbf{e}_j | i \neq j\}.$$

Let Δ consist of the vectors $\alpha_i = \mathbf{e}_i - \mathbf{e}_{i+1}$ with $i = 1, 2, \ldots, \ell$. These vectors form a basis of E. Furthermore, each element in Φ is an integral linear combination of vectors in Δ with only nonnegative or only nonpositive coefficients, so it satisfies the requirements of a system of simple roots; we also observe that clearly the first two axioms of a root system are satisfied. Next $<\beta,\alpha>= 2(\beta,\alpha)/(\alpha,\alpha) = (\beta,\alpha) \in \{0, \pm 1, 2\}$ so that also the fourth axiom holds.

Since $<\alpha_i,\alpha_{i+1}>= (\alpha_i,\alpha_{i+1}) = -1$ but $<\alpha_i,\alpha_j>= 0$ for $j \neq i\pm 1$ we see that the Dynkin diagram is really the diagram A_ℓ listed in the appendix B; one can then check by direct computation that the root system corresponding to the Cartan subalgebra of diagonal matrices in $\mathbf{sl}(\ell+1,\mathbb{F})$, with the choice of simple roots corresponding to the root vectors $e_{i,i+1} \in \mathbf{sl}(\ell+1,\mathbb{F})$, leads to the system (E,Φ,Δ).

Theorem 6.10 *Let $E = \mathbb{R}^\ell$ with $\ell \geq 2$ and Φ the set of vectors α in its integral lattice L such that $(\alpha,\alpha) = 1$ or $(\alpha,\alpha) = 2$. Then (E,Φ) is an irreducible system of roots with a Dynkin diagram corresponding to the Lie algebra B_ℓ of complex antisymmetric $(2\ell+1)\times(2\ell+1)$ matrices.*

Proof. Now $\Phi = \{\pm\mathbf{e}_i | 1 \leq i \leq \ell\} \cup \{\pm(\mathbf{e}_i \pm \mathbf{e}_j) | i \neq j\}$. The subset Δ of vectors $\alpha_i = \mathbf{e}_i - \mathbf{e}_{i+1}, i \leq \ell - 1$, and $\alpha_\ell = \mathbf{e}_\ell$ is linearly independent and the number of vectors is equal to the dimension of E, thus it is a basis of E. Furthermore,

$$\begin{aligned} \pm\mathbf{e}_i &= \pm(\alpha_i + \ldots \alpha_\ell) \\ \pm(\mathbf{e}_i - \mathbf{e}_j) &= \pm(\alpha_i + \cdots + \alpha_j) \text{ for } i < j \\ \pm(\mathbf{e}_i + \mathbf{e}_j) &= \pm(\alpha_i + \cdots + \alpha_{j-1} + 2\alpha_j + 2\alpha_{j+1} + \ldots 2\alpha_\ell) \text{ for } i < j. \end{aligned}$$

So Δ has the properties of a system of simple roots. When $i, j \leq \ell - 1$ the length of the roots α_i, α_j is equal to $\sqrt{2}$ and $<\alpha_i,\alpha_j>= 0$ for $i \neq j\pm 1, i \neq j$. For $j = i+1$ we have $<\alpha_i,\alpha_{i+1}><\alpha_{i+1},\alpha_i>= 1$. The length of α_ℓ is 1 and $<\alpha_{\ell-1},\alpha_\ell><\alpha_\ell,\alpha_{\ell-1}>= 2$. It follows that the Dynkin diagram is the diagram B_ℓ in the appendix. One can then check (a useful exercise!) that with a choice of a Cartan subalgebra this system indeed comes from the Lie algebra $\mathbf{so}(2\ell+1,\mathbb{C})$, the complexification of the Lie algebra of the group $SO(2\ell+1)$.

Theorem 6.11 *Let $E = \mathbb{R}^\ell$ with $\ell \geq 3$ and $\Phi = \{\pm 2\mathbf{e}_i | 1 \leq i \leq \ell\} \cup \{\pm(\mathbf{e}_i \pm \mathbf{e}_j) | i \neq j\}$. Then (E,Φ) is an irreducible root system corresponding to the Dynkin diagram C_ℓ.*

Remark We could have defined also C_2 but then $C_2 = B_2$. This root system corresponds to the Lie algebra of the complex symplectic group $Sp(2\ell)$. This group plays a central role in the formulation of classical Hamiltonian mechanics.

Theorem 6.12 *Let $E = \mathbb{R}^\ell$ for $\ell \geq 4$ and define Φ as the set of vectors α in the integral lattice with $(\alpha, \alpha) = 2$. Then $\Phi\{\pm(\mathbf{e}_i \pm \mathbf{e}_j) | i \neq j\}$ and it is an irreducible root system with Dynkin diagram D_ℓ corresponding to the Lie algebra of antisymmetric $2\ell \times 2\ell$ matrices, the complexified Lie algebra of the rotation group $SO(2\ell)$.*

Proof. This is actually a subalgebra of B_ℓ, by leaving out the short roots $\pm\mathbf{e}_i$. The simple roots are $\alpha_i = \mathbf{e}_i - \mathbf{e}_{i+1}$ for $i = 1, 2, \ldots, \ell - 1$ and $\alpha_\ell = \mathbf{e}_{\ell-1} + \mathbf{e}_\ell$.

We shall now describe the exceptional Lie algebras in terms of their root systems.

$\underline{G_2}$ Let $\{v_1, v_2, v_3\}$ be the standard basis of $\mathbb{R}^3$ and let E be the plane orthogonal to $v_1 + v_2 + v_3$. A basis of E is given by $\{v_1 - v_2, -2v_1 + v_2 + v_3\} = \Delta$. This is a system of simple roots for G_2. The positive roots are $\Phi^+ = \{v_1 - v_2, -v_1 + v_3, -v_2 + v_3, -2v_1 + v_2 + v_3, v_1 - 2v_2 + v_3, -v_1 - v_2 + 2v_3\}$.

$\underline{F_4}$ Let $E = \mathbb{R}^4$ and $\Delta = \{v_2 - v_3, v_3 - v_4, v_4, \frac{1}{2}(v_1 - v_2 - v_3 - v_4)\}$. The root system of F_4 consists of all *integral* linear combinations α of elements in Δ such that $\|\alpha\|^2 = 1$ or $\|\alpha\|^2 = 2$. One can show that $\Phi = \{\pm v_i\}_{i=1}^4 \cup \{\pm(v_i \pm v_j) \mid i \neq j\} \cup \{\pm\frac{1}{2}(v_1 \pm v_2 \pm v_3 \pm v_4) \mid \text{all signs}\}$. Thus the number of elements in Φ is 48.

Exercise What is the system of positive roots for F_4?

$\underline{E_8}$ Let $E = \mathbb{R}^8$ and $\Delta = \{\frac{1}{2}(v_1 + v_8) - \frac{1}{2}(v_2 + \ldots + v_7), v_1 + v_2, v_2 - v_1, v_3 - v_2, v_4 - v_3, v_5 - v_4, v_6 - v_5, v_7 - v_6\}$. The root system $\Phi(E_8)$ consists of all integral linear combinations α of elements in Δ such that $\|\alpha\|^2 = 2$. One can show that

$$\Phi = \{\pm(v_i \pm v_j) \mid i \neq j\} \cup \{\frac{1}{2}\sum_{i=1}^{8}(-1)^{\epsilon(i)} v_i \mid \epsilon(i) = 0, 1;\ \sum \epsilon(i) \in 2\mathbb{Z}\}.$$

There are 240 elements in Φ.

$\underline{E_7}$ Δ and Φ are defined here in a similar way as in the case of E_8 except that the last vector $v_7 - v_6$ in Δ is left out. There are 126 roots.

$\underline{E_6}$ Same as above, but now the two last vectors $v_6 - v_5$ and $v_7 - v_6$ are dropped. The number of roots is 72.

7 Representations of semisimple Lie algebras

A representation ϕ of a Lie algebra $\mathfrak{g}$ in a vector space V is *fully reducible* if V can be written as a direct sum $V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$ of invariant subspaces such that the restriction of ϕ to each V_i is an irreducible representation of $\mathfrak{g}$.

From known representations one can build new ones by taking direct sums. If ϕ_i is a representation of $\mathfrak{g}$ in V_i $(i \in \lambda)$, then a representation ϕ of $\mathfrak{g}$ in $\oplus \sum_{i\in\lambda} V_i$ is defined by

$$\phi(x)(v_i)_{i\in\lambda} = (\phi(x)v_i)_{i\in\lambda},$$

where $v_i \in V_i$. If ϕ is a representation of $\mathfrak{g}$ in a vector space V such that there is an invariant subspace W, then one can construct a representation ψ in the quotient space V/W by setting $\psi(x)(v+W) = \phi(x)v + W$, for any $v \in V$.

Theorem 7.1 *Any finite-dimensional representation of a semisimple Lie algebra is fully reducible.*

Thus for the purpose of classification of finite-dimensional representations of a semisimple Lie algebra it is sufficient to know the irreducible representations.

Let us denote by $\{x, y, h\}$ a basis of A_1, with the commutation relations

$$[x, y] = h, [h, x] = 2x, [h, y] = -2y.$$

A representation $\phi : \mathfrak{g} \to \mathrm{End}V$ of $\mathfrak{g} = A_1$ is a *highest weight representation* if there is a vector $0 \neq v \in V$ (the *highest weight vector*) such that

$$\begin{array}{lrcl} (1) & \phi(x)v & = & 0 \\ (2) & \phi(h)v & = & \lambda v \text{ for some } \lambda \in \mathbb{C} \\ (3) & V & = & \{\phi(u)v \mid u \in \mathcal{U}(\mathfrak{g})\}. \end{array}$$

Here $\mathcal{U}(\mathfrak{g})$, called *the universal enveloping algebra* of $\mathfrak{g}$, is just the associative algebra formed from all polynomials in x, y, h, subject to the commutation relations above. The content of (3) is that any vector in V can be reached from v by repeated action of $\phi(x), \phi(y)$, and $\phi(h)$.

The number λ is the *highest weight* of the representation.

Two representations $\phi : \mathfrak{g} \to \mathrm{End}\, V$, $\phi' : \mathfrak{g} \to \mathrm{End}\, V'$ of a Lie algebra $\mathfrak{g}$ are said to be *equivalent* if there is a linear isomorphism $\alpha : V \to V'$ such that

$$\alpha\phi(x)\alpha^{-1} = \phi'(x)\ \forall x \in \mathfrak{g}.$$

An irreducible highest weight representation (ψ, V) of A_1 is uniquely determined, up to an equivalence, by the highest weight λ.

1) Let $\lambda \notin \mathbb{N}$. By (1) - (3) the space V is spanned by the vectors $\{\psi(y)^n v \mid n = 0, 1, 2, \dots\}$. Using the commutation relations above we get

$$\psi(x)^n \psi(y)^n v = \alpha_n v\,,\ \alpha_n \neq 0.$$

Thus $\psi(y)^n v \neq 0$ for $n = 0, 1, 2, \dots$. The system $\{\psi(y)^n \mid n \in \mathbb{N}\}$ is linearly independent since different vectors correspond to different eigenvalues of the operator $\psi(h)$,

$$\psi(h)\psi(y)^n v = (\lambda - 2n)\psi(y)^n v.$$

A little more abstractly, we may think of V as a quotient space $\mathcal{U}(\mathfrak{g})/I_\lambda$, where I_λ is the *left-ideal* generated by the elements $h - \lambda(h)$ and x, that is, they consist of linear combibations $u_1(h-\lambda(h))+u_2x$, where u_1, u_2 are arbitrary polynomials of the generators x, y, h. A representation ϕ of $\mathfrak{g}$ in the quotient space is defined by $\phi(z)(u+I_\lambda)=zu+I_\lambda$ for $z=x,y,h$ and $u\in\mathcal{U}(\mathfrak{g})$. This is often (also for other Lie algebras) a powerful method to construct representations; one needs only specify a left-ideal which should kill a fixed vector in the representation. The method does not give any guarantee of the irreducibility.

We can define a linear isomorphism $\alpha:\mathcal{U}(\mathfrak{g})/I_\lambda\to V$ by $\alpha(v_n)=\psi(y)^n v$, where $v_n=y^n+I_\lambda$. We can check that $\alpha\phi(z)\alpha^{-1}=\psi(z)\ \forall z\in\mathfrak{g}$. For example,

$$\begin{aligned}\alpha\phi(y)\alpha^{-1}[\psi(y)^n v] &= \alpha\phi(y)v_n=\alpha v_{n+1}\\ &= \psi(y)^{n+1}v=\psi(y)[\psi(y)^n v].\end{aligned}$$

2) The case $\lambda\in\mathbb{N}$. Using the commutation relations we have

$$\begin{aligned}\psi(x)^n\psi(y)^n v &= \alpha_n v\neq 0 \qquad n=0,1,2,\ldots,\lambda\\ \psi(x)\psi(y)^{\lambda+1}v &=0\quad .\end{aligned}$$

It follows that $\{\psi(y)^n v\mid n=0,1,2,\ldots,\lambda\}$ is a basis of V. The rest of the proof goes like in the case 1.

Remark An irreducible finite-dimensional representation (ψ, V) of A_1 is always a highest weight representation. Let $0\neq w\in V$ be any eigenvector of $\psi(h)$ in V (which exists because of $\dim V<\infty$). If $\psi(x)w=0$ then w is a highest weight vector; otherwise set $v=\psi(x)^n w$, where n is the largest integer such that $\psi(x)^n w\neq 0$. Then w is a highest weight vector. [Because of the irreducibility of the representation, the invariant subspace $\{\psi(u)w\mid u\in\mathcal{U}(\mathfrak{g})\}$ must be the whole space V, and so also the condition (3) is satisfied.]

We shall now generalize the results obtained earlier for A_1, to the case of an arbitrary semisimple Lie algebra $\mathfrak{g}$ over $\mathbb{C}$. Note first that a semisimple Lie algebra is always spanned by subalgebras of the type A_1. Namely, let $\mathfrak{h}\subset\mathfrak{g}$ be a Cartan subalgebra and Φ the system of nonzero roots. If $\alpha\in\Phi$ then also $-\alpha\in\Phi$ (just look at the various root systems listed earlier). Choose $0\neq x_\alpha\in\mathfrak{g}_\alpha$ and $0\neq y_\alpha\in\mathfrak{g}_{-\alpha}$, remembering that $\dim\mathfrak{g}_\alpha=1\,\forall\alpha\in\Phi$. Set $k_\alpha=[x_\alpha,y_\alpha]$. If $h\in\mathfrak{h}$ then

$$\begin{aligned}[h,k_\alpha] &=[h,[x_\alpha,y_\alpha]]=[x_\alpha,[y_\alpha,h]]-[y_\alpha,[h,x_\alpha]]\\ &=-[x_\alpha,\alpha(h)y_\alpha]-[y_\alpha,\alpha(h)x_\alpha]=0.\end{aligned}$$

Since the Cartan subalgebra is a maximal commutative subalgebra of $\mathfrak{g}$, we have $k_\alpha\in\mathfrak{h}$. Since

$$[k_\alpha,x_\alpha]=\lambda x_\alpha \qquad [k_\alpha,y_\alpha]=-\lambda y_\alpha$$

with $\lambda=\alpha(k_\alpha)$, the subspace spanned by $\{y_\alpha,k_\alpha,x_\alpha\}$ is a subalgebra of $\mathfrak{g}$. We want to show that $\lambda\neq 0$.

Theorem 7.2 *(1) If $\alpha, \beta \in \Phi \cup \{0\}$ and $\alpha + \beta \neq 0$ then $\mathfrak{g}_\alpha \perp \mathfrak{g}_\beta$ with respect to the Killing form.*

(2) $[x_\alpha, y_\alpha] = (x_\alpha, y_\alpha) h_\alpha \, \forall \alpha \in \Phi$.

Proof. (1) Let $h \in \mathfrak{h}$ such that $(\alpha+\beta)(h) \neq 0$. Choose $0 \neq x_\alpha \in \mathfrak{g}_\alpha$ and $0 \neq x_\beta \in \mathfrak{g}_\beta$. Now

$$\begin{aligned} \alpha(h)(x_\alpha, x_\beta) &= ([h, x_\alpha], x_\beta) = (x_\alpha, [x_\beta, h]) \\ &= -\beta(h)(x_\alpha, x_\beta) \end{aligned}$$

and so $(\alpha + \beta)(h)(x_\alpha, x_\beta) = 0$ and $(x_\alpha, x_\beta) = 0$.

(2) Let $h \in \mathfrak{h}$. Then

$$(h, [x_\alpha, y_\alpha] - (x_\alpha, y_\alpha)h_\alpha) = (h, [x_\alpha, y_\alpha]) - (x_\alpha, y_\alpha)(h, h_\alpha)$$

$$= ([y_\alpha, h], x_\alpha) - (x_\alpha, y_\alpha)\alpha(h)$$

$$= \alpha(h)(y_\alpha, x_\alpha) - \alpha(h)(x_\alpha, y_\alpha) = 0.$$

Thus $\mathfrak{h} \perp [x_\alpha, y_\alpha] - (x_\alpha, y_\alpha)h_\alpha$. Since the restriction of the Killing form to $\mathfrak{h}$ is nondegenerate, the assertion follows.

Renormalizing the basis by $x = \sqrt{a} x_\alpha$, $y = \sqrt{a} y_\alpha$, $h = a k_\alpha$, where $a = 2/\lambda$ (and $\lambda = (x_\alpha, y_\alpha)\alpha(h_\alpha) = (x_\alpha, y_\alpha)(\alpha, \alpha) \neq 0$) we get the familiar commutation relations $[x, y] = h$, $[h, x] = 2x$, $[h, y] = -2y$ of A_1. We have now proven:

Theorem 7.3 *If $\alpha \in \Phi$ and $0 \neq x_\alpha \in \mathfrak{g}_\alpha$, $0 \neq y_\alpha \in \mathfrak{g}_{-\alpha}$ then $\{y_\alpha, h_\alpha, x_\alpha\}$ spans a subalgebra isomorphic to A_1.*

A representation $\phi :\to \mathrm{End} V$ is a *highest weight representation* if there is $0 \neq v \in V$ such that

1. $\phi(x_\alpha)v = 0 \, \forall \alpha \in \Phi^+$
2. $\phi(h)v = \lambda(h)v \, \forall h \in \mathfrak{h}$
3. $V = \{\phi(u)v \mid u \in \mathcal{U}(\mathfrak{g})\}$

where $\lambda : \mathfrak{h} \to \mathbb{C}$ is some linear form, the *highest weight* of the representation. From now on, when there is no danger of confusion, we shall write shortly zv instead of $\phi(z)v$, when $v \in V$ and $z \in \mathfrak{g}$. Consider a finite-dimensional highest weight representation of $\mathfrak{g}$ in V, with highest weight vector v. Then for each $\alpha \in \Phi^+$ there has to be $n_\alpha \in \mathbb{N}$ such that $y_\alpha^{n_\alpha+1} v = 0$; otherwise $\{y_\alpha^i v \mid i \in \mathbb{N}\}$ would span an infinite-dimensional subspace because these vectors are linearly independent by the eigenvector property

$$h_\alpha y_\alpha^i v = [\lambda(h_\alpha) - i\alpha(h_\alpha)] y_\alpha^i v.$$

On the other hand, by a similar calculation as was done for A_1,

$$x_\alpha y_\alpha^k v = k(x_\alpha, y_\alpha)\left[\lambda(h_\alpha) - \frac{1}{2}(k-1)\alpha(h_\alpha)\right] y_\alpha^{k-1} v.$$

It follows that in the finite-dimensional case (denoting by n_α the smallest number n for which $y_\alpha^{n+1}v = 0$) we must have $\lambda(h_\alpha) = \frac{1}{2}n_\alpha \cdot \alpha(h_\alpha)$. Using the notation $\langle\lambda,\alpha\rangle = 2\frac{(\lambda,\alpha)}{(\lambda,\alpha)}$ (where the inner product in the dual $\mathbf{h}^*$ is defined using the Killing form as $(\lambda,\lambda') = (h_\lambda, h_{\lambda'})$, see section 6) this relation can be written as

$$n_\alpha = \langle\lambda,\alpha\rangle.$$

In fact, one can prove a stronger result:

Theorem 7.4 *An irreducible highest weight representation of a semi-simple Lie algebra is finite-dimensional if and only if the highest weight satisfies $\langle\lambda,\alpha\rangle \in \mathbb{N}$ for all $\alpha \in \Delta$, where $\Delta \subset \Phi^+$ is a system of simple roots; these weights form the set Λ^+ of* **positive integral weights.** *For any linear form $\lambda : \mathfrak{h} \to \mathbb{C}$ which satisfies the above condition there is a unique, up to an equivalence, irreducible finite-dimensional representation with highest weight λ.*

One can define *the fundamental weights* for a system (Φ,Δ) as the weights λ^i with the property that $<\lambda^i, \alpha_j>=\delta_{ij}$ where both $i,j = 1,2,\ldots,\ell$. By defition, the fundamental weights are positive integral weights, they form a basis in Λ^+ so that any positive integral weight is a linear combination, with nonnegative integer coefficients, of the fundamental weights.

Example Let us consider the case of $\mathfrak{g} = A_2$. This is important in particle physics for at least two reasons. First, it is an approximate *internal symmetry* of strongly interacting particles ('eightfold way' of Gell-Mann and Ne'eman). Second, it is a *gauge symmetry* , in a similar way as in ordinary electrodynamics one can replace a vector potential A_μ by the equivalent potential $A_\mu + \partial_\mu\chi$, where the gauge transformation χ is a smooth function in space-time. In the case of A_2 (or $SU(3)$ in the group language) gauge symmetry the transformation is somewhat more complicated and instead of one potential there is a multiplet of eight potentials.

In this case the bases of the Cartan subalgebra is $h_1 = e_{11} - c, h_2 = e_{22} - c$, where $c = \frac{1}{3}(e_{11} + e_{22} + e_{33})$ commutes with everything. The simple roots are α_{12}, α_{23} corresponding to the root vectors e_{12}, e_{23}. The integrality conditions read

$$\lambda(h_{\alpha_{12}}), \lambda(h_{\alpha_{23}}) \in \frac{1}{6}\mathbb{N}.$$

The factor 1/6 comes from our normalization $(\alpha_{12},\alpha_{12}) = 1/3 = (\alpha_{23},\alpha_{23})$ (note that the normalization of the inner product is completely arbitrary, the only things which really matters are the ratios $<\alpha,\beta>= 2(\alpha,\beta)/(\beta,\beta)$) and from

$$\langle\lambda,\alpha\rangle = 2\frac{(\lambda,\alpha)}{(\alpha,\alpha)}.$$

Since $h_{\alpha_{12}} = \frac{1}{6}(h_1 - h_2)$ and $h_{\alpha_{23}} = \frac{1}{6}(h_2 - h_3) = \frac{1}{6}(h_1 + 2h_2)$ the condition for λ being dominant integral are that the highest weight vector is an eigenvector of h_1, h_2 with eigenvalues λ_1, λ_2 satisfying $\lambda_1 - \lambda_2 \in \mathbb{N}$ and $\lambda_1 + 2\lambda_2 \in \mathbb{N}$. The case $\lambda_1 = \lambda_2 = 0$ corresponds to the trivial one-dimensional representation. The lowest nontrivial case is $\lambda_1 = 2/3, \lambda_2 = -1/3$ corresponds to the defining representation of A_2 as complex 3×3 matrices. This weight is also the fundamental weight λ^1 for A_2. One can check from the definitions that the second fundamental weight λ^2 has the property $\lambda^2(h_1) = \frac{1}{3} = \lambda^2(h_2)$, and this corresponds to the complex conjugate of the defining representation. That representation is actually also the antisymmetrized tensor product of the defining representation with itself. Another important low dimensional example is $\lambda_1 = 1, \lambda_2 = 0$. This is the adjoint 8 dimensional representation. The highest weight vector is the root vector e_{13} and the weight is equal to $\lambda^1 + \lambda^2$.

In the case of A_1 a complete set of vectors in an irreducible representation was obtained by applying the shift operator $L_- = y$ step by step, starting from the highest weight vector. The relation of the angular momentum operators L_1, L_2, L_3 to the standard basis x, y, h in A_1 is $h = 2iL_3, x = L_1 + iL_2, y = -L_1 + iL_2$. In the case of A_2 one needs 3 independent shift operators (because there are 3 negative roots). These can be taken to be $S_{32} = e_{32}, S_{31} = e_{32}e_{21} + e_{31}(e_{11} - e_{22})$, and $S_{21} = e_{21}$. The first two connect vectors which are of highest weight with respect to a A_1 subalgebra. The last one decreases the A_1 weight inside of a representation of A_1.

The linear groups $SU(n)$ and $SO(n)$ appear in physics often as symmetries of many particle systems. This could be for example a nucleus exhibiting various kinds of particle interchange and combined rotational symmetries. If the symmetry is exact, that is, the group commutes with the hamiltonian, then one can classify eigenvectors of the hamiltonian belonging to the same eigenvalue using the representation theory of the symmetry group G. Even in the case when the symmetry is only approximate it might still be of advantage to classify the physical states according to representations of G ('supermultiplets').

8 Reduction of tensor products of representations

Let $D^{(i)}$ be a complete set of inequivalent representations of a compact group G. The in general the representation $D^{(i)} \otimes D^{(j)}$ is reducible and we can write

$$D^{(i)} \otimes D^{(j)} = \oplus_k c_{ijk} D^{(k)}$$

where the nonnegative integer c_{ijk} is the *multiplicity* of the representation $D^{(k)}$ in the tensor product.

Let us start by analyzing in detail the example $G = SU(2)$. We have seen that the irreducible representations are labelled by the highest weight $\lambda = 0, 1, 2, \ldots$, the highest eigenvalue of the element h in the Cartan subalgebra $\mathfrak{h}$. (In physics, $\lambda/2$

is interpreted as the angular momentum of a quantum mechanical system.) So in this case we can take the index i equal to λ, the representations are labelled by $0, 1, 2, \ldots$. Since the element h is represented as $h \otimes 1 + 1 \otimes h$ in the tensor product, the weights of $D^{(i)} \otimes D^{(j)}$ are the sums $m_i + m_j$ where $m_i = -i, -i+2, \ldots, +i$ and $m_j = -j, -j+2, \ldots, +j$. The weight spaces in any irreducible representation of $SU(2)$ have dimension equal to 1, so in the tensor product the highest weight is $i+j$ with multiplicity 1, the next highest is $i+j-2$ with multiplicity 2 (since $i+j-2 = (i-2)+j = i+(j-2)$,) the next is $i+j-4$ with multiplicity 3 and so on.

It follows that in the tensor product $D^{(i)} \otimes D^{(j)}$ the representation $D^{(i+j)}$ occurs with multiplicity 1. But since that representation contains a weight vector with weight $i+j-2$, we have only one remaining linearly independent vector with weight $i+j-2$. Thus the multiplicity of $D^{(i+j-2)}$ is also equal to 1! By induction, one can prove that this is the case up to the weight $|i-j|$ and after that we exhausted the possible highest weights. Thus

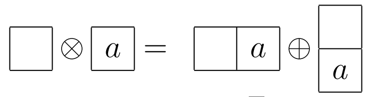
$$D^{(i)} \otimes D^{(j)} = \oplus \sum_{k=|i-j|}^{i+j} D^{(k)}.$$

With the aid of Young tableaux, we can give an algorithm for the Clebsch-Gordan decomposition of a tensor product of representations of $SU(n)$ for any n. The method is a bit complicated to explain, but stick with it. We'll do lots of examples later.

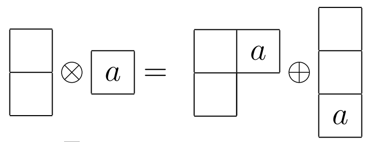
Suppose we want to decompose the tensor product of irreducible representations α and β corresponding to tableaux A and B. Put a's in the top row of B, b's in the second row, c's in the third row, and so on. Take the boxes from the top row of B and add them to A, each in different column, to form new tableaux. Then, take the second row and add them to form tableaux, again each box in a different column, then go to third row and so on until all rows in B have been exhausted. There is one additional restriction. Reading from right to left and from the top down, the number of a's must be greater than or equal to the number of b's, greater or equal to the number of c's ..... This avoids double counting of tensors. The tableaux formed in this way correspond to the irreducible representations in $\alpha \otimes \beta$.

We take some examples in the case of $SU(3)$. Here $\mathbf{3}$ denotes the defining 3-dimensional representation, $\overline{\mathbf{3}}$ the complex conjugate of the defining representation (which is the same as the representation on completely antisymmetric tensors of rank 2); the adjoint representation $\mathbf{8}$ corresponds to the Young tableau with row lengths (2,1).

Examples:

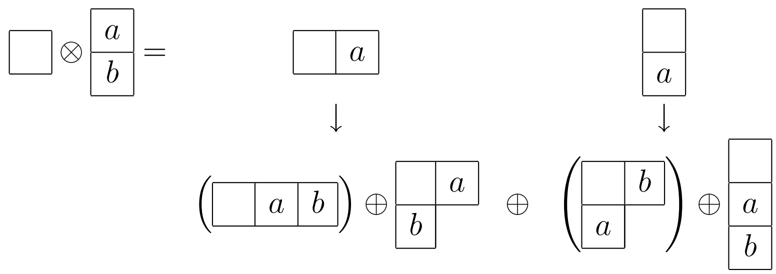


$$3 \otimes 3 = 6 \oplus \overline{3}$$



$$\overline{3} \otimes 3 = 8 \oplus 1$$

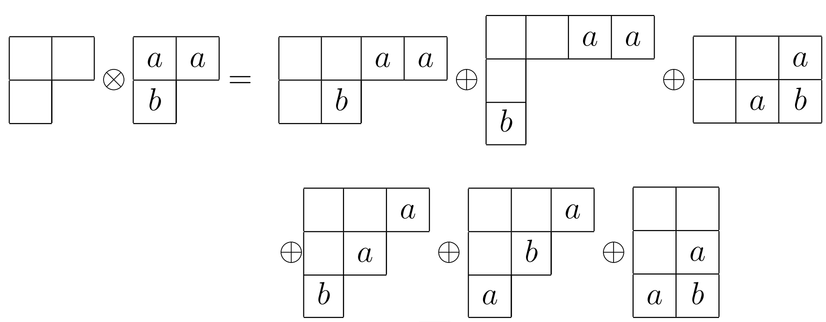
A less trivial example:



$$3 \otimes \overline{3} = \quad 8 \quad \oplus \quad 1$$

The first and third tableau do not satisfy the constraint that the number of a's is greater than the number of b's. Note that we did this one in a stupid way (on purpose to illustrate the constraint). A sensible person would work out $\overline{3} \otimes 3$ and not move so many boxes around.

Example: Finally, let us work out $8 \otimes 8$:



$$8 \otimes 8 = 27 \oplus 10 \oplus \overline{10} \oplus 8 \oplus 8 \oplus 1$$

The two 8's are different because they have a different pattern of a's and b's.

9 Differentiable Manifolds

9.1 Topological Spaces

The **topology** of a space X is defined via its open sets.

Let X= set, $\tau = \{X_\alpha\}_{\alpha \in I}$ a (finite or infinite) collection of subsets of X. (X, τ) is a **topological space**, if

T1 $\emptyset \in \tau,\ X \in \tau$

T2 <u>all</u> possible unions of X_α's belong to τ $\left(\bigcup_{\alpha \in I'} X_\alpha \in \tau,\ I' \subseteq I\right)$

T3 all intersections of a <u>finite</u> number of X_α's belong to τ. $(\bigcap_{i=1}^n X_{\alpha_i} \in \tau)$

The X_α are called the **open sets** of X in topology τ, and τ is said to give a **topology** to X.
So: topology $\hat{=}$ specify which subsets of X are open.
The same set X has several possible definitions of topologies (see examples).

Examples

(i) $\tau = \{\emptyset, X\}$ "trivial topology"

(ii) $\tau = \{\text{all subsets of } X\}$ "discrete topology"

(iii) Let $X = \mathbb{R}$, $\tau = \{\text{open intervals }]a, b[\text{ and their unions}\}$ "usual topology"

(iv) $X = \mathbb{R}^n$, $\tau = \{\ ]a_1, b_1[\times \ldots \times]a_n, b_n[$ and unions of these.$\}$

Definition: A **metric** on X is a function $d : X \times X \to \mathbb{R}$ such that

M1 $d(x, y) = d(y, x)$

M2 $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$.

M3 $d(x, y) + d(y, z) \geq d(x, z)$ "triangle inequality"

Example:

$$X = \mathbb{R}^n, \quad d_p(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p\right)^{\frac{1}{p}},\ p > 0$$

If $p = 2$ we call it the <u>Euclidean</u> metric.

If X has a metric, then the **metric topology** is defined by choosing all the "open disks"

$$U_\epsilon(x) = \{\ y \in X |\ d(x, y) < \epsilon\ \}$$

and all their unions as open sets.
The metric topology of $\mathbb{R}^n$ with metric d_p is equivalent with the usual topology (for all $p > 0$!)

Let (X, τ) be a topological space, $A \subset X$ a subset. The topology τ induces the **relative topology** τ' in A,

$$\tau' = \{ \ U_i \cap A \mid U_i \in \tau \ \}$$

This is how we obtain a topology for all subsets of $\mathbb{R}^n$ (like S^n).

9.1.1 Continuous Maps

Let (X, τ) and (Y, σ) be topological spaces. A map $f : X \to Y$ is **continuous** if and only if the inverse image of every open set $V \in \sigma$, $f^{-1}(V) = \{ \ x \in X \mid f(x) \in V \}$, is an open set in X: $f^{-1}(V) \in \tau$.
A function $f : X \to Y$ is a **homeomorphism** if f is continuous, and has an inverse $f^{-1} : Y \to X$ which is also continuous.

If there exists a homeomorphism $f : X \to Y$, then we say that X is **homeomorphic** to Y and vice versa. Denote $X \simeq Y$.
This $(\simeq)$ is an equivalence relation.

Intuitively : X and Y are homeomorphic if we can continuously deform X to Y (without cutting or pasting).
Example: coffee cup $\simeq$ donut.

[The fundamental question of topology : classify all homeomorphic spaces.]

One method of classification: **topological invariants** i.e. quantities which are invariant under homeomorphisms.
If a topological invariant for $X_1 \neq$ for X_2 then $X_1 \not\simeq X_2$.

The **neighbourhood** N of a point $x \in X$ is a subset $N \subset X$ such that there exists an open set $U \in \tau$, $x \in U$ and $U \subset N$.
(N does not have to be an open set).

(X, τ) is a **Hausdorff** space if for an arbitrary pair $x, x' \in X$, $x \neq x'$, there always exists neighbourhoods $N \ni x$, $N' \ni x'$ such that $N \cap N' = \emptyset$.
We'll assume from now on that all topological spaces (that we'll consider) are Hausdorff.

Example: $\mathbb{R}^n$ with the usual topology is Hausdorff.
All spaces X with metric topology are Hausdorff.

A subset $A \subset X$ is **closed** if its complement $X - A = \{x \in X \mid x \notin A\}$ is open.
N.B. X and $\emptyset$ are both open and closed.

A collection $\{A_i\}$ of subsets $A_i \subset X$ is called a **covering** of X if $\bigcup_i A_i = X$.
If all A_i are open sets in the topology τ of X, $\{A_i\}$ is an **open covering**.

A topological space (X, τ) is **compact** if, for every open covering $\{U_i \mid i \in I\}$ there exists a finite subset $J \subset I$ such that $\{U_i \mid i \in J\}$ is also a covering of X, i.e. every open covering has a finite subcovering.

Basic example According to the Heine-Borel theorem a subset $A \subset \mathbb{R}^n$ is compact (in the standard Euclidean metric topology) if and only if it is closed and bounded. Bounded means that $|x| < R$ for some $R > 0$ and for all $x \in A$.
X is **connected** if it cannot be written as $X = X_1 \bigcup X_2$, with X_1, X_2 both open, nonempty and disjoint, i.e. $X_1 \bigcap X_2 = \emptyset$.

A loop in topological space X is a continuous map $f : [0,1] \to X$ such that $f(0) = f(1)$. If any loop in X can be continuously shrunk to a point, X is called **simply connected**.

Examples: $\mathbb{R}^2$ is simply connected.
Torus T^2 is not simply connected.

Examples of topological invariants = quantities or properties invariant under homeomorphisms:

1. Connectedness
2. Simply connectedness
3. Compactness
4. Hausdorff
5. Euler characteristic (see below)

Let $X \subset \mathbb{R}^3$, $X \simeq$ polyhedron K. (monitahokas)
Euler characteristic:

$$\chi(X) = \chi(K) = (\#\text{ vertices in } K) - (\#\text{ edges in } K) + (\#\text{ faces in } K)$$
$$(= K\text{:n kärkien lkm.} - K\text{:n sivujen lkm.} + K\text{:n tahkojen lkm.})$$

Example: $\chi(T^2) = 16 - 32 + 16 = 0$.
$\chi(S^2) = \chi(\text{cube}) = 8 - 12 + 6 = 2$.

9.2 Homotopy Groups

9.2.1 Paths and Loops

Let X be a topological space, $I = [0,1] \subset \mathbb{R}$.
A continuous map $\alpha : I \to X$ is a **path** in X. The path α starts at $\alpha_0 = \alpha(0)$ and ends at $\alpha_1 = \alpha(1)$.
If $\alpha_0 = \alpha_1 \equiv x_0$, then α is a **loop** with **base point** x_0. We will focus on loops.

Definition: A **product** of two loops α, β with the same base point x_0, denoted by $\alpha \star \beta$, is the loop

$$(\alpha \star \beta)(t) = \begin{cases} \alpha(2t) & \text{for } 0 \leq t \leq \frac{1}{2} \\ \beta(2t-1) & \text{for } \frac{1}{2} \leq t \leq 1 \end{cases}$$

9.2.2 Homotopy

Let α, β be two loops in X with base point x_0. α and β are **homotopic**, $\alpha \sim \beta$, if there exists a continuous map $F : I \times I \to X$ such that

$$\begin{aligned} F(s,0) &= \alpha(s) && \forall s \in I \\ F(s,1) &= \beta(s) && \forall s \in I \\ F(0,t) &= F(1,t) = x_0 && \forall t \in I. \end{aligned}$$

F is called a **homotopy** between α and β.
Homotopy is an equivalence relation:

1. $\alpha \sim \alpha$: choose $F(s,t) = \alpha(s) \;\; \forall t \in I$

2. $\alpha \sim \beta$, homotopy $F(s,t) \Rightarrow \beta \sim \alpha$, homotopy $F(s, 1-t)$

3. $\alpha \sim \beta$, homotopy $F(s,t)$; $\beta \sim \gamma$, homotopy $G(s,t)$. Then choose

$$H(s,t) = \begin{cases} F(s,2t) & 0 \leq t \leq \frac{1}{2} \\ G(s,2t-1) & \frac{1}{2} \leq t \leq 1 \end{cases}$$

$\Rightarrow$ $H(s,t)$ is a homotopy between α and γ, so $\alpha \sim \gamma$.

The equivalence class $[\alpha]$ is called the **homotopy class** of α.
($[\alpha] = \{$ all paths homotopic with α $\}$).

Lemma: If $\alpha \sim \alpha'$ and $\beta \sim \beta'$, then $\alpha \star \beta \sim \alpha' \star \beta'$.
Proof: Let $F(s,t)$ be a homotopy between α and α' and let $G(s,t)$ be a homotopy between β and β'. Then

$$H(s,t) = \begin{cases} F(2s,t) & 0 \leq s \leq \frac{1}{2} \\ G(2s-1,t) & \frac{1}{2} \leq s \leq 1 \end{cases}$$

is a homotopy between $\alpha \star \beta$ and $\alpha' \star \beta'$. $\square$

By the lemma, we can define a product of homotopy classes: $\quad [\alpha] \star [\beta] \equiv [\alpha \star \beta]$.

Theorem: The set of homotopy classes of loops at $x_0 \in X$, with the product defined as above, is a group called the **fundamental group** (or **first homotopy group**) of X at x_0. It is denoted by $\Pi_1(X, x_0)$
Proof:

(0) Closure under multiplication: For all $[\alpha], [\beta] \in \Pi_1(X, x_0)$ we have $[\alpha] \star [\beta] = [\alpha \star \beta] \in \Pi_1(X, x_0)$, since $\alpha \star \beta$ is also a loop at x_0.

(1) Associativity: We need to show $(\alpha \star \beta) \star \gamma \sim \alpha \star (\beta \star \gamma)$.

$$\text{Homotopy } F(s,t) = \begin{cases} \alpha\left(\frac{4s}{1+t}\right) & 0 \leq s \leq \frac{1+t}{4} \\ \beta(4s - t - 1) & \frac{1+t}{4} \leq s \leq \frac{2+t}{4} \\ \gamma\left(\frac{4s-t-2}{2-t}\right) & \frac{2+t}{4} \leq s \leq 1 \end{cases}$$

$\Rightarrow \; [(\alpha \star \beta) \star \gamma] = [\alpha \star (\beta \star \gamma)] \equiv [\alpha \star \beta \star \gamma]$.

(2) Unit element: Let us show that the unit element is $e = [C_{x_0}]$, where C_{x_0} is the constant path $C_{x_0}(s) = x_0 \;\; \forall s \in I$. This follows since we have the homotopies:

$$\alpha \star C_{x_0} \sim \alpha : \qquad F(s,t) = \begin{cases} \alpha\left(\frac{2s}{1+t}\right) & 0 \leq s \leq \frac{1+t}{2} \\ x_0 & \frac{1+t}{2} \leq s \leq 1 \end{cases}$$

$$C_{x_0} \star \alpha \sim \alpha : \qquad F(s,t) = \begin{cases} x_0 & 0 \leq s \leq \frac{1-t}{2} \\ \alpha\left(\frac{2s-1+t}{1+t}\right) & \frac{1-t}{2} \leq s \leq 1 \end{cases}.$$

$\Rightarrow \; [\alpha \star C_{x_0}] = [C_{x_0} \star \alpha] = [\alpha]$.

(3) Inverse: Define $\alpha^{-1}(s) = \alpha(1-s)$. We need to show that α^{-1} is really the inverse of α: $[\alpha \star \alpha^{-1}] = [C_{x_0}]$. Define:

$$F(s,t) = \begin{cases} \alpha(2s(1-t)) & 0 \leq s \leq \frac{1}{2} \\ \alpha(2(1-s)(1-t)) & \frac{1}{2} \leq s \leq 1 \end{cases}$$

Now we have $F(s, 0) = \alpha \star \alpha^{-1}$ and $F(s, 1) = C_{x_0}$ so $\alpha \star \alpha^{-1} \sim C_{x_0}$. Similarly $\alpha^{-1} \star \alpha \sim C_{x_0}$ so we have proven the claim: $[\alpha^{-1} \star \alpha] = [\alpha \star \alpha^{-1}] = [C_{x_0}]$. $\square$

In general, a homotopy between two maps $f, g : X \to Y$ is defined as a map $F : X \times [0, 1]$ such that $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$. The topological spaces X, Y are said to be homotopy equivalent if there is a pair of maps $f : X \to Y$ and $g : Y \to X$ such that $f \circ g$ is homotopic to id_Y and $g \circ f$ id homotopic to id_X. Many topological notions are invariant under a homotopy equivalence; for example, the fundamental groups of X and Y are isomorphic if X, Y are homotopy equivalent.

9.2.3 Properties of the Fundamental Group

1. If x_0 and x_1 can be connected by a path, then $\Pi_1(X, x_0) \cong \Pi_1(X, x_1)$. If X is arcwise connected, then the fundamental group is independent of the choice of x_0 up to an isomorphism: $\Pi_1(X, x_0) \cong \Pi_1(X)$.

   (A space X is arcwise connected if any two points $x_0, x_1 \in X$ can be connected with a path. It can be shown that an arcwise connected space is always connected, but the converse is not true. However a connected metric space is also arcwise connected.)

2. $\Pi_1(X)$ is a topological invariant: $X \simeq Y \Rightarrow \Pi_1(X) \cong \Pi_1(Y)$.

3. Examples:
   - $\Pi_1(\mathbb{R}^2) = 0$ (= the trivial group)
   - $\Pi_1(T^2) = \Pi_1(S^1 \times S^1) = \mathbb{Z} \oplus \mathbb{Z}$.

   (One can show that $\Pi_1(X \times Y) = \Pi_1(X) \oplus \Pi_1(Y)$ for arcwise connected spaces X and Y.)

The real projective space is defined as $\mathbb{R}P^n = \{$ lines through the origin in $\mathbb{R}^{n+1}\}$. If $x = (x^0, x^1, \ldots, x^n) \neq 0$, then x defines a line. All $y = \lambda x$ for some nonzero $\lambda \in \mathbb{R}$ are on the same line and thus we have an equivalence relation: $y \sim x \leftrightarrow y = \lambda x,\ \lambda \in \mathbb{R} - \{0\} \leftrightarrow$ (x and y are on the same line.)
So $\mathbb{R}P^n = \{[x] |\ x \in \mathbb{R}^{n+1} - 0\}$ with the above equivalence relation.

Example: $\mathbb{R}P^2 \simeq$ (S^2 with opposite points identified)
$\Pi_1(\mathbb{R}P^2) = \mathbb{Z}_2$.

9.2.4 Higher Homotopy Groups

Define: $I^n = \{(s_1, \ldots, s_n) |\ 0 \leq s_i \leq 1,\ 1 \leq i \leq n\}$
$\partial I^n =$ boundary of $I^n = \{(s_1, \ldots, s_n) |$ some $s_i = 0$ or $1\}$

A map $\alpha : I^n \to X$ which maps every point on ∂I^n to the same point $x_0 \in X$ is called an **n-loop** at $x_0 \in X$. Let α and β be n-loops at x_0. We say that α is homeotopic to β, $\alpha \sim \beta$, if there exists a continuous map $F : I^n \times I \to X$ such that

$$F(s_1, \ldots, s_n, 0) = \alpha(s_1, \ldots, s_n)$$
$$F(s_1, \ldots, s_n, 1) = \beta(s_1, \ldots, s_n)$$
$$F(s_1, \ldots, s_n, t) = x_0 \quad \forall t \in I \text{ when } (s_1, \ldots, s_n) \in \partial I^n.$$

Homotopy $\alpha \sim \beta$ is again an equivalence relation with respect to homotopy classes $[\alpha]$.

$$\text{Define: } \alpha \star \beta : \ \alpha \star \beta(s_1, \ldots, s_n) = \begin{cases} \alpha(2s_1, s_2, \ldots, s_n) & 0 \le s_1 \le \frac{1}{2} \\ \beta(2s_1 - 1, s_2, \ldots, s_n) & \frac{1}{2} \le s_1 \le 1. \end{cases}$$

$\alpha^{-1} : \ \alpha^{-1}(s_1, \ldots, s_n) = \alpha(1 - s_1, \ldots, s_n)$
$[\alpha] \star [\beta] = [\alpha \star \beta]$
$\Rightarrow \Pi_n(X, x_0)$, the $\mathbf{n^{th}}$ **homotopy group** of X at x_0. (This classifies continuous maps $S^n \to X$.)
Example: $\Pi_n(S^n) = \mathbb{Z}$ for all $n \ge 1$ whereas $\Pi_k(S^n) = 0$ for $k < n$.

One of the classic problems in algebraic topology is to compute the homotopy groups $\pi_k(S^n)$ of spheres. This is still unsolved for general values of n, k; partial answers are known when k is not too large compared to n. For example, $\Pi_4(S^3) = \Pi_5(S^3) = \mathbb{Z}_2$.

For more information on homotopy groups and algebraic topology in general, see M. Greenberg and J.R. Harper: *Algebraic Topology, A First Course.*

One can prove (and you might be able to provide the proof) that the higher homotopy groups $\Pi_n(X)$, for $n \ge 2$, **are all commutative.**

9.3 Differentiable manifolds

9.3.1 The definition of a differentiable manifold

Let M be a topological space. We normally assume also the *Hausdorff property:* For any pair x, y of distinct points there is a pair of nonoverlapping open sets U, V such that $x \in U$ and $y \in V$.

Actually, all the spaces we study in (finite dimensional) differential geometry are *locally homeomorphic* to $\mathbb{R}^n$.

Definition A topological space M is called a *smooth manifold* of dimension n if

- there is a family of open sets U_α (with $\alpha \in \Lambda$) such that the union of all U_α's is equal to M
- for each α there is a homeomorphism $\phi_\alpha : U_\alpha \to V_\alpha \subset \mathbb{R}^n$ such that
- the *coordinate transformations* $\phi_\alpha \circ \phi_\beta^{-1}$ on their domains of definition are smooth functions in $\mathbb{R}^n$.

Example 1 $\mathbb{R}^n$ is a smooth manifold. We need only one *coordinate chart* $U = M$ with $\phi : U \to \mathbb{R}^n$ the identity mapping.

Example 2 The same as above, but take $M \subset \mathbb{R}^n$ any open set.

Example 3 Take $M = S^1$, the unit circle. Set U equal to the subset parametrized by the polar angle $-0.1 < \phi < \pi + 0.1$ and V equal to the set $\pi < \phi < 2\pi$. Then $U \cap V$

consists of two intervals $\pi < \phi < \pi + 0.1$ and $-0.1 < \phi < 0 \sim 2\pi - 0.1 < \phi < 2\pi$. The coordinate transformation is the identity map $\phi \mapsto \phi$ on the former and the translation $\phi \mapsto \phi + 2\pi$ on the latter interval.

Example 4 The unit sphere S^n in n dimensions. Let's realize S^n as a subset of $\mathbb{R}^{n+1}$: $S^n = \{x \in \mathbb{R}^{n+1} | \sum_{i=0}^{n} (x^i)^2 = 1\}$.
One possible atlas:

- coordinate neighbourhoods:

$$U_{i+} \equiv \{x \in S^n | x^i > 0\}$$
$$U_{i-} \equiv \{x \in S^n | x^i < 0\}$$

- coordinates:

$$\varphi_{i+}(x^0, \ldots, x^n) = (x^0, \ldots, x^{i-1}, x^{i+1}, \ldots, x^n) \in \mathbb{R}^n$$
$$\varphi_{i-}(x^0, \ldots, x^n) = (x^0, \ldots, x^{i-1}, x^{i+1}, \ldots, x^n) \in \mathbb{R}^n$$

(so these are projections on the plane $x^i = 0$.)

The transition functions $(i \neq j,\ \alpha = \pm,\ \beta = \pm)$,

$$\begin{aligned} \psi_{i\alpha j\beta} =& \varphi_{i\alpha} \circ \varphi_{j\beta}^{-1}, \\ (x^0, \ldots, & x^i, \ldots, x^{j-1}, x^{j+1}, \ldots, x^n) \\ & \mapsto (x^0, \ldots, x^{i-1}, x^{i+1}, \ldots, x^{j-1}, \beta \sqrt{1 - \sum_{k \neq j} (x^k)^2}, x^{j+1}, \ldots, x^n) \end{aligned}$$

are C^∞.
There are other compatible atlases, e.g. the stereographic projection.

Example 5 The group $GL(n, \mathbb{R})$ of invertible real $n \times n$ matrices is a smooth manifold as an open subset of $\mathbb{R}^{n^2}$. It is an open subset since it is a complement of the closed surface determined by the polynomial equation $\det A = 0$.

9.3.2 Differentiable maps

Let M, N be a pair of smooth manifolds (of dimensions m, n) and $f : M \to N$ a continuous map. If (U, ϕ) is a local coordinate chart on M and (V, ψ) a coordinate chart on N then we have a map $\psi \circ f \circ \phi^{-1}$ from some open subset of $\mathbb{R}^m$ to an open subset of $\mathbb{R}^n$. If the composite map is smooth for any pair of coordinate charts we say that f is smooth. The reader should convince himself that the condition of smoothness for f does not depend on the choice of coordinate charts. From elementary results in differential calculus it follows that if $g : N \to P$ is another smooth map then also $g \circ f : M \to P$ is smooth.

Note that we can write the map $\psi \circ f \circ \phi^{-1}$ as

$$y = (y^1, \ldots, y^n) = (y^1(x^1, \ldots, x^m), \ldots, y^n(x^1, \ldots, x^m))$$

in terms of the Cartesian coordinates. Smoothness of f simply means that the coordinate functions $y^i(x^1, x^2, \ldots, x^m)$ are smooth functions.

Remark In a given topological space M one can often construct different inequivalent smooth structures. That is, one might be able to construct atlases $\{(U_\alpha, \phi_\alpha)\}$ and $\{(V_\alpha, \psi_\alpha)\}$ such that both define a structure of smooth manifold, say M_U and M_V, but the manifolds M_U, M_V are not diffeomorphic (see the definition below). A famous example of this phenomen are the spheres S^7, S^{11} (John Milnor, 1956). On the sphere S^7 there are exactly 28 and on S^{11} 992 inequivalent differentiable structures! On the Euclidean space $\mathbb{R}^4$ there is an infinite number of differentiable structures (S.K. Donaldson, 1983). In dimensions $n = 1, 2, 3, 5, 6$ there is exactly one sphere (up to a diffeomorphism) whereas the case $n = 4$ is still open.

A diffeomorphism is a one-to-one smooth map $f : M \to N$ such that its inverse $f^{-1} : N \to M$ is also smooth. The set of diffeomorphisms $M \to M$ forms a group $\mathrm{Diff}(M)$. A smooth map $f : M \to N$ is *an immersion* if at each point $p \in M$ the rank of the derivative $\frac{dh}{dx}$ is equal to the dimension of M. Here $h = \psi \circ f \circ \phi^{-1}$ with the notation as before. Finally $f : M \to N$ is *an embedding* if f is injective and it is an immersion; in that case $f(M) \subset N$ is *an embedded submanifold.*

A smooth curve on a manifold M is a smooth map γ from an open interval of the real axes to M. Let $p \in M$ and (U, ϕ) a coordinate chart with $p \in U$. Assume that curves γ_1, γ_2 go through p, let us say $p = \gamma_i(0)$. We say that the curves are equivalent at p, $\gamma_1 \sim \gamma_2$, if

$$\frac{d}{dt}\phi(\gamma_1(t))|_{t=0} = \frac{d}{dt}\phi(\gamma_2(t))|_{t=0}.$$

This relation does not depend on the choice of (U, ϕ) as is easily seen by the help of the chain rule:

$$\frac{d}{dt}\psi(\gamma_1(t)) - \frac{d}{dt}\psi(\gamma_2(t)) = (\psi \circ \phi^{-1})' \cdot \left(\frac{d}{dt}\phi(\gamma_1(t)) - \frac{d}{dt}\phi(\gamma_2(t))\right) = 0$$

at the point $t = 0$. Clearly if $\gamma_1 \sim \gamma_2$ and $\gamma_2 \sim \gamma_3$ at the point p then also $\gamma_1 \sim \gamma_3$ and $\gamma_2 \sim \gamma_1$. Trivially $\gamma \sim \gamma$ for any curve γ through p so that " $\sim$ " is an equivalence relation.

A tangent vector v at a point p is an equivalence class of smooth curves $[\gamma]$ through p. For a given chart (U, ϕ) at p the equivalence classes are parametrized by the vector

$$\frac{d}{dt}\phi(\gamma(t))|_{t=0} \in \mathbb{R}^n.$$

Thus the space T_pM of tangent vectors $v = [\gamma]$ inherits the natural linear structure of $\mathbb{R}^n$. Again, it is a simple exercise using the chain rule that the linear structure does not depend on the choice of the coordinate chart.

We denote by TM the disjoint union of all the tangent spaces T_pM. This is called the *tangent bundle* of M. We shall define a smooth structure on TM. Let $p \in M$ and (U, ϕ) a coordinate chart at p. Let $\pi : TM \to M$ the natural projection, $(p, v) \mapsto p$. Define $\tilde{\phi} : \pi^{-1}(U) \to \mathbb{R}^n \times \mathbb{R}^n$ as

$$\tilde{\phi}(p, [\gamma]) = (\phi(p), \frac{d}{dt}\phi(\gamma(t))|_{t=0}).$$

If now (V, ψ) is another coordinate chart at p then

$$(\tilde{\phi} \circ \tilde{\psi}^{-1})(x, v) = (\phi(\psi^{-1}(x)), (\phi \circ \psi^{-1})'(x)v),$$

by the chain rule. It follows that $\tilde{\phi} \circ \tilde{\psi}^{-1}$ is smooth in its domain of definition and thus the pairs $(\pi^{-1}(U), \tilde{\phi})$ form an atlas on TM, giving TM a smooth structure.

Example 1 If M is an open set in $\mathbb{R}^n$ then $TM = M \times \mathbb{R}^n$.

Example 2 Let $M = S^1$. Writing $z \in S^1$ as a complex number of unit modulus, consider curves through z written as $\gamma(t) = ze^{ivt}$ with $v \in \mathbb{R}$. This gives in fact a parametrization for the equivalence classes $[\gamma]$ as vectors in $\mathbb{R}$. The tangent spaces at different points z_1, z_2 are related by the phase shift $z_1 z_2^{-1}$ and it follows that TM is simply the product $S^1 \times \mathbb{R}$.

Example 3 In general, $TM \neq M \times \mathbb{R}^n$. The simplest example for this is the unit sphere $M = S^2$. Using the spherical coordinates, for example, one can identify the tangent space at a given point (θ, ϕ) as the plane $\mathbb{R}^2$. However, there is no natural way to identify the tangent spaces at different points on the sphere; the sphere is not parallelizable. This is the content of the famous hairy ball theorem. Any smooth vector field on the sphere has zeros. (If there were a globally nonzero vector field on S^2 we would obtain a basis in all the tangent spaces by taking a (oriented) unit normal vector field to the given vector field. Together they would form a basis in the tangent spaces and could be used for identifying the tangent spaces as a standard $\mathbb{R}^2$.)

Exercise The unit 3-sphere S^3 can be thought of complex unitary 2×2 matrices with determinant =1. Use this fact to show that the tangent bundle is trivial, $TS^3 = S^3 \times \mathbb{R}^3$.

Let $f : M \to N$ be a smooth map. We define a linear map

$$T_pf : T_pM \to T_{f(p)}N, \text{ as } T_pf \cdot [\gamma] = [f \circ \gamma],$$

where γ is a curve through the point p. This map is expressed in terms of local coordinates as follows. Let (U, ϕ) be a coordinate chart at p and (V, ψ) a chart at $f(p) \in N$. Then the coordinates for $[\gamma] \in T_pM$ are $v = \frac{d}{dt}\phi(\gamma(t))|_{t=0}$ and the coordinates for $[f \circ \gamma] \in T_{f(p)}N$ are $w = \frac{d}{dt}\psi(f(\gamma(t)))|_{t=0}$. But by the chain rule,

$$w = (\psi \circ f \circ \phi^{-1})'(x) \cdot \frac{d}{dt}\phi(\gamma(t))|_{t=0} = (\psi \circ f \circ \phi^{-1})'(x) \cdot v$$

with $x = \phi(p)$. Thus in local coordinates the linear map $T_p f$ is the derivative of $\psi \circ f \circ \phi^{-1}$ at the point x. Putting together all the maps $T_p f$ we obtain a map

$$Tf : TM \to TN.$$

Theorem 9.1 *The map $Tf : TM \to TN$ is smooth.*

Proof. Recall that the coordinate charts $(U, \phi), (V, \psi)$ on M, N, respectivly, lead to coordinate charts $(\pi^{-1}(U), \tilde{\phi})$ and $(\pi^{-1}(V), \tilde{\psi})$ on TM, TN. Now

$$(\tilde{\psi} \circ Tf \circ \tilde{\phi}^{-1})(x, v) = ((\psi \circ f \circ \phi^{-1})(x), (\psi \circ f \circ \phi^{-1})'(x)v)$$

for $(x, v) \in \tilde{\phi}(\pi^{-1}(U)) \in \mathbb{R}^m \times \mathbb{R}^m$. Both component functions are smooth and thus Tf is smooth by definition.

If $f : M \to N$ and $g : N \to P$ are smooth maps then $g \circ f : M \to P$ is smooth and

$$T(g \circ f) = Tg \circ Tf.$$

To see this, the curve γ through $p \in M$ is first mapped to $f \circ \gamma$ through $f(p) \in N$ and further, by Tg, to the curve $g \circ f \circ \gamma$ through $g(f(p)) \in P$.

In terms of local coordinates x_i at p, y_i at $f(p)$ and z_i at $g(f(p))$ the chain rule becomes the standard formula,

$$\frac{\partial z^i}{\partial x^j} = \sum_k \frac{\partial z^i}{\partial y^k} \frac{\partial y^k}{\partial x^j}.$$

9.3.3 Manifold with a Boundary

Let $\mathbb{H}$ be the "upper" half-space: $\mathbb{H}^m = \{(x^1, \ldots, x^m) \in \mathbb{R}^m \mid x^m \geq 0\}$.
Now require for the coordinate functions: $\varphi_i : U_i \to U_i' \subset \mathbb{H}^m$, where U_i' is open in $\mathbb{H}^m$. (The topology on $\mathbb{H}^m$ is the relative topology induced from $\mathbb{R}^m$.)
Points with coordinate $x^m = 0$ belong to the **boundary** of M (denoted by ∂M). The transition functions must now satisfy: $\psi_{ij} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$ are C^∞ in an open set of $\mathbb{R}^m$ which contains $\varphi_j(U_i \cap U_j)$. .

9.3.4 Vector fields

We denote by $C^\infty(M)$ the algebra of smooth real valued functions on M. *A derivation* of the algebra $C^\infty(M)$ is a linear map $d : C^\infty(M) \to C^\infty(M)$ such that

$$d(fg) = d(f)g + fd(g)$$

for all f, g. Let $v \in T_p M$ and $f \in C^\infty(M)$. Choose a curve γ through p representing v. Set

$$v \cdot f = \frac{d}{dt} f(\gamma(t))|_{t=0}.$$

Clearly $v : C^\infty(M) \to \mathbb{R}$ is linear. Furthermore,

$$v \cdot (fg) = \frac{d}{dt} f(\gamma(t))|_{t=0} g(\gamma(0)) + f(\gamma(0)) \frac{d}{dt} g(\gamma(t))|_{t=0} = (v \cdot f) g(p) + f(p)(v \cdot g).$$

A vector field on a manifold M is a smooth distribution of tangent vectors on M, that is, a smooth map $X : M \to TM$ such that $X(p) \in T_pM$. From the previous formula follows that a vector field defines a derivation of $C^\infty(M)$; take above $v = X(p)$ at each point $p \in M$ and the right- hand-side defines a smooth function on M and the operation satisfies the Leibnitz' rule.

We denote by $D^1(M)$ the space of vector fields on M. As we have seen, a vector field gives a linear map $X : C^\infty(M) \to C^\infty(M)$ obeying the Leibnitz' rule. Conversely, one can prove that any derivation of the algebra $C^\infty(M)$ is represented by a vector field.

One can develop an algebraic approach to manifold theory. In that the commutative algebra $\mathcal{A} = C^\infty(M)$ plays a central role. Points in M correspond to *maximal ideals* in the algebra $\mathcal{A}$. Namely, any point p defines the ideal $I_p \subset \mathcal{A}$ consisting of all functions which vanish at the point p.

The action of a vector field on functions is given in terms of local coordinates $x^1, \dots, x^n$ as follows. If $v = X(p)$ is represented by a curve γ then

$$(X \cdot f)(p) = \frac{d}{dt} f(\gamma(t))_{t=0} = \sum_k \frac{\partial f}{\partial x^k} \frac{dx^k}{dt}(t=0) \equiv \sum_k X^k(x) \frac{\partial f}{\partial x^k}.$$

Thus a vector field is locally represented by the vector valued function $(X^1(x), \dots, X^n(x))$.

In addition of being a real vector space, $D^1(M)$ is a *left module* for the algebra $C^\infty(M)$. This means that we have a linear left multiplication $(f, X) \mapsto fX$. The value of fX at a point p is simply the vector $f(p)X(p) \in T_pM$.

As we have seen, in a coordinate system x_i a vector field defines a derivation with local representation $X = \sum_k X^k \frac{\partial}{\partial x^k}$. In a second coordinate system x'^k we have a representation $X = \sum X'^k \frac{\partial}{\partial x'^k}$. Using the chain rule for differentiation we obtain the coordinate transformation rule

$$X'^k(x') = \sum_j \frac{\partial x'^k}{\partial x^j} X^j(x),$$

for $x'^k = x'^k(x^1, \dots, x^n)$.

We shall denote $\partial_k = \frac{\partial}{\partial x^k}$ and we use Einstein's summation convention over repeated indices,

Let $X, Y \in D^1(M)$. We define a new derivation of $C^\infty(M)$, the commutator $[X, Y] \in D^1(M)$, by

$$[X, Y]f = X(Yf) - Y(Xf).$$

We prove that this is indeed a derivation of $C^\infty(M)$.

$$\begin{aligned}[X,Y](fg) &= X(Y(fg)) - Y(X(fg)) = X(fYg + gYf) - Y(fXg + gXf) \\ &= (Xf)(Yg) + fX(Yg) + (Xg)(Yf) + gX(Yf) - (Yf)(Xg) - fY(Xg) \\ &\quad -(Yg)(Xf) - gY(Xf) = f[X,Y]g + g[X,Y]f.\end{aligned}$$

Writing $X = X^k\partial_k$ and $Y = Y^k\partial_k$ we obtain the coordinate expression

$$[X,Y]^k = X^j\partial_j Y^k - Y^j\partial_j X^k.$$

Thus we may view $D^1(M)$ simply as the space of first order linear partial differential operators on M with the ordinary commutator of differential operators. The commutator $[X,Y]$ is also called the *Lie bracket* on $D^1(M)$. It has the basic properties of a Lie algebra:

- $[X,Y]$ is linear in both arguments
- $[X,Y] = -[Y,X]$
- $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0.$

Exercise Check the relations

$$[X,fY] = f[X,Y] + (Xf)Y, \text{ and } [fX,Y] = f[X,Y] - (Yf)X$$

for $X,Y \in D^1(M)$ and $f \in C^\infty(M)$.

Let $f : M \to N$ be a diffeomorphism and $X \in D^1(M)$. We can define a vector field $Y = f_*X$ on N by setting $Y(q) = T_pf \cdot X(p)$ for $q = f(p)$. In terms of local coordinates,

$$Y = Y^k\frac{\partial}{\partial y^k} = X^j\frac{\partial y^k}{\partial x^j}\frac{\partial}{\partial y^k}.$$

In the case $M = N$ this gives back the coordinate transformation rule for vector fields.

9.3.5 The flow generated by a vector field

Let $X \in D^1(M)$. Consider the differential equation

$$X(\gamma(t)) = \frac{d}{dt}\gamma(t)$$

for a smooth curve γ. In terms of local coordinates this equation is written as

$$X^k(x(t)) = \frac{d}{dt}x^k(t), k = 1,2,\ldots,n. \qquad (1)$$

By the theory of ordinary differential equations this system has locally, at a neighborhood of an initial point $p = \gamma(0)$, a unique solution. However, in general the solution does not need to extend to $-\infty < t < +\infty$ except in the case when M is a compact manifold. The (local) solution γ is called *an integral curve* of X through the point p.

The integral curves for a vector field X define a (local) *flow* on the manifold M. This is a (local) map

$$f : \mathbb{R} \times M \to M$$

given by $f(t,p) = \gamma(t)$ where γ is the integral curve through p. We have the identity

$$f(t+s,p) = f(t, f(s,p)),$$

which follows from the uniqueness of the local solution to the first order ordinary differential equation. In coordinates,

$$\frac{d}{dt} f^k(t, f(s,x)) = X^k(f(t, f(s,x)))$$

and

$$\frac{d}{dt} f^k(t+s,x) = X^k(f(t+s,x)).$$

Thus both sides of (1) obey the same differential equation. Since the initial conditions are the same, at $t=0$ both sides are equal to $f(0, f(s,x)) = f(s,x)$, the solutions must agree.

Denoting $f_t(p) = f(t,p)$, observe that the map $\mathbb{R} \to \text{Diff}_{loc}(M)$, $t \mapsto f_t$, is a homomorphism,

$$f_t \circ f_s = f_{t+s}.$$

Thus we have a *one parameter group of (local) transformations* f_t on M. In the case when M is compact we actually have globally globally defined transformations on M. The vector field X is called the infinitesimal generator of the flow $(t,p) \mapsto f_t(p)$.

Example Let $X(r,\phi) = (-r\sin\phi, r\cos\phi)$ be a vector field on $M = \mathbb{R}^2$. The integral curves are solutions of the equations

$$\begin{aligned} x'(t) &= -r(t)\sin\phi(t) \\ y'(t) &= r(t)\cos\phi(t) \end{aligned}$$

and the solutions are easily seen to be given by $(x(t), y(t)) = (r_0\cos(\phi+\phi_0), r_0\sin(\phi+\phi_0))$, where the initial condition is specified by the constants ϕ_0, r_0. The one parameter group of tranformations generated by the vector field X is then the group of rotations in the plane.

Given a vector field X, the corresponding flow in local coordinates is often denoted by

$$\sigma_t^\mu(x) = \sigma^\mu(t,x) = \exp(tX)x^\mu = (e^{tX})x^\mu$$

and called the exponentiation of X. This is because

$$\begin{aligned}\sigma_t^\mu(x) &= x^\mu + t\frac{d\sigma^\mu(s,x)}{ds}\bigg|_{s=0} + \frac{1}{2!}t^2\,\frac{d^2\sigma^\mu(s,x)}{ds^2}\bigg|_{s=0} + \cdots \\ &= \left(1 + t\frac{d}{ds} + \frac{1}{2!}t^2\frac{d^2}{ds^2} + \cdots\right)\sigma^\mu(s,x)\bigg|_{s=0} \\ &= e^{t\frac{d}{ds}}\sigma^\mu(s,x)\bigg|_{s=0} = e^{tX}x^\mu \ .\end{aligned}$$

Further reading: M. Nakahara: *Geometry, Topology and Physics,* Institute of Physics Publ. (1990), sections 5.1 - 5.3 . S. S. Chern, W.H. Chen, K.S. Lam: *Lectures on Differential Geometry,* World Scientific Publ. (1999), Chapter 1.

9.3.6 Dual Vector Space

Let V be a complex vector space and f a linear function $V \to \mathbb{C}$. Now $V^* = \{f|f \text{ is a linear function } V \to \mathbb{C}\}$ is also a complex vector space, the **dual vector space** to V:

- $(f_1 + f_2)(\vec{v}) = f_1(\vec{v}) + f_2(\vec{v})$
- $(af)(\vec{v}) = a(f(\vec{v}))$
- $\vec{0}_{V^*}(\vec{v}) = 0 \quad \forall \vec{v} \in V$

The elements of V^* are called the **dual vectors**.
Let $\{\vec{e}_1, \ldots, \vec{e}_n\}$ be a basis of V. Then any vector $\vec{v} \in V$ can be written as $\vec{v} = v^i\vec{e}_i$. We define a **dual basis** in V^* such that $e^{*i}(\vec{e}_j) = \delta^i{}_j$. From this it follows that $\dim V = \dim V^* = n$ (dual basis $= \{e^{*1}, \ldots, e^{*n}\}$). We can then expand any $f \in V^*$ as $f = f_i e^{*i}$ for some coefficients $f_i \in \mathbb{C}$. Now we have

$$f(\vec{v}) = f_i e^{*i}(v^j\vec{e}_j) = f_i v^j e^{*i}(\vec{e}_j) = f_i v^i.$$

This can be interpreted as an **inner product**:

$$\begin{aligned}&< \, , \, > : \ V^* \times V \to \mathbb{C} \\ &< f, \vec{v} > = \ f_i v^i.\end{aligned}$$

(Note that this is not the same inner product $< \ | \ >$ which we discussed before: $< , > : \ V^* \times V \to \mathbb{C}$ but $< \ | \ > : \ V \times V \to \mathbb{C}$.)

Pullback: Let $f: V \to W$ and $g: W \to \mathbb{C}$ be linear maps ($g \in W^*$). It follows that $g \circ f: V \to \mathbb{C}$ is a linear map, i.e. $g \circ f \in V^*$.

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ & \searrow & \downarrow g \\ g \circ f & & \mathbb{C} \end{array}$$

Now f induces a map $f^*: W^* \to V^*, g \mapsto g \circ f$ i.e. $f^*(g) = g \circ f \in V^*$. $f^*(g)$ is called the **pullback** (takaisinveto) of g.

Dual of a Dual: Let $\omega: V^* \to \mathbb{C}$ be a linear function ($\omega \in (V^*)^*$). Every $\vec{v} \in V$ induces via inner product $\omega_{\vec{v}} \in (V^*)^*$ defined by $\omega_{\vec{v}}(f) =< f, \vec{v} >$. On the other hand, it can be shown this gives all $\omega \in (V^*)^*$. So we can identify $(V^*)^*$ with V.

Tensors: A tensor of type (p, q) is a function of p dual vectors and q vectors, and is linear in its every argument[1]

$$T: \overbrace{V^* \times \ldots \times V^*}^{p} \times \overbrace{V \times \ldots \times V}^{q} \to \mathbb{C}.$$

Examples: (0,1) tensor = dual vector : $V \to \mathbb{C}$
(1,0) tensor = (dual of a dual) vector
(1,2) tensor: $T: V^* \times V \times V \to \mathbb{C}$. Choose basis $\{\vec{e}_i\}$ in V and $\{e^{*i}\}$ in V^*:

$$T(f, \vec{v}, \vec{w}) = T(f_i e^{*i}, v^j \vec{e}_j, w^k \vec{e}_k) = f_i v^j w^k \overbrace{T(e^{*i}, \vec{e}_j, \vec{e}_k)}^{\equiv T^i{}_{jk}} = T^i{}_{jk} f_i v^j w^k,$$

where $T^i{}_{jk}$ are the components of the tensor and they uniquely determine the tensor. Note the positioning of the indices.
In general, (p, q) tensor components have p upper and q lower indices.

Tensor product: Let R be a (p, q) tensor and S be a (p', q') tensor. Then $T = R \otimes S$ is defined as the $(p + p', q + q')$ tensor:

$$\begin{aligned} &T(f_1, \ldots, f_p; f_{p+1}, \ldots, f_{p+p'}; \vec{v}_1, \ldots, \vec{v}_q; \vec{v}_{q+1}, \ldots, \vec{v}_{q+q'}) \\ &\qquad = R(f_1, \ldots, f_p; \vec{v}_1, \ldots, \vec{v}_q) S(f_{p+1}, \ldots, f_{p+p'}; \vec{v}_{q+1}, \ldots, \vec{v}_{q+q'}). \end{aligned}$$

In terms of components:

$$T^{i_1 \ldots i_p i_{p+1} \ldots i_{p+p'}}_{j_1 \ldots j_q j_{q+1} \ldots j_{q+q'}} = R^{i_1 \ldots i_p}_{j_1 \ldots j_q} S^{i_{p+1} \ldots i_{p+p'}}_{j_{q+1} \ldots j_{q+q'}}$$

[1] So T is a multilinear object.

Contraction: This is an operation that produces a $(p-1,q-1)$ tensor from a (p,q) tensor:

$$\underbrace{T}_{(p,q)} \mapsto \underbrace{T_{c(ij)}}_{(p-1,q-1)} ,$$

where the $(p-1,q-1)$ tensor $T_{c(ij)}$ is

$$T_{c(ij)}(f_1,\dots,f_{p-1};\vec{v}_1,\dots,\vec{v}_{q-1}) = T(f_1,\dots,\overbrace{e^{*k}}^{i^{th}},\dots,f_{p-1};\vec{v}_1,\dots,\overbrace{\vec{e}_k}^{j^{th}},\dots,\vec{v}_{q-1}).$$

Note the sum over k in the formula above. In component form this is

$$T_{c(ij)}{}^{l_1\dots l_{p-1}}{}_{m_1\dots m_{q-1}} = T^{l_1\dots l_{i-1}kl_i\dots l_{p-1}}{}_{m_1\dots m_{j-1}km_j\dots m_{q-1}}$$

Now we can return to calculus on manifolds.

9.3.7 1-forms (i.e. cotangent vectors)

Tangent vectors of a differentiable manifold M at point p were elements of the vector space T_pM. **Cotangent vectors** or **1-forms** are their dual vectors, i.e. linear functions $T_pM \to \mathbb{R}$. In other words, they are elements of the dual vector space T_p^*M. Let $w \in T_p^*M$ and $v \in T_pM$, then the inner product $< \, , \, >: T_p^*M \times T_pM \to \mathbb{R}$ is

$$< w, v >= w(v) \in \mathbb{R}.$$

The inner product is bilinear:

$$\begin{aligned} < w, \alpha_1 v_1 + \alpha_2 v_2 > &= w(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 < w, v_1 > +\alpha_2 < w, v_2 > \\ < \alpha_1 w_1 + \alpha_2 w_2, v > &= (\alpha_1 w_1 + \alpha_2 w_2)(v) = \alpha_1 < w_1, v > +\alpha_2 < w_2, v > . \end{aligned}$$

Let $\{e_\mu\} = \{\frac{\partial}{\partial x^\mu}\}$ be a coordinate basis of T_pM. (Note that the correct notation would be $\{\left(\frac{\partial}{\partial x^\mu}\right)_p\}$, but this is somewhat cumbersome so we use the shorter notation.) The dual basis is denoted by $\{dx^\mu\}$ and it satisfies by definition

$$< dx^\mu, \frac{\partial}{\partial x^\nu} >= dx^\mu(\frac{\partial}{\partial x^\nu}) = \delta^\mu{}_\nu.$$

Now we can expand $w = w_\mu dx^\mu$ and $v = v^\nu \frac{\partial}{\partial x^\nu}$. Then

$$w(v) =< w, v >= w_\mu v^\nu dx^\mu(\frac{\partial}{\partial x^\nu}) = w_\mu v^\mu.$$

Consider now a function $f \in \mathcal{F}(M)$ (i.e. f is a smooth map $M \to \mathbb{R}$). Its **differential** $df \in T_p^*M$ is the map

$$df(v) =< df, v >\equiv v(f) = v^\mu \frac{\partial f}{\partial x^\mu}.$$

Thus the components of df are $\frac{\partial f}{\partial x^\mu}$ and

$$df = \frac{\partial f}{\partial x^\mu} dx^\mu.$$

Consider two coordinate patches U_i and U_j with $p \in U_i \cap U_j$. Let $x = \varphi_i(p)$ and $y = \varphi_j(p)$ be the coordinates in U_i and U_j respectively. We can derive how the components of a 1-form transform under the change of coordinates:
Let $w = w_\mu dx^\mu = \tilde{w}_\nu dy^\nu \in T_p^* M$ and $v = v^\rho \frac{\partial}{\partial x^\rho} = \tilde{v}^\sigma \frac{\partial}{\partial y^\sigma} \in T_p M$ be a 1-form and a vector. We already know that $\tilde{v}^\nu = \frac{\partial y^\nu}{\partial x^\mu} v^\mu$, so we get

$$w(v) = w_\mu v^\mu = \tilde{w}_\nu \tilde{v}^\nu = \tilde{w}_\nu \frac{\partial y^\nu}{\partial x^\mu} v^\mu,$$

so we find the transformed components

$$w_\mu = \tilde{w}_\nu \frac{\partial y^\nu}{\partial x^\mu} \qquad \text{or} \quad \tilde{w}_\mu = w_\nu \frac{\partial x^\nu}{\partial y^\mu}.$$

The dual basis vectors transform as

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

9.3.8 Tensors

A tensor of type (q, r) is a multilinear map

$$T : \overbrace{T_p^* M \times \ldots \times T_p^* M}^{q} \times \overbrace{T_p M \times \ldots \times T_p M}^{r} \to \mathbb{R}.$$

Denote the set of type (q, r) tensors at $p \in M$ by $T_{r,p}^q(M)$. Note that $T_{0,p}^1 = (T_p^* M)^* = T_p M$ and $T_{1,p}^0(M) = T_p^* M$.
The basis of $T_{r,p}^q$ is

$$\left\{ \frac{\partial}{\partial x^{\mu_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_q}} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_r} \right\}.$$

The basis vectors satisfy (as a mapping $T_p^* M \times \ldots \times T_p^* M \times T_p M \times \ldots \times T_p M \to \mathbb{R}$):

$$\left(\frac{\partial}{\partial x^{\mu_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_q}} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_r} \right) \left(dx^{\alpha_1}, \ldots, dx^{\alpha_q}, \frac{\partial}{\partial x^{\beta_1}}, \ldots, \frac{\partial}{\partial x^{\beta_r}} \right)$$
$$= \delta^{\alpha_1}{}_{\mu_1} \ldots \delta^{\alpha_q}{}_{\mu_q} \delta^{\nu_1}{}_{\beta_1} \ldots \delta^{\nu_r}{}_{\beta_r}.$$

(Note that $\frac{\partial}{\partial x^\mu}(dx^\alpha) \equiv < dx^\alpha, \frac{\partial}{\partial x^\mu} >= \delta^\alpha{}_\mu$. On the left $\frac{\partial}{\partial x^\mu}$ is interpreted as an element of $(T_p^* M)^*$.)
We can expand as $T = T^{\mu_1 \ldots \mu_q}{}_{\nu_1 \ldots \nu_r} \left\{ \frac{\partial}{\partial x^{\mu_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_q}} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_r} \right\}$ so

$$T(w_1, \ldots, w_q; v_1, \ldots, v_r) = T^{\mu_1 \ldots \mu_q}{}_{\nu_1 \ldots \nu_r} w_{1\mu_1} \ldots w_{q\mu_q} v_1^{\nu_1} \ldots v_r^{\nu_r}.$$

The **tensor product** of tensors $T \in T^q_{r,p}(M)$ and $U \in T^s_{t,p}(M)$ is the tensor $T \otimes U \in T^{q+s}_{r+t,p}(M)$ with

$$
\begin{aligned}
(T \otimes U)(w_1, \ldots, w_q, w_{q+1}, \ldots, w_{q+s}; v_1, \ldots, v_r, v_{r+1}, \ldots, v_{r+t}) \\
= T(w_1, \ldots, w_q; v_1, \ldots, v_r) U(w_{q+1}, \ldots, w_{q+s}; v_{r+1}, \ldots, v_{r+t}). \\
= T^{\mu_1 \ldots \mu_q}{}_{\nu_1 \ldots \nu_r} w_{1\mu_1} \ldots w_{q\mu_q} v_1^{\nu_1} \ldots v_r^{\nu_r} \cdot \\
U^{\alpha_1 \ldots \alpha_s}{}_{\beta_1 \ldots \beta_t} w_{(q+1)\alpha_1} \ldots w_{(q+s)\alpha_s} v_{r+1}^{\beta_1} \ldots v_{r+t}^{\beta_t}.
\end{aligned}
$$

Contraction maps a tensor $T \in T^q_{r,p}(M)$ to a tensor $T' \in T^{q-1}_{r-1,p}(M)$ with components

$$
T'^{\mu_1 \ldots \mu_{q-1}}{}_{\nu_1 \ldots \nu_{r-1}} = T^{\mu_1 \ldots \mu_{i-1} \rho \mu_i \ldots \mu_{q-1}}{}_{\nu_1 \ldots \nu_{j-1} \rho \nu_j \ldots \nu_{r-1}}
$$

Under a coordinate transformation, a tensor of type (q, r) transforms like a product of q vectors and r one-forms (note that $v_1 \otimes \cdots \otimes v_q \otimes w_1 \otimes \ldots \otimes w_r$ is one example of a (q, r) tensor). For example $T \in T^1_{2,p}(M)$ tensor of type $(1, 2)$:

$$
T = T^\alpha{}_{\beta_1 \beta_2} \frac{\partial}{\partial x^\alpha} \otimes dx^{\beta_1} \otimes dx^{\beta_2} = \tilde{T}^\mu{}_{\nu_1 \nu_2} \frac{\partial}{\partial y^\mu} \otimes dy^{\nu_1} \otimes dy^{\nu_2}
$$

gives us the transformation rule for the components

$$
\tilde{T}^\mu{}_{\nu_1 \nu_2} = \frac{\partial y^\mu}{\partial x^\alpha} \frac{\partial x^{\beta_1}}{\partial y^{\nu_1}} \frac{\partial x^{\beta_2}}{\partial y^{\nu_2}} T^\alpha{}_{\beta_1 \beta_2}
$$

9.3.9 Tensor Fields

Suppose that a vector $v(p)$ has been assigned to every point p in M. This is a **(smooth) vector field**, if for every C^∞ function $f \in \mathcal{F}$ the function $v(p)(f) : M \to \mathbb{R}$ is also a smooth function. We denote $v(p)(f)$ by $v[f]$. The set of smooth vector fields on M is denoted by $\chi(M)$.

Smooth cotangent vector field : For every $p \in M$ there is $w(p) \in T^*_p M$ such that if $V \in \chi(M)$, then the function

$$
\begin{aligned}
w[V] \quad &: \quad M \to \mathbb{R} \\
p \quad &\mapsto \quad w[V](p) = w(p)(V(p))
\end{aligned}
$$

is smooth. The set of cotangent vector fields is denoted by $\Omega^1(M)$.

Smooth (q, r)-tensor field : If for all $p \in M$ there is $T(p) \in T^q_{r,p}(M)$ such that if $w_1, \ldots, w_q$ are smooth cotangent vector fields and $v_1, \ldots, v_r$ are smooth tangent vector fields, then the map

$$
p \mapsto T[w_1, \ldots, w_q; v_1, \ldots, v_r](p) = T(p)(w_1(p), \ldots, w_q(p); v_1(p), \ldots, v_r(p))
$$

is smooth on M.

9.3.10 Differential Map and Pullback

Let M and N be differentiable manifolds and $f : M \to N$ smooth.
f induces a map called the **differential map** (työntökuvaus) $f_* : T_pM \to T_pN$. It is defined as follows:
If $g \in \mathcal{F}(N)$ (i.e. $g : N \to \mathbb{R}$ smooth), and $v \in T_pM$, then

$$(f_*v)[g] = v[g \circ f].$$

In other words, if v characterizes the rate of change of a function along a curve $c(t)$, then f_*v characterizes the rate of change of a function along the curve $f(c(t))$.
Let x be local coordinates on M and y be local coordinates on N, "$y = f(x)$". Also let $v = v^\mu \frac{\partial}{\partial x^\mu}$ and $(f_*v)^\nu \frac{\partial}{\partial y^\nu}$. Then

$$v[g \circ f] = v^\mu \frac{\partial(g(f(x)))}{\partial x^\mu} = v^\mu \frac{\partial g}{\partial y^\nu} \frac{\partial y^\nu}{\partial x^\mu} \equiv (f_*v)^\nu \frac{\partial g}{\partial y^\nu}$$

and we get

$$(f_*v)^\nu = v^\mu \frac{\partial y^\nu}{\partial x^\mu}, \text{ where } y = f(x).$$

[More precisely $x^\mu = \varphi^\mu(p)$, $y^\nu = \psi^\nu(f(p))$ and $\frac{\partial y^\nu}{\partial x^\mu} = \frac{\partial(\psi \circ f \circ \varphi^{-1})^\nu}{\partial x^\mu}$.]

Example. Let (x^1, x^2) and (y^1, y^2, y^3) be the coordinates in M and N, respectively, and let $V = a\frac{\partial}{\partial x^1} + b\frac{\partial}{\partial x^2}$ be a tangent vector at (x^1, x^2). Let $f : M \to N$ be a map whose coordinate presentation is $y = (x^1, x^2, \sqrt{1 - (x^1)^2 - (x^2)^2})$. Then

$$f_*V = V^\mu \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial}{\partial y^\alpha} = a\frac{\partial}{\partial y^1} + b\frac{\partial}{\partial y^2} - (a\frac{y^1}{y^3} + b\frac{y^2}{y^3})\frac{\partial}{\partial y^3}.$$

The function f also induces the map

$$f^* : T^*_{f(p)}N \to T^*_pM, \quad (f^*w)(v) = w(f_*v),$$

where $v \in T_pM$ and $w \in T^*_{f(p)}N$ are arbitrary. f^* is called the **pullback**.
In local coordinates, $w = w_\nu dy^\nu$,

$$w(f_*v) = w_\nu dy^\nu \left(v^\mu \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial}{\partial y^\alpha} \right) = w_\nu v^\mu \frac{\partial y^\nu}{\partial x^\mu} = (f^*w)_\mu v^\mu = (f^*w)(v),$$

from which we get

$$(f^*w)_\mu = w_\nu \frac{\partial y^\nu}{\partial x^\mu}.$$

The pullback f^* can also be generalized to $(0, r)$ tensors and similarly the differential map f_* can be generalized to $(q, 0)$ tensors.

9.3.11 Lie Derivative

Let $\sigma_t(x)$ be a flow on M generated by vector field X: $\frac{d\sigma_t^\mu(x)}{dt} = X^\mu(\sigma_t(x))$. Let Y be another vector field on M. We want to calculate the rate of change of Y along the curve $x^\mu(t) = \sigma_t^\mu(x)$.

The **Lie derivative** of a vector field Y is defined by

$$\mathcal{L}_X Y =_{\epsilon\to 0} \frac{1}{\epsilon}\left((\sigma_{-\epsilon})_* Y|_{\sigma_\epsilon(x)} - Y|_x\right).$$

Let's rewrite this in a more user-friendly form: First

$$\begin{aligned} Y|_x &= Y^\mu(x)\frac{\partial}{\partial x^\mu} \\ Y|_{\bar{x}} &= Y^\mu(\bar{x})\frac{\partial}{\partial \bar{x}^\mu}, \end{aligned}$$

where we have for the coordinates

$$\begin{aligned} \bar{x}^\mu \equiv& \sigma_\epsilon^\mu(x) = x^\mu + \epsilon X^\mu(x) + O(\epsilon^2) \\ &\Rightarrow x^\mu = \bar{x}^\mu - \epsilon X^\mu(\bar{x}^\mu) + O(\epsilon^2). \end{aligned}$$

Thus

$$Y|_{\bar{x}} = (Y^\mu(x+\epsilon X))\frac{\partial}{\partial\bar{x}^\mu} = \left(Y^\mu(x) + \epsilon X^\nu \frac{\partial Y^\mu(x)}{\partial x^\nu}\right)\frac{\partial}{\partial x^\mu}.$$

Differential map from $\bar{x}$ to x:

$$\begin{aligned} ((\sigma_{-\epsilon})_* Y|_{\bar{x}})^\alpha &= Y^\mu|_{\bar{x}}\frac{\partial x^\alpha}{\partial\bar{x}^\mu} = \left(Y^\mu(x) + \epsilon X^\nu(x)\frac{\partial Y^\mu(x)}{\partial x^\nu}\right)\Bigg(\delta^\alpha{}_\mu - \epsilon\overbrace{\frac{\partial X^\alpha(\bar{x})}{\partial\bar{x}^\mu}}^{\frac{\partial X^\alpha}{\partial x^\mu}+O(\epsilon)}\Bigg) \\ &= Y^\alpha(x) + \epsilon\left(X^\nu(x)\frac{\partial Y^\alpha}{\partial x^\nu} - Y^\mu(x)\frac{\partial X^\alpha}{\partial x^\mu}\right) + O(\epsilon^2) \\ \Rightarrow \mathcal{L}_X Y &= \left(X^\nu\frac{\partial Y^\mu}{\partial x^\nu} - Y^\nu\frac{\partial X^\mu}{\partial x^\nu}\right)\frac{\partial}{\partial x^\mu}. \end{aligned}$$

So we got

$$\mathcal{L}_X Y = \left(X^\nu\frac{\partial Y^\mu}{\partial x^\nu} - Y^\nu\frac{\partial X^\mu}{\partial x^\nu}\right)\frac{\partial}{\partial x^\mu} = [X,Y],$$

where the commutator ("Lie bracket") acts on functions by

$$[X,Y]\ f = X[Y[f]] - Y[X[f]].$$

Note that XY is not a vector field but $[X,Y]$ is:

$$XY\ f = X[Y[f]] = X^\mu\partial_\mu[Y^\nu\partial_\nu f] = \underbrace{X^\mu(\partial_\mu Y^\nu)\partial_\nu}_{\text{vector field}} f + \underbrace{X^\mu Y^\nu\partial_\mu\partial_\nu}_{\text{not a vector field}}\ f.$$

Lie derivative of a one-form: Let $w \in \Omega^1(M)$ be a one-form (cotangent vector). Define the Lie derivative of w along X as

$$\mathcal{L}_X w =_{\epsilon\to 0} \frac{1}{\epsilon}\left(\sigma_\epsilon^* w|_{\sigma_\epsilon(x)} - w|_x\right).$$

Let's simplify this. The coordinates at $\sigma_\epsilon(x)$: $y^\mu \equiv \sigma_\epsilon^\mu(x) \simeq x^\mu + \epsilon X^\mu(x)$.

$$\begin{aligned}(\sigma_\epsilon^* w)_\alpha &= w_\beta(y)\frac{\partial y^\beta}{\partial x^\alpha} = w_\beta(x+\epsilon X)\frac{\partial}{\partial x^\alpha}(x^\beta + \epsilon X^\beta)\\ &= (w_\beta(x) + \epsilon X^\mu \partial_\mu w_\beta(x))(\delta^\beta{}_\alpha + \epsilon\partial_\alpha X^\beta)\\ &= w_\alpha + \epsilon(X^\mu\partial_\mu w_\alpha + w_\mu \partial_\alpha X^\mu)\end{aligned}$$

Thus we find

$$\mathcal{L}_X w = (X^\mu \partial_\mu w_\alpha + w_\mu \partial_\alpha x^\mu)\, dx^\alpha.$$

Lie derivative of a function: A natural guess would be $\mathcal{L}_X f = X[f]$. Let's check if this works:

$$\mathcal{L}_x f =_{\epsilon\to 0} \frac{1}{\epsilon}(f(\sigma_\epsilon(x)) - f(x)) =_{\epsilon\to 0} \frac{1}{\epsilon}(f(x+\epsilon X) - f(x)) = X^\mu\partial_\mu f = Xf = X[f].$$

Thus the definition works.

Lie derivative of a tensor field: We define these using the Leibnitz rule: we require that

$$\mathcal{L}_X(t_1\otimes t_2) = (\mathcal{L}_X t_1)\otimes t_2 + t_1 \otimes (\mathcal{L}_X t_2).$$

This is true if t_1 is a function ((0,0) tensor) and t_2 is a one form or a vector field, or vice versa. (exercise)
Example: Let's find the Lie derivative of a (1,1) tensor: $t = t_\mu{}^\nu dx^\mu \otimes e_\nu$; $e_\nu = \frac{\partial}{\partial x^\nu}$.

$$\begin{aligned}\mathcal{L}_X t &= (\mathcal{L}_X t_\mu{}^\nu) dx^\mu\otimes e_\nu + t_\mu{}^\nu(\mathcal{L}_X dx^\mu)\otimes e_\nu + t_\mu{}^\nu dx^\mu \otimes(\mathcal{L}_X e_\nu)\\ &= (X^\alpha\partial_\alpha t_\mu{}^\nu)dx^\mu\otimes e_\nu + t_\mu{}^\nu(\partial_\alpha X^\mu)dx^\alpha\otimes e_\nu - t_\mu{}^\nu dx^\mu\otimes(\partial_\nu X^\alpha)e_\alpha\\ &= (X^\alpha\partial_\alpha t_\mu{}^\nu + t_\alpha{}^\nu\partial_\mu X^\alpha - t_\mu{}^\alpha\partial_\alpha X^\nu)dx^\mu\otimes e_\nu.\end{aligned}$$

[We used here $e_\nu = \frac{\partial}{\partial x^\nu}$, $(e_\nu)^\alpha = \delta_\nu{}^\alpha$, $(dx^\mu)_\alpha = \delta^\mu{}_\alpha$, $(\mathcal{L}_X e_\nu)^\alpha = X^\mu\partial_\mu(e_\nu)^\alpha - (e_\nu)^\mu\partial_\mu X^\alpha = -\partial_\nu X^\alpha$ and also $(\mathcal{L}_X dx^\mu)_\alpha = X^\nu\partial_\nu(dx^\mu)_\alpha + (dx^\mu)_\nu\partial_\alpha X^\nu = \partial_\alpha X^\mu$.]

9.3.12 Differential Forms

A **differential form** of order r (or **r-form**) is a totally antisymmetric $(0,r)$-tensor:

$$p\in S_r:\ w(v_{p(1)},\ldots,v_{p(r)}) = \mathrm{sgn}(p)\ w(v_1,\ldots,v_r),$$

where sgn(p) is the sign of the permutation p:

$$\mathrm{sgn}(p) = (-1)^{\text{number of exchanges}} = \begin{cases} +1 & \text{for an even permutation} \\ -1 & \text{for an odd permutation.} \end{cases}$$

Example: $p : (123) \to (231)$: Two exchanges $[(231) \to (213) \to (123)]$ to (123), thus p is an even permutation.
$\tilde{p} : (123) \to (321)$: One exchange to (231) and then two exchanges to (123), thus $\tilde{p}$ is an odd permutation.

The r-forms at point $p \in M$ form a vector space $\Omega_p^r(M)$. What is its basis?
We define the **wedge product** of 1-forms:

$$dx^{\mu_1} \wedge dx^{\mu_2} \wedge \ldots \wedge dx^{\mu_r} = \sum_{p \in S_r} \mathrm{sgn}(p)\; dx^{\mu_{p(1)}} \otimes \ldots \otimes dx^{\mu_{p(r)}}$$

Then $\{\, dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r} |\, \mu_1 < \mu_2 < \ldots < \mu_r \,\}$ forms the basis of $\Omega_p^r(M)$.

Examples: $dx^\mu \wedge dx^\nu = dx^\mu \otimes dx^\nu - dx^\nu \otimes dx^\mu$
$$dx^1 \wedge dx^2 \wedge dx^3 = dx^1 \otimes dx^2 \otimes dx^3 + dx^2 \otimes dx^3 \otimes dx^1 + dx^3 \otimes dx^1 \otimes dx^2 - dx^2 \otimes dx^1 \otimes dx^3 - dx^3 \otimes dx^2 \otimes dx^1 - dx^1 \otimes dx^3 \otimes dx^2.$$

Note:

- $dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r} = 0$ if the same index appears twice (or more times).
- $dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r} = \mathrm{sgn}(p) dx^{\mu_{p(1)}} \wedge \ldots \wedge dx^{\mu_{p(r)}}$. (reshuffling of terms.)

In the above basis, an r-form $w \in \Omega_p^r(M)$ is expanded

$$w = \frac{1}{r!} w_{\mu_1 \ldots \mu_r} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r}.$$

Note: the components $w_{\mu_1 \ldots \mu_r}$ are totally antisymmetric in the indices
(e.g. $w_{\mu_1 \mu_2 \mu_3 \ldots \mu_r} = -w_{\mu_2 \mu_1 \mu_3 \ldots \mu_r}$).

One can show that $\dim \Omega_p^r(M) = \frac{m!}{r!(m-r)!} = \binom{m}{r}$, where $m = \dim M$.
Note also: $\Omega_p^1(M) = T_p^*(M)$ cotangent space
$\Omega_p^0(M) = \mathbb{R}$ by convention

Now we generalize the wedge product for the products of a q-form and an r-form and call it **exterior product**:

Definition: The exterior product of a q-form ω and an r-form η is a $(q+r)$-form $\omega \wedge \eta$:

$$(\omega \wedge \eta)(v_1, \ldots, v_{q+r}) = \frac{1}{q!r!} \sum_{p \in S_{q+r}} \mathrm{sgn}(p)\omega(v_{p(1)}, \ldots, v_{p(q)}) \cdot \eta(v_{p(q+1)}, \ldots, v_{p(q+r)}).$$

If $q+r > m = \dim(M)$, then $\omega \wedge \eta = 0$. The exterior product satisfies the properties:

(i) $\omega \wedge \omega = 0$, if q is odd.

(ii) $\omega \wedge \eta = (-1)^{qr} \eta \wedge \omega$.

(iii) $(\omega \wedge \eta) \wedge \xi = \omega \wedge (\eta \wedge \xi)$.

[Proof: exercise]

Example. Take the Cartesian coordinates (x, y) in $\mathbb{R}^2$. The two-form $dxdy$ is the oriented area element (the vector product in elementary vector algebra). In polar coordinates this becomes

$$\begin{aligned} dxdy = &\ (\cos\theta dr - r\sin\theta d\theta)(\sin\theta dr + r\cos\theta d\theta) \\ = &\ \cos\theta\sin\theta drdr + r(\cos\theta)^2 drd\theta - r(\sin\theta)^2 d\theta dr - r^2 \sin\theta\cos\theta d\theta d\theta \\ = &\ rdrd\theta. \end{aligned}$$

We may assign an r-form smoothly at each point p on a manifold M, to obtain an r-form field. The r-form field will also be called an r-form for short.

The corresponding vector spaces of r-forms (r-form fields) are called $\Omega^r(M)$:

$$\begin{aligned} &\Omega^0(M) = \mathcal{F}(M) \quad \text{smooth functions on } M \\ &\Omega^1(M) = T^*(M) \quad \text{cotangent vector fields on } M \\ &\Omega^2(M) = sp\{dx^\mu \wedge dx^\nu \mid \mu < \nu\} \\ &\quad \vdots \end{aligned}$$

9.3.13 Exterior derivative

The **exterior derivative** d is a map $\Omega^r(M) \to \Omega^{r+1}(M)$,

$$\omega = \frac{1}{r!}\omega_{\mu_1 \ldots \mu_r} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r} \mapsto d\omega = \frac{1}{r!}\frac{\partial \omega_{\mu_1 \ldots \mu_r}}{dx^\nu} dx^\nu \wedge dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r}.$$

Example: dim $M = m = 3$. We have the following r-forms:

- $r = 0: \quad \omega_0 = f(x, y, z),$

- $r = 1:\quad \omega_1 = \omega_x(x,y,z)dx + \omega_y(x,y,z)dy + \omega_z(x,y,z)dz,$
- $r = 2:\quad \omega_2 = \omega_{xy}(x,y,z)dx\wedge dy + \omega_{yz}dy\wedge dz + \omega_{zx}dz\wedge dx,$
- $r = 3:\quad \omega_3 = \omega_{xyz}dx\wedge dy\wedge dz.$

The exterior derivatives are:

- $d\omega_0 = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$. Thus the components are the components of ∇f.
- $d\omega_1 = \frac{\partial \omega_x}{\partial y}dy\wedge dx + \frac{\partial \omega_x}{\partial z}dz\wedge dx + \frac{\partial \omega_y}{\partial x}dx\wedge dy + \frac{\partial \omega_y}{\partial z}dz\wedge dy + \frac{\partial \omega_z}{\partial x}dx\wedge dz + \frac{\partial \omega_z}{\partial y}dy\wedge dz$
  $= \left(\frac{\partial \omega_y}{\partial x} - \frac{\partial \omega_x}{\partial y}\right)dx\wedge dy + \left(\frac{\partial \omega_z}{\partial y} - \frac{\partial \omega_y}{\partial z}\right)dy\wedge dz + \left(\frac{\partial \omega_x}{\partial z} - \frac{\partial \omega_z}{\partial x}\right)dz\wedge dx$
  These are the components of $\nabla\times\vec{\omega}\quad (\vec{\omega} = (\omega_x,\omega_y,\omega_z))$
- $d\omega_2 = \frac{\partial \omega_{xy}}{\partial dz}dz\wedge dx\wedge dy + \frac{\partial \omega_{yz}}{\partial x}dx\wedge dy\wedge dz + \frac{\partial \omega_{zx}}{\partial y}dy\wedge dz\wedge dx$
  $= \left(\frac{\partial \omega_{yz}}{\partial x} + \frac{\partial \omega_{zx}}{\partial y} + \frac{\partial \omega_{xy}}{\partial z}\right)dx\wedge dy\wedge dz$
  The component is a divergence: $\nabla\cdot\vec{\omega}'\quad$ (where $\vec{\omega}' = (\omega_{yz},\omega_{zx},\omega_{xy})$)
- Thus the exterior derivatives correspond to the gradient, curl and divergence! $[d\omega_3 = 0]$

What is $d(d\omega)$?

$$d(d\omega) = \frac{1}{r!}\left(\underbrace{\frac{\partial^2}{\partial x^\alpha \partial x^\beta}}_{\text{symmetric in }\alpha\text{ and }\beta} w_{\mu_1\ldots\mu_r}\overbrace{dx^\alpha\wedge dx^\beta}^{\text{antisymmetric in }\alpha\text{ and }\beta}\wedge dx^{\mu_1}\wedge\ldots\wedge dx^{\mu_r}\right) = 0.$$

So $d^2 = 0$. Note that (for dim $M = 3$)

$$d(df) = d(\partial_x f dx + \partial_y f dy + \partial_z f dz) = \left(\frac{\partial^2}{\partial x\partial y} - \frac{\partial^2}{\partial y\partial x}\right)dx\wedge dy + \ldots = 0,$$

so we recover $\nabla\times\nabla f = 0$. Similarly $d(d\omega_1) = 0 \leftrightarrow \nabla\cdot\nabla\times\vec{\omega} = 0$.

If $d\omega = 0$, we say that ω is a **closed** r-form. If there exists an (r-1)-form ω_{r-1} such that $\omega_r = d\omega_{r-1}$, then we say that ω_r is an **exact** r-form.

The exterior derivative induces the sequence of maps

$$0 \xrightarrow{i} \Omega^0 \xrightarrow{d_0} \Omega^1 \xrightarrow{d_1} \Omega^2 \xrightarrow{d_2} \ldots \xrightarrow{d_{m-2}} \Omega^{m-1} \xrightarrow{d_{m-1}} \Omega^m \xrightarrow{d_m} 0,$$

where $\Omega^r = \Omega^r(M)$, i is the inclusion map $0 \hookrightarrow \Omega^0(M)$ and d_r denotes the map $d_r : \Omega^{r-1} \to \Omega^r$, $\omega \mapsto d\omega$. Since $d^2 = 0$, we have $\underbrace{Im\ d_r}_{\text{exact r-forms}} \subset \underbrace{ker\ d_{r+1}}_{\text{closed r+1 forms}}$. Such a sequence is called an **exact sequence**. This particular sequence is called the **de**

Rham complex. The quotient space $Ker\ d_{r+1}/Im\ d_r$ is called the r^{th} **de Rham cohomology group.**

A differential form ω of degree k can be thought of as a totally antisymmetric multilinear function of k vector fields, with values in $C^\infty(M)$. To evaluate $\omega(X_1, X_2, \ldots, X_k)$ one simply evaluates the form ω at an arbitrary point x on the manifold M and feeds into the form ω the tangent vectors $X_i(x)$ at the same point. The result is a smooth function of the argument x. Using this point of view one can prove (Exercise!) that the exterior derivative is given in a coordinate independent form as

$$\begin{aligned} d\omega(X_1, \ldots, X_{k+1}) = & \sum_{i=1}^{k+1} (-1)^{i-1} X_i \cdot \omega(X_1, \ldots, \hat{X}_i, \ldots, X_{k+1}) \\ & + \sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, \hat{X}_i, \ldots, \hat{X}_j, \ldots, X_{k+1}) \end{aligned}$$

where the hat means that the corresponding argument is removed. For example, when $k = 1$

$$d\omega(X, Y) = X \cdot \omega(Y) - Y \cdot \omega(X) - \omega([X, Y]).$$

Again, $X \cdot \omega(Y)$ is the derivative of the smooth function $\omega(Y)$ in the direction of X.

There is one more basic operation on differential forms, the **interior product** by a vector field X. In the coordinate free notation above this is given by

$$(i_X \omega)(X_1, \ldots, X_{k-1}) = \omega(X, X_1, \ldots, X_{k-1}).$$

This i_X maps linearly k forms to $k-1$ forms. One can prove a basic identity (Exercise!)

$$\mathcal{L}_X = d \circ i_X + i_X \circ d$$

where $\mathcal{L}_X$ is the Lie derivative acting on forms.

9.3.14 Lie Groups and Algebras

A **Lie group** G is a differentiable manifold with a group structure,

(i) product $G \times G \to G$, $(g_1, g_2) \mapsto g_1 g_2$, such that $g_1(g_2 g_3) = (g_1 g_2) g_3$,

(ii) unit element: point $e \in G$ such that $eg = ge = g\ \forall g \in G$,

(iii) inverse element: $\forall g \in G\ \exists g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$,

in such a way that the map $G \times G \to G$, $(g_1, g_2) \mapsto g_1 g_2$ is differentiable. We already know some examples: GL, SL, O, U, SU and SO.

Example: Coordinates on $\mathrm{GL}(n, \mathbb{R})$: $x^{ij}(g) = g^{ij}$ (and thus $x^{ij}(e) = \delta^{ij}$.) One chart is sufficient : $U = \mathrm{GL}(n, \mathbb{R})$. (thus U is open in any topology.)

- To be exact we don't yet have a topology on $\mathrm{GL}(n, \mathbb{R})$. We can define the topology in several (inequivalent) ways. One way would be to choose a topology manually, for instance choose the discrete or trivial topology. This is rarely a useful method. A better way of defining the topology is to choose a map f from $\mathrm{GL}(n, \mathbb{R})$ to some known topological space N and then choose the topology on $\mathrm{GL}(n, \mathbb{R})$ so that the map f is continuous, i.e. define

$$V \subset \mathrm{GL}(n, \mathbb{R}) \text{ is open} \leftrightarrow V = f^{-1}W \text{ for some } W \text{ open in } N.$$

(check that this defines a topology). Here are two possible topologies:

1. Choose $f : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}, \quad g \mapsto \det(g)$. (So we choose $N = \mathbb{R}$). The induced topology is:
$V \subset \mathrm{GL}(n, \mathbb{R})$ is open $\leftrightarrow V = f^{-1}(W)$ for some W open in $\mathbb{R}$.
Note that $\mathrm{GL}(n, \mathbb{R})$ is not Hausdorff with respect to this topology, since if $g_1, g_2 \in \mathrm{GL}(n, \mathbb{R}), \ g_1 \neq g_2$, and $\det g_1 = \det g_2$, then any open set containing g_1 also contains g_2.
2. Choose $N = \mathbb{R}^{n^2}, \quad f : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^{n^2}$ defined by

$$\begin{pmatrix} x^{11} & \cdots & x^{1n} \\ \vdots & \ddots & \vdots \\ x^{n1} & \cdots & x^{nn} \end{pmatrix} \mapsto (x^{11}, x^{12}, \ldots, x^{1n}, x^{21}, \ldots, x^{nn}) \in \mathbb{R}^{n^2}.$$

This is clearly injective, and when we define topology as above, we see that f is a homeomorphism from $\mathrm{GL}(n, \mathbb{R})$ to an open subset of $\mathbb{R}^{n^2}$. Since $\mathbb{R}^{n^2}$ is Hausdorff, so is $\mathrm{GL}(n, \mathbb{R})$ with this topology. Thus this topology is not equivalent to the one defined in the first example. This is the usual topology one has on $\mathrm{GL}(n, \mathbb{R})$.

Let $a \in G$ be a given element. We can define the **left-translation**

$$L_a : G \to G, \quad L_a(g) = ag \quad \text{(group action on itself from the left)}.$$

This is a diffeomorphism $G \to G$.
A vector field X on G is **left-invariant**, if the push satisfies

$$(L_a)_\star X|_g = X|_{ag}$$

Using coordinates, this means

$$(L_a)_\star X|_g = X^\mu(g) \frac{\partial x^\alpha(ag)}{\partial x^\mu(g)} \frac{\partial}{\partial x^\alpha}\bigg|_{ag} = X|_{ag} = X^\alpha(ag) \left.\frac{\partial}{\partial x^\alpha}\right|_{ag},$$

and thus

$$X^\alpha(ag) = X^\mu(g) \frac{\partial x^\alpha(ag)}{\partial x^\mu(g)}.$$

A left-invariant vector field is uniquely defined by its value at a point, for example at $e \in G$, because

$$X|_g = (L_g)_\star X_e \equiv L_{g\star} V,$$

where $V = X|_e \in T_e G$. Let's denote the set of left-invariant vector fields by $\mathcal{G}$. It is a vector space (since $L_{g\star}$ is a linear map); it is isomorphic with $T_e G$. Thus we have $\dim \mathcal{G} = \dim G$.

Example: The left-invariant fields of $\mathrm{GL}(n, \mathbb{R})$:

$$V = V^{ij} \left.\frac{\partial}{\partial x^{ij}}\right|_e \in T_e \mathrm{GL}(n, \mathbb{R}),$$

$$X|_g = L_{g\star} V = V^{ij} \frac{\partial(\overbrace{x^{kl}(g) x^{lm}(e)}^{=x^{km}(g)})}{\partial x^{ij}(e)} \frac{\partial}{\partial x^{km}(e)} = V^{ij} x^{kl}(g) \delta^l_i \delta^m_j \frac{\partial}{\partial x^{km}(g)}$$

$$V^{ij} x^{ki}(g) \frac{\partial}{\partial x^{kj}(g)} = \underbrace{x^{ki}(g) V^{ij}}_{(gV)^{kj}} \frac{\partial}{\partial x^{kj}(g)} = (gV)^{kj} \frac{\partial}{\partial x^{kj}(g)},$$

where V^{ij} is an arbitrary $n \times n$ real matrix.

Since $\mathcal{G}$ is a collection of vector fields, we can compute their commutators. The result is again left-invariant!

$$L_{a\star} \left.[X, Y]\right|_g = [L_{a\star} X|_g, L_{a\star} Y|_g] \stackrel{\text{l. inv.}}{=} [X|_{ag}, Y|_{ag}] \equiv [X, Y]|_{ag}.$$

So if $X, Y \in \mathcal{G}$, also $[X, Y] \in \mathcal{G}$.

Definition: The set of left-invariant vector fields $\mathcal{G}$ with the commutator (Lie bracket) $[\ ,\ ] : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ is called the **Lie algebra** of a Lie group G.

Examples:

1. $\mathrm{gl}(n, \mathbb{R}) = n \times n$ real matrices (Lie algebras are written with lower case letters).

2. $\mathrm{sl}(n, \mathbb{R})$: Take a curve $c(t)$ that passes through $e \in \mathrm{SL}(n, \mathbb{R})$ and compute its tangent vector ($c(0) = e = 1_n$). For small t: $c(t) = 1_n + tA$, $\left.\frac{dc}{dt}\right|_{t=0} = A \in T_e \mathrm{SL}(n, \mathbb{R})$. Now $\det c(t) = \det\,(1_n + tA) = 1 + t \operatorname{tr} A + \ldots = 1$. Thus $\operatorname{tr} A = 0$ and $\mathrm{sl}(n, \mathbb{R}) = \{A |\ A \text{ is a } n \times n \text{ real matrix, } \operatorname{tr} A = 0\}$.

3. $\mathrm{so}(n)$: $c(t) = 1_n + tA$. We need $c(t)$ to be orthogonal: $c(t)c(t)^T = (1 + tA)(1 + tA^T) = 1 + t(A + A^T) + O(t^2) = 1$. Thus we need to have $A = -A^T$ and so $\mathrm{so}(n) = \{A |\ A \text{ is an antisymmetric } n \times n \text{ matrix }\}$.

For complex matrices, the coordinates are taken to be the real and imaginary parts of the matrix

4. $\mathrm{u}(n) : c(t) = 1_n + tA$. Thus $c(t)c(t)^\dagger = (1+tA)(1+tA^\dagger) = 1+t(A+A^\dagger)+O(t^2) = 1$. So $A = -A^\dagger$ and $\mathrm{u}(n) = \{A|A$ is an antihermitean $n \times n$ complex matrix $\}$.

Note: In physics, we usually use the convention $c(t) = 1 + itA \Rightarrow A^\dagger = A$
$\Rightarrow \mathrm{u}(n) = \{$Hermitean $n \times n$ matrices $\}$.

5. $\mathrm{su}(n) = \{n \times n$ antihermitean traceless matrices $\}$.

9.3.15 Structure Constants of the Lie Algebra

Let $\{V_1, \ldots, V_n\}$ be a basis of T_eG (assume dim $G = n < \infty$). Then $X_\mu|_g = L_{g\star}V_\mu, \quad \mu = 1, \ldots, n$ is a basis of T_gG (usually it is not a coordinate basis). Since the vectors $\{V_1, \ldots, V_n\}$ are linearly independent, $\{X_1|_g, \ldots, X_n|_g\}$ are also linearly independent. ($L_{g\star}$ is an isomorphism between T_eG and T_gG; $(L_{g\star})^{-1} = L_{g^{-1}\star}$). Since V_μ are basis vectors of T_eG, we can expand

$$[V_\mu, V_\nu] = c_{\mu\nu}{}^\lambda V_\lambda.$$

Let's then push this to T_gG:

$$\begin{aligned} &L_{g\star}[V_\mu, V_\nu] = [L_{g\star}V_\mu, L_{g\star}V_\nu] = [X_\mu|_g, X_\nu|_g] \\ &L_{g\star}(c_{\mu\nu}{}^\lambda V_\lambda) = c_{\mu\nu}{}^\lambda X_\lambda|_g \\ &\qquad \Rightarrow [X_\mu|_g, X_\nu|_g] = c_{\mu\nu}{}^\lambda X_\lambda|_g. \end{aligned}$$

Letting g vary over all G, we get the same equation everywhere on G with the same numbers $c_{\mu\nu}{}^\lambda$. Thus we can write

$$[X_\mu, X_\nu] = c_{\mu\nu}{}^\lambda X_\lambda.$$

The $c_{\mu\nu}{}^\lambda$ are called the **structure constants** of the Lie algebra. Evidently we have $c_{\mu\nu}{}^\lambda = -c_{\nu\mu}{}^\lambda$. We also have the Jacobi identity (of commutators)

$$c_{\mu\nu}{}^\tau c_{\tau\rho}{}^\sigma + c_{\nu\rho}{}^\tau c_{\tau\mu}{}^\sigma + c_{\rho\mu}{}^\tau c_{\tau\nu}{}^\sigma = 0.$$

9.3.16 The adjoint representation of G

Let b be some element of G, $b \in G$. Let us define the map

$$ad_b : G \to G, \quad ad_b(g) \equiv ad_b g = bgb^{-1}.$$

This is a homomorphism: $ad_b g_1 \cdot ad_b g_2 = ad_b(g_1 g_2)$, and at the same time defines an action of G on itself (conjugation): $ad_b \cdot ad_c = ad_{bc}, \quad ad_e = id_G$. (Note that this is really a combined map: $ad_b \cdot ad_c \equiv ad_b \circ ad_c$). The differential map $ad_{b\star}$ pushes vectors from T_gG to $T_{ad_b g}G$. If $g = e$, $ad_b e = beb^{-1} = e$, so $ad_{b\star}$ maps T_eG to itself. Lets denote this map by Ad_b:

$$Ad_b : T_eG \to T_eG, \quad Ad_b = ad_{b\star}|_{T_eG}$$

One can easily show that $(f \circ g)_\star = f_\star \circ g_\star$, thus $ad_{b\star} ad_{c\star} = ad_{bc\star}$. It then follows that Ad_b is a representation of G in the vector space $\mathcal{G} \cong T_eG$, the so-called **adjoint representation**:

$$Ad : G \to \mathrm{Aut}(\mathcal{G}), \quad b \mapsto Ad_b.$$

If G is a matrix group (O, SO,...), then $V \in T_eG \cong \mathcal{G}$ is a matrix and

$$Ad_g V = gVg^{-1}.$$

(This follows from $ad_g(e + tV) = e + tgVg^{-1}$.) So, if $\{V_\mu\}$ is a basis of $\mathcal{G}$,

$$gV_\mu g^{-1} = V_\nu D^{(\mathrm{adj})\nu}{}_\mu(g).$$

9.4 de Rham cohomology

Recall the definition of the de Rham cohomology groups: First, $d : \Omega^k(M) \to \Omega^{k+1}(M)$ is a linear map with $d^2 = 0$. We set $B^k(M) = d(\Omega^{k-1}(M)) \subset \Omega^k(M)$ and $Z^k(M) = \ker d = \{\omega \in \Omega^k | d\omega = 0\} \subset \Omega^k(M)$. These are linear subspaces with the property $B^k(M) \subset Z^k(M)$, because of $d^2 = 0$. Elements of Z^k are called *closed forms* and elements of B^k are *exact forms.* We set

$$H^k(M) = Z^k(M)/B^k(M), \text{ with } k = 0, 1, 2, \ldots$$

where $H^0(M) \equiv Z^0(M)$. Note that $H^k(M) = 0$ for $k > n$ since $\Omega^k(M) = 0$ for $k > n$. The vector spaces $H^k(M)$ are called the *de Rham cohomology groups* of M. In case when M is compact, one can prove that $\dim H^k(M) < \infty$ for all k.

Example $M = \mathbb{R}^3$. Since $df = 0$ for $f \in C^\infty(M) = \Omega^0(M)$ means that f is a constant function, we get $H^0(\mathbb{R}^3) = \mathbb{R}$. If $\omega = \omega_i dx^i$ satisfies $d\omega = 0$ then the vector field $(\omega_1, \omega_2, \omega_3)$ has zero curl, and we know vector analysis that there is a scalar potential f such that $\nabla f = \omega$, in other words, $df = \omega$. Thus $B^1 = Z^1$ and so $H^1(\mathbb{R}) = 0$. If $\omega = \frac{1}{2}\omega_{ij} dx^i \wedge dx^j$ is a 2-form with $d\omega = 0$ then $\mathrm{div}\, \omega = 0$ with $\omega = (\omega_{23}, \omega_{31}, \omega_{12})$. This implies that there is a vector potential $\mathbf{A}$ such that $\nabla \times \mathbf{A} = \omega$, or in other words, $dA = \omega, A = A_i dx^i$. Again, $Z^2 = B^2$ and $H^2(\mathbb{R}^3) = 0$. In the same vein one can show that $H^3(\mathbb{R}^3) = 0$.

Theorem 9.2 *(Poincare's lemma) Let $M \subset \mathbb{R}^n$ be a star shaped open set. This means that there is a point $z \in M$ such that the line $tx + (1-t)z, 0 \le t \le 1$, belongs to M for any $x \in M$. Let ω be a closed $k-$form on M, $k > 0$. Then there exists a $(k-1)$-form θ such that $d\theta = \omega$.*

Proof. Write

$$\omega = \frac{1}{k!}\omega_{i_1\cdots i_k}(x)dx^{i_1}\wedge\cdot\wedge dx^{i_k}$$

and define

$$\theta_{i_1\dots i_{k-1}}(x) = \frac{1}{(k-1)!}\int_0^1 t^{k-1}(x^j - z^j)\omega_{ji_1i_2\dots i_{k-1}}(tx+(1-t)z)dt.$$

We claim that $d\theta = \omega$. Now

$$\begin{aligned} d\theta &= \frac{1}{(k-1)!}\int_0^1 t^{k-1}\omega_{ji_1\dots i_{k-1}}(tx+(1-t)z)dx^j\wedge dx^{i_1}\wedge\dots dx^{i_{k-1}}dt &(11)\\ &+ \frac{1}{(k-1)!}\int t^k(x^j-z^j)\partial_l\omega_{ji_1\dots i_{k-1}}(tx+(1-t)z)dx^l\wedge dx^{i_1}\wedge\dots dx^{i_{k-1}}dt &(12)\end{aligned}$$

The equation $d\omega = 0$ gives

$$\partial_l\omega_{ji_1\dots i_{k-1}} \pm \text{ cyclic permutations of } lji_1\dots i_{k-1} = 0,$$

where the signs are given by the parity of the cyclic permutation. From this equation one can reduce, by setting the contraction $i_{\partial_j}d\omega$ equal to zero,

$$k\cdot\partial_l\omega_{ji_1\dots i_{k-1}}dx^l\wedge dx^{i_1}\wedge\cdots\wedge dx_{i^{k-1}} = \partial_j\omega_{li_1\dots i_{k-1}}dx^l\wedge\cdots\wedge dx^{i_{k-1}}.$$

Note that in local coordinates

$$i_{\partial_j}d\omega_{\dots} + di_{\partial^j}\omega_{\dots} = \mathcal{L}_{\partial_j}\omega_{\dots} = \partial_j\omega_{\dots}.$$

Inserting this to the second term I_2 in (12) we obtain

$$\begin{aligned} I_2 &= \frac{1}{k!}\int_0^1(x^j-z^j)t^k\partial_j\omega_{li_1\dots i_{k-1}}(tx+(1-t)z)dx^l\wedge dx^{i_1}\cdots\wedge dx^{i_{k-1}}dt\\ &= \frac{1}{k!}\int_0^1 t^k\frac{d}{dt}\omega_{li_1\dots i_{k-1}}(tx+(1-t)z)dx^l\wedge dx^{i_1}\dots dx^{i_{k-1}}dt\\ &= -\frac{1}{(k-1)!}\int_0^1 t^{k-1}\omega_{li_1\dots i_{k-1}}(tx+(1-t)z)dx^l\wedge dx^{i_1}\wedge\dots dx^{i_{k-1}}dt\\ &+ \frac{1}{k!}\omega_{li_1\dots i_{k-1}}dx^l\wedge dx^{i_1}\cdots\wedge dx^{i_{k-1}}.\end{aligned}$$

Insertion to (12) completes the proof of $d\theta = \omega$.

The above result extends (by a use of coordinates) to the case when M is a contractible subset of a smooth manifold: contractibility means that the identity map on M can be smoothly deformed to a constant map $x \mapsto X_0$ on M. Let $f_t : M \to M$ be such a contraction, $f_0(x) = x_0$ and $f_1(x) = x$, $0 \le t \le 1$. Then one can repeat the proof but with the straight lines $t \mapsto tx + (1-t)z$ replaced by $t \mapsto f_t(x), z = x_0$, see Nakahara, section 6.3, for details.

Example 1 Let $M = S^1$. The 1-form $d\phi$ is closed but $d\phi \neq df$ for any smooth function f on S^1. Note that the polar angle ϕ is not a function on S^1 since it is nonperiodic. Any 1-form on S^1 is given as $f(\phi)d\phi$ for some periodic function f of ϕ. The integral of f over the interval $[0, 2\pi]$ gives a real number λ_f. If $\lambda_f = \lambda_g$ for any two functions f, g then we can write $f - g = h'$ for a periodic function h, that is, $fd\phi - gd\phi = dh$. It follows that the cohomology classes $[f] \in H^1(S^1)$ are parametrized by the integral λ_f and so $H^1(S^1) = \mathbb{R}$.

Example 2 On the unit sphere S^2 the area form is given as $\omega = \sin\theta d\theta \wedge d\phi$ in spherical coordinates. Locally, $\omega = d(-\cos\theta d\phi) = d(-\phi \sin\theta d\theta)$. Note that the first expression becomes singular at the poles $\theta = 0, \pi$ whereas the second is nonperiodic in the coordinate ϕ. One can prove that $H^2(S^2) = \mathbb{R}$ and that the cohomology classes are parametrized by the integral of the 2-form over S^2. In general, it is known that $H^k(S^n) = 0$ for $1 \leq k \leq n-1$ and that $H^0(S^n) = \mathbb{R} = H^n(S^n)$.

Example 3 $H^1(S^1 \times S^1) = \mathbb{R}^2$ (basis of 1-forms $d\phi_1, d\phi_2$) and $H^2(S^1 \times S^1) = \mathbb{R}$, basis $d\phi_1 \wedge d\phi_2$.

9.5 Integration of differential forms

We define an *orientation* on a manifold M of dimension n. The manifold is oriented if we have a complete system of local coordinates such that all coordinate transformations $x'^i = x'^i(x^1, \ldots, x^n)$ satisfy the condition $\det(\frac{\partial x'^i}{\partial x^j}) > 0$.

Not every manifold can be oriented. The standard spheres S^n inherit an orientation from $\mathbb{R}^{n+1}$. The orientation on $\mathbb{R}^n$ is given by the ordered set of Cartesian coordinates $(x^1, x^2, \ldots, x^n)$. A coordinate system $(y^1, \ldots, y^n)$ on the embedded unit sphere in $\mathbb{R}^{n+1}$ is then oriented if the vectors $(v, \partial_1, \ldots, \partial_n)$ (in y coordinates) are compatible with the orientation of $\mathbb{R}^{n+1}$. Here v is the outward unit normal vector field on the sphere and compatibility means that the matrix relating the given tangent vectors to the standard basis has positive determinant. On the other hand, *the real projective plane* $P\mathbb{R}^2 = S^2/\mathbb{Z}_2 = (\mathbb{R}^3 - \{0\})/\mathbb{R}_+$, consisting of lines through the origin in $\mathbb{R}^3$, has no orientation.

Let M be a smooth oriented manifold of dimension n. We fix an atlas of coordinate neighborhoods compatible with the given orientation. Let $x^1, \ldots, x^n$ be local coordinates on an open set $U \subset M$. Asssume that $f \in C^\infty(M)$ is such that $f(x) = 0$ when x is outside of a compact subset K of U. Then $\omega = f(x)dx^1 \wedge dx^2 \cdots \wedge dx^n$ is a $n-$ form on M. We **define** the integral

$$\int \omega = \int f(x) dx^1 dx_2 \ldots dx^n,$$

as the ordinary Riemann integral in $\mathbb{R}^n$.

Let us assume that we have a *locally finite* atlas (U_α, ϕ_α). This means that for any $x \in M$ there is an open neighborhood V of x such that V intersects only a finite number of the sets U_α. A space which has a locally finite cover is said to be *paracompact.* In fact, finite-dimensional manifolds are normally defined to be paracompact. A locally finite atlas has a subordinate *partition of unity.* That is, there is a family of smooth nonnegative functions $\rho_\alpha : M \to \mathbb{R}$ such that

- $\mathrm{supp}\rho_\alpha \subset U_\alpha$
- $\sum_\alpha \rho_\alpha(x) = 1$ for all $x \in M$.

The support $\mathrm{supp}f$ of a function f is defined as a closure of the set of points x for which $f(x) \neq 0$.

Let $\omega \in \Omega^n(M)$. we define

$$\int_M \omega = \sum_\alpha \int \rho_\alpha \omega,$$

and we apply the previous definition to each term on the right-hand-side. The integral converges always when M is compact.

Exercise Show that the above definition does not depend on the choice of the partition of unity or of the locally finite atlas.

Example Let $M = S^1$, $U_1 = S^1 - \{(1,0)\}$, $U_2 = S^1 - \{(-1,0)\}$. Choose the (inverse) coordinate functions as

$$\begin{aligned} \varphi_1^{-1} &: (0, 2\pi) \to U_1, \quad \theta_1 \mapsto (\cos\theta_1, \sin\theta_1) \\ \varphi_2^{-1} &: (-\pi, \pi) \to U_2, \quad \theta_2 \mapsto (\cos\theta_2, \sin\theta_2) \end{aligned}$$

Partition of unity: $\rho_1(\theta_1) = \sin^2\frac{\theta_1}{2}$, $\rho_2(\theta_2) = \cos^2\frac{\theta_2}{2}$. (Note that this satisfies (i) - (iii)). Choose $f : S^1 \to \mathbb{R}$ as $f(\theta) = \sin^2\theta$ and $\omega = f \cdot d\theta_1$ on U_1 and $\omega = f \cdot d(\theta_2 + 2\pi) = 1 \cdot d\theta_2$ on U_2. Now

$$\int_{S^1} \omega = \sum_{i=1}^{2} \int_{U_i} \rho_i \omega = \int_0^{2\pi} d\theta_1 \sin^2\frac{\theta_1}{2} \sin^2\theta_1 + \int_{-\pi}^{\pi} d\theta_2 \cos^2\frac{\theta_2}{2} \sin^2\theta_2 = \frac{\pi}{2} + \frac{\pi}{2} = \pi,$$

as expected.

Next we want to define the integral of a form $\omega \in \Omega^k(M)$ over a parametrized $k-$ surface for arbitrary $0 \leq k \leq n$.

A *standard k-simplex* in $\mathbb{R}^k$ is the subset

$$\sigma_k = \{(x^1, \ldots, x^k) \in \mathbb{R}^k | \sum x^i \leq 1, x^j \geq 0\}.$$

So σ_0 is just a point, σ_1 is the unit interval, σ_2 is a triangle, etc.

A singular k-simplex is any smooth map $s_k : \sigma_k \to M$. A *k-chain* is a formal linear combination $\sum a_\alpha s_{k,\alpha}$, with $a_\alpha \in \mathbb{R}$ and each $s_{k,\alpha}$ is a singular $k-$ simplex.

We define an affine map $F_k^i : \sigma_{k-1} \to \sigma_k$ where $i = 0, 1, \ldots, k$. Note that the subset of points in σ_k with the coordinate $x^i = 0$ can be naturally identified as a $k-1$ simplex

σ_{k-1} for $1 \leq i \leq k$. This defines the map (as an identity map) for $i = 1, 2 \ldots, k$. The remaining map F_k^0 sends the $(k-1)$-simplex σ_{k-1} to the face of the k-simplex which is not parallel to any of the coordinate axes. The map is completely fixed by requiring it to be affine and *compatible with the orientations,* and such that the origin of σ_{k-1} is mapped to the vertex of σ_k lying on the first coordinate axes, and the vertex of σ_{k-1} lying on the i :th coordinate axes is mapped to the vertex of σ_k on the $(i+1)$:th coordinate axes, for $i = 1, 2, \ldots, k-1$. So

$$
\begin{aligned}
F_k^0(x^1, \ldots, x^{k-1}) &= (1 - \sum_{i=1}^{k-1} x^i, x^1, \ldots, x^{k-1}) \\
F_k^i(x^1, \ldots, x^{k-1}) &= (x^1, \ldots, x^{i-1}, 0, x^i, \ldots, x^{k-1}) \text{ with } i = 1, 2, \ldots, k.
\end{aligned}
$$

The *boundary* of a singular k-simplex $s_k : \sigma_k \to M$ is the singular k-chain defined as

$$\partial s_k = \sum_{i=0}^{k} (-1)^i s_k \circ F_k^i.$$

we extend the definition, by linearity, to the space C_k of singular k-chains, $\partial : C_k \to C_{k-1}$.

Theorem 9.3 $\partial^2 = 0$.

Proof. We first observe that

$$F_k^i \circ F_{k-1}^j = F_k^j \circ F_{k-1}^{i-1}, \text{ for } j < i.$$

Let $s = \sum_\alpha a_\alpha s_{k,\alpha} \in C_k$. Then

$$
\begin{aligned}
\partial^2 s &= \partial \sum_\alpha a_\alpha \sum_{i=0}^{k} (-1)^i s_{k,\alpha} \circ F_k^i \\
&= \sum_\alpha a_\alpha \sum_{i=0}^{k} (-1)^i \sum_{j=0}^{k} s_{k,\alpha} \circ F_k^i \circ F_{k-1}^j (-1)^j \\
&= \sum_\alpha a_\alpha \left(\sum_{0 \leq i \leq j \leq k-1} (-1)^{i+j} s_{k,\alpha} F_k^i \circ F_{k-1}^j \right. \\
&\quad \left. + \sum_{0 \leq j < i \leq k} (-1)^{i+j} s_{k,\alpha} F_k^i \circ F_{k-1}^j \right) \\
&= \sum_\alpha a_\alpha \left(\sum_{0 \leq i \leq j \leq k-1} (-1)^{i+j} s_{k,\alpha} F_k^i \circ F_{k-1}^j \right. \\
&\quad \left. + \sum_{0 \leq j < i \leq k} (-1)^{i+j} s_{k,\alpha} F_k^j \circ F_{k-1}^{i-1} \right).
\end{aligned}
$$

Relabel $i \mapsto j, j \mapsto i-1$ in the first term of right-hand-side of the last equality; then the terms cancel.

A *cycle* is a singular chain s such that $\partial s = 0$. A *boundary* is a singular chain b such that $b = \partial s$ for some singular chain s. Denote by Z_k the space of k-cycles and by B_k the space of k-boundaries. Finally, the *singular k-homology group* is the space

$$H_k(M) = H_k(M, \mathbb{R}) = Z_k(M)/B_k(M).$$

Sometimes one considers also the homology group $H_k(M, \mathbb{Z})$ which is defined as the real homology group but one restricts to integral linear combinations of the singular k-simplexes.

Exercise Show that $H_0(M)$ is isomorphic with $\mathbb{R}^k$, where k is the number of path connected components of M.

The homology groups H_k of contractible manifolds vanish for $k > 0$, so in particular $H_k(\mathbb{R}^n) = 0$ for $k > 0$. On the other hand, $H_n(S^n) = \mathbb{R}$ but $H_k(S^n) = 0$ for $0 < k < n$.

We define the integral of a k-form over a singular k-chain $s = \sum_\alpha a_\alpha s_{k,\alpha}$,

$$\int_s \omega = \sum_\alpha a_\alpha \int_{\sigma_k} s^*_{k,\alpha}\omega.$$

Each of the integral on the right is an ordinary Riemann integral of a smooth function defined in the standard simplex $\sigma_k \subset \mathbb{R}^k$, after writing each of the pull-back forms as $f(x)dx_1 \wedge \dots dx_k$.

Theorem 9.4 *(Stokes' theorem)*

$$\int_s d\omega = \int_{\partial s} \omega$$

for any $\omega \in \Omega^{k-1}(M)$ and for any singular k-chain s.

Proof. By linearity, it is sufficient to give the proof for a single singular k-simplex s_k. But in this case a typical term in $s_k^*\omega$ can be written as

$$s_k^*\omega = \sum_{j=1}^k b^j(x)dx^1 \wedge \dots \hat{dx^j} \wedge \dots dx^k(-1)^{j-1}$$

for some smooth functions b^j. Then

$$d(s_k^*\omega) = s_k^*(d\omega) = \sum(\partial_j b^j)dx^1 \wedge \dots \wedge dx^k = f(x)dx^1 \wedge \dots \wedge dx^k.$$

We can now apply the familiar Gauss' theorem for vector fields in $\mathbb{R}^k$,

$$\int_{\sigma_k} \partial_j b^j dx^1 \dots dx^k = \int_{\partial\sigma_k} \mathbf{b} \cdot \mathbf{n} dS,$$

where $\mathbf{n}$ is the outward normal vector field on σ_k and dS is the Euclidean area measure on the surface $\partial\sigma_k$ of the k-simplex. But the right-hand-side of the equation is equal to the integral $\int_{\partial\sigma_k} s_k^*\omega$, which proves the theorem.

We have a pairing $H_k(M) \times H^k(M) \to \mathbb{R}$ which is given as

$$< [s], [\omega] >= \int_s \omega.$$

Because of Stokes' theorem the right-hand-side does not depend on particular representatives of the (co)homology classes, i.e., if $s - s'$ is a boundary and $\omega - \omega'$ is a coboundary then

$$\int_s \omega = \int_{s'} \omega'.$$

For compact oriented manifolds one can prove that the pairing is nondegenerate, i.e., if $< [s], [\omega] >= 0$ for all $[\omega]$ (resp. for all $[s]$) then $[s] = 0$ (resp. $[\omega] = 0$).

There is a more refined version of Stokes' theorem (which we are not going to prove). This uses the idea of a *closed submanifold with boundary.* A manifold M with boundary is defined using the half space $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n | x^n \geq 0\}$ as a model instead of the vector space $\mathbb{R}^n$. That is, M should be equipped with a cover by open sets U and coordinate maps $\phi : U \to \mathbb{R}^n_+$ which are homeomorphism to open subsets of the half space. The coordinate transformations $\phi \circ \psi^{-1}$ are again required to be smooth in their domain of definition. Note that the derivative in the x^n direction at the boundary points $x^n = 0$ is only defined to the positive direction.

Example The closed unit ball $B^n = \{x \in \mathbb{R}^n | ||x|| \leq 1\}$ is a manifold with boundary. The set of boundary points is the manifold S^{n-1}.

Let $N \subset M$ be an oriented manifold with boundary (dimension n) embedded in M. Its boundary ∂N is a manifold of dimension $n-1$. Let $\omega \in \Omega^{n-1}(M)$. Then one can prove

$$\int_N d\omega = \int_{\partial N} \omega.$$

Note that the integral on the left is an integral of a n-form over a manifold of dimension n (and this we have already defined) and on the right we have an integral of a $(n-1)$-form over a manifold of dimension $n-1$.

Additional reading: Nakahara: 5.4, 5.5, and Chapter 6
Chern, Chen, and Lam: Chapters 2 and 3

10 Riemannian Geometry (Metric Manifolds)

(Chapter 7 of Nakahara's book)

10.1 The Metric Tensor

Let M be a differentiable manifold. The Riemannian metric on M is a $(0,2)$-tensorfield, which satisfies

(i) $g_p(U,V) = g_p(V,U) \qquad \forall p \in M, \;\; U,V \in T_pM$ (i.e. g is symmetric)

(ii) $g_p(U,U) \geq 0,$ and $g_p(U,U) = 0 \leftrightarrow U = 0$ (g is positive definite).

If instead of (ii) g satisfies

(ii') If $g_p(U,V) = 0$ for all $U \in T_pM$, then $V = 0$,

we say that g is a pseudo-Riemannian metric (symmetric and non-degenerate).
(M,g) with a (pseudo-) Riemannian metric is called a (pseudo-) Riemannian manifold. The spacetime in general relativity is an example of a pseudo-Rimannian manifold. In local coordinates $g = g_{\mu\nu} dx^\mu \otimes dx^\nu$. (The Euclidean metric: $g_{\mu\nu} = \delta_{\mu\nu}$. Then $g(U,V) = \sum_{i=1}^n U^i V^i$.)

10.2 The Induced Metric

Let (N, g_N) be a Riemannian manifold, dim $N = n$. We define an m dimensional **submanifold** M of N:
Let $f : M \to N$ be a smooth map such that f is an injection and the push $f_\star : T_pM \to T_{f(p)}N$ is also an injection. Then f is an **embedding** of M in N and the image $f(M)$ is a **submanifold** of N. However, it follows that M and $f(M)$ are diffeomorphic, so we can call M a submanifold of N.

Now the pullback f^* of f induces the natural metric g_M on M:

$$g_M = f^* g_N.$$

The components of g_M are given by

$$g_{M\mu\nu}(x) = g_{N\alpha\beta}(f(x)) \frac{\partial f^\alpha}{\partial x^\mu} \frac{\partial f^\beta}{\partial x^\nu}.$$

[By the chain rule: $g_{M\mu\nu} dx^\mu \otimes dx^\nu = g_{N\alpha\beta} \frac{\partial f^\alpha}{\partial x^\mu} \frac{\partial f^\beta}{\partial x^\nu} dx^\mu \otimes dx^\nu$]

Example: Let (θ, φ) be the polar coordinates on S^2 and $f : S^2 \to \mathbb{R}^3$ the usual embedding: $f(\theta,\varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$. On $\mathbb{R}^3$ we have the Euclidean metric $\delta_{\mu\nu}$. We denote $y^1 = \theta, y^2 = \varphi$. We obtain the induced metric on S^2:

$$g_{\mu\nu} dy^\mu \otimes dy^\nu = \delta_{\alpha\beta} \frac{\partial f^\alpha}{\partial y^\mu} \frac{\partial f^\beta}{\partial y^\nu} dy^\mu \otimes dy^\nu = d\theta \otimes d\theta + \sin^2\theta d\varphi \otimes d\varphi.$$

Thus the components of the metric are $g_{11}(\theta,\varphi) = 1, \;\; g_{22}(\theta,\varphi) = \sin^2\theta, \;\; g_{12}(\theta,\varphi) = g_{21}(\theta,\varphi) = 0.$

Why the notation ds^2 is often used for the metric?
Often the metric is denoted $ds^2 = g_{\mu\nu} dx^\mu \otimes dx^\nu$. The reason for this is as follows. Let $c(t)$ be a curve on manifold M with the metric g. The tangent vector of the curve is $\dot{c}(t)$, which in local coordinates is $\dot{c}(t) = (\frac{dx^\mu(t)}{dt})$. $[c(t) = (x^\mu(t))]$
If $M = \mathbb{R}^3$ with the Euclidean metric $g_{\mu\nu} = \delta_{\mu\nu}$, the length of the curve between t_0 and t_1 would be

$$L_{\mathbb{R}^3} = \int_{t_0}^{t_1} dt \sqrt{(\dot{x}^1)^2 + (\dot{x}^2)^2 + (\dot{x}^3)^2} = \int_{t_0}^{t_1} dt \sqrt{\delta_{\mu\nu} \dot{x}^\mu \dot{x}^\nu}.$$

In general case the length of the part of the curve between t_0 and t_1 is then

$$length L = \int_{t_0}^{t_1} dt \sqrt{g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu}. \tag{13}$$

If t_0 and t_1 are infinitesimally close : $t_1 = t_0 + \Delta t$, then

$$\Delta s \equiv L \simeq \Delta t \sqrt{g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} \simeq \Delta t \sqrt{g_{\mu\nu} \frac{\Delta x^\mu}{\Delta t} \frac{\Delta x^\nu}{\Delta t}} = \sqrt{g_{\mu\nu} \Delta x^\mu \Delta x^\nu}.$$

Thus $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ is the square of an "infinitesimal length element" ds. We will have more to say about the formula for the length later.

10.3 An application to Maxwell's equations

We arrange the Cartesian coordinates of the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ as an antisymmetric 4×4 matrix,

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -cB_z & +cB_y \\ E_y & cB_z & 0 & -cB_x \\ E_z & -cB_y & cB_x & 0 \end{pmatrix}.$$

We label the rows and columns by $\mu, \nu = 0, 1, 2, 3$ and we set $F = \frac{1}{2} F^{\mu\nu} dx_\mu \wedge dx_\nu$.

Let ϕ be an electric scalar potential and $\mathbf{A}$ a magnetic vector potential. Then

$$\mathbf{E} = -\nabla\phi - \partial^0 \mathbf{A} \text{ and } \mathbf{B} = \nabla \times \mathbf{A},$$

where $\partial_0 = \frac{1}{c}\frac{\partial}{\partial t}$ but we shall work in units with speed of light $c = 1$. Define the 1-form $A = A_\mu dx^\mu$ with $A_0 = \phi$ and $A_i = c\mathbf{A}_i$. Thus we may write

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$

that is, $F = dA$.

Since $d^2 = 0$ we have automatically $dF = 0$. Written in electric and magnetic field components this gives the second set of Maxwell's equations,

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}. \end{aligned}$$

In order to obtain a differential form expression for the first set of Maxwell's equations,

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \rho/\epsilon_0 \\ \nabla \times \mathbf{B} &= \mu_0 \left(\epsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j} \right), \end{aligned}$$

with $\mu_0 \epsilon_0 = 1/c^2$, we must first fix a metric tensor $(g_{\mu\nu})$ in space-time; this could be just the Minkowski metric $\mathrm{diag}(1, -1, -1, -1)$ but we may take any (pseudo) Riemannian metric. Note that the second set of Maxwell's equations is intrinsic to any smooth manifold, it does not depend on the choice of metric.

We shall denote $g_{ij} = g(\partial_i, \partial_j)$ for a (pseudo) Riemannian metric $g_x : T_xM \times T_xM \to \mathbb{R}$. A metric tensor is **pseudo-Riemannian** if it has the properties of a metric tensor except that we do not require that it is positive definite; for example, in relativity theory (in four space-time dimensions) the metric tensor has one positive eigenvalue and three negative eigenvalues. We assume that the manifold is oriented.

A metric defines also **a duality operation** $* : \Omega^k(M) \to \Omega^{n-k}(M)$ on differential forms. In local coordinates,

$$\begin{aligned} *\omega &= \theta_{i_1 i_2 \dots i_{n-k}} dx^{i_1} \wedge dx^{i_2} \dots dx^{i_{n-k}} \text{ with} \\ \theta_{i_1 \dots i_{n-k}} &= |\det(g_{ij})|^{1/2} \frac{1}{k!} \epsilon_{i_1 \dots i_{n-k}}{}^{j_1 \dots j_k} \, \omega_{j_1 \dots j_k}, \end{aligned}$$

where $\epsilon_{i_1 \dots i_n}$ is the totally antisymmetric tensor with $\epsilon_{12\dots n} = +1$ and the raising of indices is done with the help of the metric tensor as in general relativity, i.e., $A^{\alpha_1 \dots \alpha_k} = g^{\alpha_1 \beta_1} \cdots g^{\alpha_k \beta_k} A_{\beta_1 \dots \beta_k}$ where (g^{ij}) is the inverse of the matrix (g_{ij}).

Example Let $M = \mathbb{R}^4$ and g_{ij} the Minkowski metric. Then $vol_M = dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$. The dual of the Maxwell 2-form $F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu$ is given by

$$(*F)_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu}{}^{\alpha\beta} F_{\alpha\beta},$$

so $(*F)_{12} = -F_{03}$, and cyclic permutations of 123, and $(*F)_{01} = F_{23}$, and cyclic permutations of 123. That is, the magnetic components of the dual are equal to $(-1)\times$ the electric components of the original and the electric components of the dual are equal to the magnetic components of the original field.

The complete set of Maxwell's equations can now be written as

$$\begin{aligned} d * F &= J \\ dF &= 0, \end{aligned}$$

where the 3-form J is defined as $\frac{1}{3!} \epsilon_{\mu\alpha\beta\gamma} J^\mu dx^\alpha \wedge dx^\beta \wedge dx^\gamma$ with $J^0 = \rho/\epsilon_0$ and $J^k = c\mu_0 j^k$. Here ρ is the charge density and $\mathbf{j}$ is the electric current density.

10.4 Affine Connection

Recall that $\chi(M) = \{$ vector fields on $M\}$. An (affine) **connection** ∇ is a map $\chi(M) \times \chi(M) \to \chi(M)$, $(X, Y) \mapsto \nabla_X Y$ such that

1. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$ (linear in the 2^{nd} argument)
2. $\nabla_{(X+Y)} Z = \nabla_X Z + \nabla_Y Z$ (linear in the 1^{st} argument)
3. f is a function on M ($f \in \mathcal{F}(M)$) $\Rightarrow \nabla_{fX} Y = f \nabla_X Y$
4. $\nabla_X(fY) = X[f]Y + f\nabla_X Y$.

Now take a chart (U, φ) with coordinates $x = \varphi(p)$. Let $\{e_\nu = \frac{\partial}{\partial x^\nu}\}$ be the coordinate basis of T_pM. We define $(\dim M)^3$ **connection coefficients** ${\Gamma^\lambda}_{\mu\nu}$ by

$$\nabla_{e_\mu} e_\nu = {\Gamma^\lambda}_{\mu\nu} e_\lambda.$$

We can express the connection in the coordinate basis with the help of connection coefficients: Let $X = X^\mu e_\mu$ and $Y = Y^\nu e_\nu$ be two vector fields. Denote $\nabla_\mu \equiv \nabla_{e_\mu}$. Now

$$\begin{aligned}\nabla_X Y &\overset{2,3}{=} X^\mu \nabla_\mu (Y^\nu e_\nu) \overset{4}{=} X^\mu e_\mu[Y^\nu] e_\nu + X^\mu Y^\nu \nabla_\mu e_\nu = X^\mu \frac{\partial Y^\nu}{\partial x^\mu} e_\nu + X^\mu Y^\nu {\Gamma^\lambda}_{\mu\nu} e_\lambda \\ &= X^\mu \left(\frac{\partial Y^\lambda}{\partial x^\mu} + {\Gamma^\lambda}_{\mu\nu} Y^\nu\right) e_\lambda \equiv X^\mu (\nabla_\mu Y)^\lambda e_\lambda,\end{aligned}$$

where we have

$$(\nabla_\mu Y)^\lambda = \frac{\partial Y^\lambda}{\partial x^\mu} + {\Gamma^\lambda}_{\mu\nu} Y^\nu.$$

Note that $\nabla_X Y$ contains no derivatives of X unlike $\mathcal{L}_X Y$.

10.5 Parallel Transport and Geodesics

Let $c : (a, b) \to M$ be a curve on M with coordinate representation $x^\mu = x^\mu(t)$. Its tangent vector is

$$V = V^\mu e_\mu \big|_{c(t)} = \frac{dx^\mu(c(t))}{dt} e_\mu \Big|_{c(t)}.$$

If a vector field X satisfies

$$\nabla_V X = 0 \qquad \text{(along } c(t)),$$

then we say that X is **parallel transported** along the curve $c(t)$. In component form this is

$$\frac{dX^\mu}{dt} + {\Gamma^\mu}_{\nu\lambda} \frac{dx^\nu(t)}{dt} X^\lambda = 0.$$

If the tangent vector V itself is parallel transported along the curve $c(t)$,

$$geodesic1 \nabla_V V = 0, \tag{14}$$

then the curve $c(t)$ is called a **geodesic**. The above equation is the **geodesic equation** and in component form it is

$$\frac{d^2x^\mu}{dt^2} + \Gamma^\mu{}_{\nu\lambda}\frac{dx^\nu}{dt}\frac{dx^\lambda}{dt} = 0$$

Geodesics can be interpreted as the straightest possible curves in a Riemannian manifold. If $M = \mathbb{R}^n$ and $g = \delta$ (the Euclidean metric), then the geodesics are straight lines.

10.6 The Covariant Derivative of Tensor Fields

Connection was a term that we used for the map $\nabla : (X, Y) \mapsto \nabla_X Y$. The map $\nabla_X : \chi(M) \to \chi(M)$, $Y \mapsto \nabla_X Y$ is called the covariant derivative. It is a proper generalization of the directional derivative of functions to vector fields, and as we'll discuss next, to tensor fields.

For a function, we define $\nabla_X f$ to be the same as the directional derivative:

$$\nabla_X f = X[f].$$

Thus the condition number 4 in the definition of ∇ is the Leibnitz rule:

$$\nabla_X(fY) = (\nabla_X f)Y + f(\nabla_X Y).$$

Let's require that this should be true for any product of tensors:

$$\nabla_X(T_1 \otimes T_2) = (\nabla_X T_1) \otimes T_2 + T_1 \otimes (\nabla_X T_2),$$

where T_1 and T_2 are tensor fields of arbitrary types. The formula must also be true when some of the indices are contracted. Thus we can define the covariant derivative of a one-form as follows. Let $\omega \in \Omega^1(M)$ be a one-form ((0,1) tensor field), $Y \in \chi(M)$ be a vector field ((1,0) tensor field). Then $< \omega, Y > \in \mathcal{F}(M)$ is a smooth function on M. Recall that $< \omega, Y > \equiv \omega[Y] = \omega_\mu Y^\mu$. (Here μ is the contracted index.) Then

$$\nabla_X < \omega, Y > = X(\omega[Y]) = X^\mu \frac{\partial}{\partial x^\mu}(\omega_\nu Y^\nu) = X^\mu \frac{\partial \omega_\nu}{\partial x^\mu} Y^\nu + X^\mu \omega_\nu \frac{\partial Y^\nu}{\partial x^\mu}.$$

On the other hand because of the Leibnitz rule we must have

$$\begin{aligned}\nabla_X < \omega, Y > &= < \nabla_X \omega, Y > + < \omega, \nabla_X Y > = (\nabla_X \omega)_\nu Y^\nu + \omega_\nu (\nabla_X Y)^\nu \\ &= (\nabla_X \omega)_\nu Y^\nu + \omega_\nu X^\mu \frac{\partial Y^\nu}{\partial x^\mu} + \omega_\nu \Gamma^\nu{}_{\mu\alpha} X^\mu Y^\alpha\end{aligned}$$

¿From these two formulas we find $(\nabla_X \omega)_\nu$. (Note that the two $X^\mu \omega_\nu \frac{\partial Y^\nu}{\partial x^\mu}$ terms cancel.)

$$\Rightarrow (\nabla_X \omega)_\nu = X^\mu \left(\frac{\partial \omega_\nu}{\partial x^\mu} - \Gamma^\alpha{}_{\mu\nu} \omega_\alpha \right).$$

When $X = \frac{\partial}{\partial x^\mu}$, this reduces to

$$(\nabla_\mu \omega)_\nu = \frac{\partial \omega_\nu}{\partial x^\mu} - \Gamma^\alpha{}_{\mu\nu} \omega_\alpha.$$

Further when $\omega = dx^\sigma$: $\nabla_\mu dx^\sigma = -\Gamma^\sigma{}_{\mu\nu} dx^\nu$.
For a generic tensor, the result turns out to be

$$\begin{aligned}\nabla_\nu t^{\lambda_1 \ldots \lambda_p}_{\mu_1 \ldots \mu_q} = \partial_\nu t^{\lambda_1 \ldots \lambda_p}_{\mu_1 \ldots \mu_q} + \Gamma^{\lambda_1}{}_{\nu\rho} t^{\rho\lambda_2 \ldots \lambda_p}_{\mu_1 \ldots \mu_q} + \ldots + \Gamma^{\lambda_p}{}_{\nu\rho} t^{\lambda_1 \ldots \lambda_{p-1}\rho}_{\mu_1 \ldots \mu_q} \\ - \Gamma^\rho{}_{\nu\mu_1} t^{\lambda_1 \ldots \lambda_p}_{\rho\mu_2 \ldots \mu_q} - \ldots - \Gamma^\rho{}_{\nu\mu_q} t^{\lambda_1 \ldots \lambda_p}_{\mu_1 \ldots \mu_{q-1}\rho}.\end{aligned}$$

(Note that we should really have written $t^{\lambda_1 \ldots \lambda_p}{}_{\mu_1 \ldots \mu_q}$, but this was not done for typographical reasons.)

10.7 The Transformation Properties of Connection Coefficients

Let U and V be two overlapping charts with coordinates:

$$\begin{aligned} \text{on } U: \quad & x \quad e_\mu = \frac{\partial}{\partial x^\mu}, \\ \text{on } V: \quad & y \quad \tilde{e}_\nu = \frac{\partial}{\partial y^\nu} = \frac{\partial x^\mu}{\partial y^\nu} e_\mu. \end{aligned}$$

Let $p \in U \cap V \neq \emptyset$. The connection coefficients on V are

$$\nabla_{\tilde{e}_\alpha} \tilde{e}_\beta = \tilde{\Gamma}^\gamma{}_{\alpha\beta} \tilde{e}_\gamma = \tilde{\Gamma}^\gamma{}_{\alpha\beta} \frac{\partial x^\nu}{\partial y^\gamma} e_\nu$$

On the other hand

$$\nabla_{\tilde{e}_\alpha} \tilde{e}_\beta = \nabla_{\tilde{e}_\alpha} (\frac{\partial x^\mu}{\partial y^\beta} e_\mu) = \left(\frac{\partial^2 x^\nu}{\partial y^\alpha y^\beta} + \frac{\partial x^\lambda}{\partial y^\alpha} \frac{\partial x^\mu}{\partial y^\beta} \Gamma^\nu{}_{\lambda\mu} \right) e_\nu$$

Thus

$$\tilde{\Gamma}^\gamma{}_{\alpha\beta} \frac{\partial x^\nu}{\partial y^\gamma} = \left(\frac{\partial^2 x^\nu}{\partial y^\alpha y^\beta} + \frac{\partial x^\lambda}{\partial y^\alpha} \frac{\partial x^\mu}{\partial y^\beta} \Gamma^\nu{}_{\lambda\mu} \right).$$

¿From this we find the transformation rule for the connection coefficients:

$$\tilde{\Gamma}^\gamma{}_{\alpha\beta} = \frac{\partial y^\gamma}{\partial x^\nu} \frac{\partial x^\lambda}{\partial y^\alpha} \frac{\partial x^\mu}{\partial y^\beta} \Gamma^\nu{}_{\lambda\mu} + \frac{\partial^2 x^\nu}{\partial y^\alpha y^\beta} \frac{\partial y^\gamma}{\partial x^\nu}.$$

We notice that the first term is just the transformation rule for the components of a (1,2)-tensor. But we also have an additional second term, which is symmetric in α and β. Thus Γ is <u>almost</u> like a (1,2)-tensor, but not quite. To construct a (1,2)-tensor out of Γ, define

$$T^{\gamma}{}_{\alpha\beta} = \Gamma^{\gamma}{}_{\alpha\beta} - \Gamma^{\gamma}{}_{\beta\alpha} \equiv 2\Gamma^{\gamma}{}_{[\alpha\beta]} = \text{ the } \textbf{torsion tensor}$$

(note: $t_{[\alpha\beta]} = \frac{1}{2}(t_{\alpha\beta} - t_{\beta\alpha})$ is the antisymmetrization of indices.)

10.8 The Metric Connection

Let c be an arbitrary curve and V its tangent vector. If a connection ∇ satisfies[2]

$$\nabla_V(g(X,Y)) = 0 \qquad \text{when} \quad \nabla_V X = 0 \;\text{ and }\; \nabla_V Y = 0,$$

then we say that ∇ is a **metric connection.** Since

$$\nabla_\nu(g(X,Y)) = (\nabla_V g)(X,Y) + g(\overbrace{\nabla_V X}^{=0}, Y) + g(X, \overbrace{\nabla_V Y}^{=0}) = 0,$$

the metric connection satisfies

$$\nabla_V g = 0.$$

In component form:

1. $(\nabla_\mu g)_{\alpha\beta} = \partial_\mu g_{\alpha\beta} - \Gamma^{\lambda}{}_{\mu\alpha} g_{\lambda\beta} - \Gamma^{\lambda}{}_{\mu\beta} g_{\alpha\lambda} = 0.$
   And by cyclic permutation of μ, α and β we get:
2. $(\nabla_\alpha g)_{\beta\mu} = \partial_\alpha g_{\beta\mu} - \Gamma^{\lambda}{}_{\alpha\beta} g_{\lambda\mu} - \Gamma^{\lambda}{}_{\alpha\mu} g_{\beta\lambda} = 0$
3. $(\nabla_\beta g)_{\mu\alpha} = \partial_\beta g_{\mu\alpha} - \Gamma^{\lambda}{}_{\beta\mu} g_{\lambda\alpha} - \Gamma^{\lambda}{}_{\beta\alpha} g_{\mu\lambda} = 0$

Let us denote the symmetrization of indices: $\Gamma^{\gamma}{}_{(\alpha\beta)} \equiv \frac{1}{2}(\Gamma^{\gamma}{}_{\alpha\beta} + \Gamma^{\gamma}{}_{\beta\alpha})$. Then adding -(1)+(2)+(3) gives

$$-\partial_\mu g_{\alpha\beta} + \partial_\alpha g_{\beta\mu} + \partial_\beta g_{\mu\alpha} + T^{\lambda}{}_{\mu\alpha} g_{\lambda\beta} + T^{\lambda}{}_{\mu\beta} g_{\lambda\alpha} - 2\Gamma^{\lambda}{}_{(\alpha\beta)} g_{\lambda\mu} = 0$$

In other words

$$\Gamma^{\lambda}{}_{(\alpha\beta)} g_{\lambda\mu} = \frac{1}{2}\left\{(\partial_\alpha g_{\beta\mu} + \partial_\beta g_{\mu\alpha} - \partial_\mu g_{\alpha\beta}) + T^{\lambda}{}_{\mu\alpha} g_{\lambda\beta} + T^{\lambda}{}_{\mu\beta} g_{\lambda\alpha}\right\}$$

Thus

$$\Gamma^{\kappa}{}_{(\alpha\beta)} = \{\kappa\alpha\beta\} + \frac{1}{2}(T_{\alpha}{}^{\kappa}{}_{\beta} + T_{\beta}{}^{\kappa}{}_{\alpha}),$$

[2]This condition means that the angle between vectors is preserved under parallel transport.

where $\{\kappa\alpha\beta\} = \frac{1}{2}g^{\kappa\mu}(\partial_\alpha g_{\beta\mu} + \partial_\beta g_{\mu\alpha} - \partial_\mu g_{\alpha\beta})$ are the **Christoffel symbols** and $T_\alpha{}^\kappa{}_\beta = g_{\alpha\lambda}g^{\kappa\mu}T^\lambda{}_{\mu\beta}$.
The coefficients of a metric connection thus satisfy

$$\Gamma^\kappa{}_{\alpha\beta} = \Gamma^\kappa{}_{(\alpha\beta)} + \Gamma^\kappa{}_{[\alpha\beta]} = \{\kappa\alpha\beta\} + \underbrace{\frac{1}{2}\left(T_\alpha{}^\kappa{}_\beta + T_\beta{}^\kappa{}_\alpha + T^\kappa{}_{\alpha\beta}\right)}_{\equiv K^\kappa{}_{\alpha\beta} = \textbf{ contorsion}}.$$

If the torsion tensor vanishes, $T^\kappa{}_{\alpha\beta} = 0$, the metric connection is called the **Levi-Civita connection**:

$$\Gamma^\kappa{}_{\alpha\beta} = \{\kappa\alpha\beta\}.$$

10.9 Curvature And Torsion

We define two new tensors:
(Riemann) curvature tensor: $R : \chi(M) \times \chi(M) \times \chi(M) \to \chi(M)$
$R(X, Y, Z) \equiv R(X, Y)Z \equiv \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z.$
Torsion tensor: $T : \chi(M) \times \chi(M) \to \chi(M)$
$T(X, Y) \equiv \nabla_X Y - \nabla_Y X - [X, Y].$
Let's check that these definitions really define tensors, i.e. multilinear maps. Obviously $R(X + X', Y, Z) = R(X, Y, Z) + R(X', Y, Z)$ etc. are true, but it is less obvious that $R(fX, hY, gZ) = fghR(X, Y, Z)$ where $f, g, h \in \mathcal{F}(M)$. Let's calculate:

$$[fX, gY] = fX[g]Y - gY[f]X + fg[X, Y] \tag{15}$$

Using this we obtain

$$\begin{aligned} R(fX, gY)(hZ) &= f\nabla_X(g\nabla_Y(hZ)) - g\nabla_Y(f\nabla_X(hZ)) \\ &\quad - fX[g]\nabla_Y(hZ) + gY[f]\nabla_X(hZ) - fg\nabla_{[X,Y]}(hZ). \end{aligned}$$

Here the first term is

$$\begin{aligned} f\nabla_X(g\nabla_Y(hZ)) &= f\nabla_X(gY[h]Z + gh\nabla_Y Z) = fX[g]Y[h]Z + fg(X[Y[h]])Z \\ &\quad + fgY[h]\nabla_X Z + fgX[h]\nabla_Y Z + fhX[g]\nabla_Y Z + fgh\nabla_X\nabla_Y Z, \end{aligned}$$

and the second term is obtained by changing $X \leftrightarrow Y$ and $f \leftrightarrow g$. Continuing

$$\begin{aligned} R(fX, gY)(hZ) &= fX[g]Y[h]Z + fg(X[Y[h]])Z + fgY[h]\nabla_X Z + fgX[h]\nabla_Y Z \\ &\quad + fhX[g]\nabla_Y Z + fgh\nabla_X\nabla_Y Z - gY[f]X[h]Z - fg(Y[X[h]])Z \\ &\quad - fgX[h]\nabla_Y Z - fgY[h]\nabla_X Z - ghY[f]\nabla_X Z - fgh\nabla_Y\nabla_X Z \\ &\quad - fX[g]Y[h]Z - fhX[g]\nabla_Y Z + gY[f]X[h]Z + ghY[f]\nabla_X Z \\ &\quad - fg([X, Y][h])Z - fgh\nabla_{[X,Y]}Z = fgh(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z) \\ &\quad = fghR(X, Y)Z. \end{aligned}$$

Thus R is a linear map. In other words, when $X = X^\mu \partial_\mu, Y = Y^\nu \partial_\nu$ and $Z = Z^\lambda \partial_\lambda$, we have

$$R(X,Y)Z = X^\mu Y^\nu Z^\lambda R(\partial_\mu, \partial_\nu)\partial_\lambda.$$

R maps three vector fields to a vector field, so it is a (1,3)-tensor. A similar (but shorter) calculation shows that $T(fX, gY) = fgT(X,Y)$, so $T(X,Y) = X^\mu Y^\nu T(\partial_\mu, \partial_\nu)$. T is a (1,2) tensor.
The operations of R and T on vectors are obtained by knowing their actions on the basis vectors $\partial_\mu \frac{\partial}{\partial x^\mu}$. Denote

$$R(e_\mu, \partial_\nu)\partial_\lambda = \text{ a vector,(expand in basis } \partial_\kappa) = R^\kappa{}_{\lambda\mu\nu}\partial_\kappa.$$

Note the placement of indices. We can derive a formula for obtaining the components $R^\kappa{}_{\lambda\mu\nu}$. Recall that $[\partial_\mu, \partial_\nu] = 0$ and $dx^\kappa(\partial_\sigma) = \delta^\kappa{}_\sigma$. Thus we get

$$\begin{aligned} R^\kappa{}_{\lambda\mu\nu} &= dx^\kappa(R(\partial_\mu, \partial_\nu)\partial_\lambda) = dx^\kappa(\nabla_\mu\nabla_\nu\partial_\lambda - \nabla_\nu\nabla_\mu\partial_\lambda) = dx^\kappa(\nabla_\mu(\Gamma^\eta{}_{\nu\lambda}\partial_\eta) - \nabla_\nu(\Gamma^\eta{}_{\mu\lambda}\partial_\eta)) \\ &= dx^\kappa((\partial_\mu\Gamma^\eta{}_{\nu\lambda})\partial_\eta) + \Gamma^\eta{}_{\nu\lambda}\Gamma^\rho{}_{\mu\eta}\partial_\rho - (\partial_\nu\Gamma^\eta{}_{\mu\lambda})\partial_\eta - \Gamma^\eta{}_{\mu\lambda}\Gamma^\rho{}_{\nu\eta}\partial_\rho) \end{aligned} \tag{16}$$

Therefore

$$\boxed{R^\kappa{}_{\lambda\mu\nu} = \partial_\mu\Gamma^\kappa{}_{\nu\lambda} - \partial_\nu\Gamma^\kappa{}_{\mu\lambda} + \Gamma^\eta{}_{\nu\lambda}\Gamma^\kappa{}_{\mu\eta} - \Gamma^\eta{}_{\mu\lambda}\Gamma^\kappa{}_{\nu\eta}}$$

Similarly if we denote $T(\partial_\mu, \partial_\nu) = T^\lambda{}_{\mu\nu}\partial_\lambda$ and derive the components $T^\lambda{}_{\mu\nu}$:

$$T^\lambda{}_{\mu\nu} = dx^\lambda(T(\partial_\mu, \partial_\nu)) = dx^\lambda(\nabla_\mu\partial_\nu - \nabla_\nu\partial_\mu) = dx^\lambda(\Gamma^\eta{}_{\mu\nu}\partial_\eta - \Gamma^\eta{}_{\nu\mu}\partial_\eta),$$

and therefore

$$T^\lambda{}_{\mu\nu} = \Gamma^\lambda{}_{\mu\nu} - \Gamma^\lambda{}_{\nu\mu}.$$

Thus this is the <u>same</u> torsion tensor as the one we had defined earlier.

Geometric interpretation:

SEE THE FIGURES IN SECTION 7.3.2. OF NAKAHARA

Let us also define:
The Ricci tensor: $Ric(X,Y) = dx^\lambda(R(e_\lambda, Y)X)$. Thus the components are: $(Ric)_{\mu\nu} = Ric(e_\mu, e_\nu) = R^\lambda{}_{\mu\lambda\nu}$. (Usual notation $(Ric)_{\mu\nu} \equiv R_{\mu\nu}$.)
The scalar curvature: $R = g^{\mu\nu}(Ric)_{\mu\nu} = R^{\lambda\nu}{}_{\lambda\nu}$.
The Einstein tensor: $G_{\mu\nu} = (Ric)_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$.

10.10 Geodesics of Levi-Civita Connections

The length of a curve $c(s) = (x^\mu(s))$ is defined by

$$I(c) = \int_c ds = \int_c \sqrt{g_{\mu\nu}\frac{dx^\mu}{ds'}\frac{dx^\nu}{ds'}}ds' \equiv \int_c L ds'$$

Thus along a curve L is constant. One can normalize s' such that $L = 1$ so $s' = s$. Curves with extremal (minimum or maximum) length satisfy $\delta I = 0$ about the curve. (Variational principle.) They satisfy the Euler-Lagrange equations (familiar from calculus of variations (FYMM II)):

$$EL \frac{d}{ds}\left(\frac{\partial L}{dx'^\mu}\right) - \frac{\partial L}{dx^\mu} = 0, \qquad \text{where } x'^\mu = \frac{dx^\mu}{ds} \tag{17}$$

L = Lagrange function or Lagrangian. Instead of L, which contains a square root, we can equivalently use a simpler Lagrange function

$$F = \frac{1}{2}g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = \frac{1}{2}L^2,$$

because

$$\frac{d}{ds}\left(\frac{\partial F}{dx'^\mu}\right) - \frac{\partial F}{dx^\mu} = L\underbrace{\left(\frac{d}{ds}\left(\frac{\partial L}{dx'^\mu}\right) - \frac{\partial L}{dx^\mu}\right)}_{=0} + \frac{\partial L}{\partial x'^\mu}\underbrace{\frac{dL}{ds}}_{=0} = 0,$$

when $x^\mu(s)$ satisfies the Euler-Lagrange equation. Then $\delta(\int F ds) = 0$ gives

$$\begin{aligned}
&\frac{d}{ds}\left(g_{\lambda\mu}\frac{dx^\mu}{ds}\right) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\lambda}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0 \\
&\Rightarrow \frac{\partial g_{\lambda\mu}}{\partial x^\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} + g_{\lambda\mu}\frac{d^2x^\mu}{ds^2} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\lambda}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0 \\
&\Rightarrow g_{\lambda\mu}\frac{d^2x^\mu}{ds^2} + \frac{1}{2}\left(\frac{\partial g_{\lambda\mu}}{\partial x^\nu} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\right)\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0.
\end{aligned}$$

Multiply this by $g^{\kappa\lambda}$ and sum over λ:

$$\textit{geodesic}\ \boxed{\frac{d^2x^\kappa}{ds^2} + \{\kappa\mu\nu\}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0.} \tag{18}$$

This is the geodesic equation with a Levi-Civita connection! The action $I = \int F ds$ sometimes provides a convenient starting point for computing the Christoffel symbols $\{\kappa\mu\nu\}$: plug in the metric to I, derive the Euler- Lagrange equations and read off the Christoffel symbols comparing the Euler-Lagrange equations with the general geodesic equations.
Note: previously when we discussed the geodesic equation in the context of general connection, we said that geodesics are the "straightest" possible curves. Now, in the

context of the Levi-Civita connection which is only based on the metric, we that the geodesics are also the shortest possible curves.

Note also that we can explicitly restore a parameter m and write the action of the length of the curve as $I = m\int \sqrt{g_{\mu\nu}\frac{dx^\mu}{ds'}\frac{dx^\nu}{ds'}}ds'$. This is the relativistic action of a free massive point particle (with mass m) moving on a curved spacetime. Thus the free point particles move along geodesics. If $m^2 > 0$ (usual particles), we say that the corresponding geodesics (on a pseudo-Riemannian manifold) are **timelike**, if $m^2 < 0$ (tachyonic particles) the geodesics are **spacelike**. Massless particles (such as the photon) move along **null** geodesics. The invariant length vanishes along a null geodesic, $ds^2 = 0$. This equation can be used to determine the null geodesics.

10.11 Lie Derivative And the Covariant Derivative

Let ${\Gamma^\mu}_{\nu\lambda}$ be an arbitrary symmetric (${\Gamma^\mu}_{\nu\lambda} = {\Gamma^\mu}_{\lambda\nu}$) connection. We can then re-express the Lie derivative with the help of the covariant derivative as follows:

$$(\mathcal{L}_X Y)^\mu = X^\nu\partial_\nu Y^\mu - Y^\nu\partial_\nu X^\mu = X^\nu\nabla_\nu Y^\mu - (\nabla_\nu X^\mu)Y^\nu$$

This is true because of the symmetry of the connection:

$$\begin{aligned} X^\nu\nabla_\nu Y^\mu - (\nabla_\nu X^\mu)Y^\nu &= X^\nu(\partial_\nu Y^\mu + {\Gamma^\mu}_{\nu\lambda}Y^\lambda) - (\partial_\nu X^\mu + {\Gamma^\mu}_{\nu\lambda}X^\lambda)Y^\nu \\ X^\nu\partial_\nu Y^\mu - Y^\nu\partial_\nu X^\mu &+ (\underbrace{{\Gamma^\mu}_{\nu\lambda} - {\Gamma^\mu}_{\lambda\nu}}_{=0})X^\nu Y^\lambda \end{aligned}$$

For a generic (p,q)-tensor:

$$\begin{aligned} \mathcal{L}_X T^{\mu_1\ldots\mu_p}_{\nu_1\ldots\nu_q} &= (X^\lambda\nabla_\lambda)T^{\mu_1\ldots\mu_p}_{\nu_1\ldots\nu_q} - (\nabla_\lambda X^{\mu_1})T^{\lambda\mu_2\ldots\mu_p}_{\nu_1\ldots\nu_q} - \ldots - (\nabla_\lambda X^{\mu_p})T^{\mu_1\ldots\mu_{p-1}\lambda}_{\nu_1\ldots\nu_q} \\ &+ (\nabla_{\nu_1}X^\lambda)T^{\mu_1\ldots\mu_p}_{\lambda\nu_2\ldots\nu_q} + \ldots + (\nabla_{\nu_q}X^\lambda)T^{\mu_1\ldots\mu_p}_{\nu_1\ldots\nu_{q-1}\lambda}. \end{aligned}$$

10.12 Isometries

Isometries are a very important concept. They are symmetries of a Riemannian manifold. If the manifold is a spacetime, we usually require a physical theory to be invariant under isometries.

Definition. Let (M, g) be a (pseudo)-Riemannian manifold. A diffemorphism $f : M \to M$ is an **isometry** if it preserves the metric,

$$f^* g_{f(p)} = g_p \ ,$$

for all $p \in M$.

If we interpret the metric as a map on vector fields, the above requirement means

$$g_{f(p)}(f_*X, f_*Y) = g_p(X, Y)$$

for all tangent vectors $X, Y \in T_pM$. In component form the above equation is

$$\frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}\, g_{\alpha\beta}(f(p)) = g_{\mu\nu}(p) \tag{19}$$

where x, y are coordinates of the points $p, f(p)$ respectively. This means that an isometry must preserve the angles between all tangent vectors and their lengths.

The identity map is trivially an isometry, also the composite map $f \circ g$ of two isometries f, g is an isometry. Further, if f is an isometry, so is its inverse f^{-1}. This means that isometries form a group with composition of maps as the product, called the **isometry group**. The isometry group is a group of symmetries of a (pseudo)-Riemannian manifold.

Examples.

- $(M, g) =$ the Euclidean space (R^n, δ) with the Euclidean metric. All translations $x^\mu \mapsto x^\mu + a^\mu$ in some direction $a = (a^\mu)$ are isometries, and so are rotations. The isometry group $\{$translations, rotations, and their combinations$\}$ is called the **Euclidean group** or **Galilean group** and denoted by E^n.
- $(M, g) =$ the $(d+1)$-dimensional Minkowski space(time) $(R^{1,d}, \eta)$ with the Minkowski metric η. Again, spacetime translations $x\mu \mapsto x^\mu + a^\mu$ are isometries, additional isometries are (combinations of these and) space rotations and boosts. The isometry group $\{$translations, rotations, boosts, and their combinations$\}$ is called the **Poincaré** group.

In typical laboratory scales, our spacetime is approximately flat (a Minkowski space) so its approximate isometry group is the Poincaré group. That's the reason for special relativity and the requirement that physics in the laboratory be relativistic, *i.e.* Poincaré invariant. More precisely, that requirement is necessary for experiments which involve scales where relativistic effects become important. For lower scales, time "decouples" and we can make a further approximation where only the Euclidean isometries of the spacelike directions are relevant. Recall also that symmetries such as the time translations and space translations lead into conservation laws, like the conservation of energy and momentum. As you can see, important physical principles are a reflection of the isometries of the spacetime.

10.13 Killing Vector Fields

Let us now consider the limit of "small" isometries, *i.e.* infinitesimal displacements $x = p \mapsto f(p) = y \simeq x + \epsilon X$. Here ϵ is an infinitesimal parameter and X is a vector field indicating the direction of the infinitesimal displacement. If the above map is an isometry, the vector field X is called a **Killing vector field**. Since the infinitesimal displacement is an isometry, eqn. must be satisfied and it now takes the form

$$\frac{\partial(x^\alpha + \epsilon X^\alpha)}{\partial x^\mu}\frac{\partial(x^\beta + \epsilon X^\beta)}{\partial x^\nu}\ g_{\alpha\beta}(x + \epsilon X) = g_{\mu\nu}(x) \tag{20}$$

By Taylor expanding the left hand side, and requiring that the leading infinitesimal term of order ϵ vanishes (there's no ϵ-dependence on the right hand side), we obtain the equation

$$X^\xi \partial_\xi g_{\mu\nu} + \partial_\mu X^\alpha g_{\alpha\nu} + \partial_\nu X^\beta g_{\mu\beta} = 0\ . \tag{21}$$

We can recognize the left hand side as a Lie derivative, so (21) can be rewritten as

$$\mathcal{L}_X g_{\mu\nu} = 0\ .$$

Expressing $\mathcal{L}_X g_{\mu\nu}$ with the help of the covariant derivative,

$$\mathcal{L}_X g_{\mu\nu} = X^\lambda \overbrace{\nabla_\lambda g_{\mu\nu}}^{=0} + (\nabla_\mu X^\lambda) g_{\lambda\nu} + (\nabla_\nu X^\lambda) g_{\mu\lambda} = 0.$$

$(\nabla_\lambda g_{\mu\nu} = 0)$ for a metric connection). Thus Killing vector field satisfies

$$\boxed{\nabla_\mu X_\nu + \nabla_\nu X_\mu = 0} \qquad \text{Killing equation.}$$

Let X and Y be two Killing vector fields. We can easily verify that

a) all linear combinations $aX + bY$ with $a, b \in R$ are also Killing vector fields

b) the Lie bracket $[X, Y]$ is a Killing vector field

It then follows that the Killing vector fields form an algebra, the **Lie algebra of the isometry group**. (The isometry group is usually a Lie group.)

Now let $x^\mu(t)$ be a geodesic, its tangent vector $U^\mu = \frac{dx^\mu}{dt}$, and let V^μ be a Killing vector. Then,

$$(U^\nu \nabla_\nu)(U^\mu V_\mu) = \underbrace{U^\mu U^\nu \nabla_\nu V_\mu}_{=\frac{1}{2}U^\mu U^\nu(\nabla_\mu V_\nu + \nabla_\nu V_\mu)} + V_\mu \underbrace{U^\nu \nabla_\nu U^\mu}_{=0\ \ \text{(geodesic)}} = 0.$$

Thus $U^\mu V_\mu = U \cdot V$ is a *constant on a geodesic.*

An m-dimensional manifold M can have at most $\frac{1}{2}m(m+1)$ linearly independent Killing vector fields. Manifold with the maximum number of Killing vector fields are called **maximally symmetric**. E.g. $\mathbb{R}^m$ is maximally symmetric ($g_{\mu\nu} = \delta_{\mu\nu} \ \Rightarrow \Gamma = 0$). The Killing equation $\partial_\mu V_\nu + \partial_\nu V_\mu = 0$ has solutions:

$$\begin{aligned} &V^\mu_{(i)} = \delta^\mu_i \quad (m \text{ of these}) \\ &V_\mu = a_{\mu\nu} x^\nu \quad \text{with} \quad \underbrace{a_{\mu\nu} = -a_{\nu\mu}}_{\frac{1}{2}m(m-1)\text{ components}} \ = \ \text{constant} \ \neq 0 \end{aligned} \tag{22}$$

Thus in total we have $m + \frac{1}{2}m(m-1) = \frac{1}{2}m(m+1)$. Ok.

10.14 APPENDIX: Einstein's Field Equations

The *Einstein tensor* is defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R, \tag{23}$$

where $R = g^{\mu\nu} R_{\mu\nu}$ is the *Ricci scalar*. We assume that the metric $g_{\mu\nu}$ is pseudo-Riemannian of signature $(1, 3)$ (one positive direction and three negative directions). The connection is the Levi-Civita connection computed from the metric and $R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu}$ is the Ricci tensor.

Exercise 1 Writing $R_{\alpha\beta\mu\nu} = g_{\alpha\lambda} R^{\lambda}_{\beta\mu\nu}$, show that

$$R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu} = -R_{\alpha\beta\nu\mu} = R_{\mu\nu\alpha\beta}.$$

Show that this implies that $R_{\mu\nu}$ is symmetric.

The Einstein tensor is symmetric. Furthermore, its *covariant divergence* vanishes,

$$\nabla_\mu G^{\mu\nu} = \partial_\mu G^{\mu\nu} + \Gamma^{\mu}_{\mu\alpha} G^{\alpha\nu} + \Gamma^{\nu}_{\mu\alpha} G^{\mu\alpha} = 0. \tag{24}$$

This is seen as follows. First, taking $Z = \partial_\alpha, X = \partial_\mu, Y = \partial_\nu$ and using the defining properties of a metric connection, we obtain

$$\partial_\alpha g_{\mu\nu} = \Gamma^{\beta}_{\alpha\mu} g_{\beta\nu} + \Gamma^{\beta}_{\alpha\nu} g_{\mu\beta} = \Gamma_{\alpha\nu\mu} + \Gamma_{\alpha\mu\nu}. \tag{25}$$

This can be also written as

$$(\nabla_\alpha g)_{\mu\nu} = 0. \tag{26}$$

For the inverse tensor $g^{\mu\nu} = (g^{-1})_{\mu\nu}$, one gets

$$\partial_\alpha g^{\mu\nu} + \Gamma^{\nu}_{\alpha\beta} g^{\mu\beta} + \Gamma^{\mu}_{\alpha\beta} g^{\beta\nu} = 0. \tag{27}$$

Note the difference in sign for the covariant derivative of the metric tensor and its inverse.

Exercise 2 For any vector field $X = X^\mu \partial_\mu$ the components of the covariant derivatives are $(\nabla_\nu X)^\mu = \partial_\nu X^\mu + \Gamma^{\mu}_{\nu\alpha} X^\alpha$. Show that the *covariant divergence* is given by

$$(\nabla_\mu X)^\mu = (-\det g)^{-1/2} \partial_\mu ((-\det g)^{1/2} X^\mu).$$

In relativity theory literature, it is a custom to use the abbreviation $X_{\mu;\nu} = (\nabla_\nu X)_\mu$ for the covariant differentiation of vector (and higher order tensor) indices. With this notation, we can write the second Bianchi identity as

$$R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\nu\lambda;\mu} + R_{\alpha\beta\lambda\mu;\nu} = 0. \tag{28}$$

Contracting the α and μ indices in this identity with the metric tensor, we get

$$g^{\alpha\mu} (R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\nu\lambda;\mu} + R_{\alpha\beta\lambda\mu;\nu}) = 0. \tag{29}$$

By the definition of the Ricci tensor, this can be written as

$$R_{\beta\nu;\lambda} + R^{\mu}_{\beta\nu\lambda;\mu} - R_{\beta\lambda;\nu} = 0, \tag{30}$$

where we have taken into account that the covariant derivative of $g^{\mu\nu}$ vanishes, implying that the multiplication with the components of the metric tensor commutes with covariant differentiation; in particular, index raising and lowering commutes with covariant derivatives. Using the results of Exercise 1, we get

$$g^{\alpha\mu} R_{\alpha\beta\lambda\mu;\nu} = -g^{\alpha\mu} R_{\alpha\beta\mu\lambda;\nu} = -R_{\beta\lambda;\nu}. \tag{31}$$

Contracting Eq. (30) once again with $g^{\beta\nu}$, we get

$$g^{\beta\nu}(R_{\beta\nu;\lambda} + R^{\mu}_{\beta\nu\lambda;\mu} - R_{\beta\lambda;\nu}) = 0, \tag{32}$$

or in other words,

$$R_{;\lambda} - R^{\mu}_{\lambda;\mu} - R^{\nu}_{\lambda;\nu} = 0. \tag{33}$$

Note that since R is a scalar, $R_{;\mu} = \partial_\mu R$. An equivalent form of the previous equation is

$$(2R^{\mu}_{\lambda} - \delta^{\mu}_{\lambda} R)_{;\mu} = 0. \tag{34}$$

Raising the index λ and dividing by 2 finally leads to

$$\left(R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R \right)_{;\mu} = 0. \tag{35}$$

Einstein's gravitational field equations are written simply as

$$G^{\mu\nu} = 8\pi \frac{G}{c^4} T^{\mu\nu}, \tag{36}$$

where G on the right-hand side (not to be confused with Einstein's tensor!) is Newton's gravitational constant and $T^{\mu\nu}$ is the *stress-energy (energy-momentum) tensor*. It describes the distribution of matter and energy in space-time. For example, the electromagnetic field gives a contribution to $T_{\mu\nu}$ defined by $T^{EM}_{\mu\nu} = \epsilon_0 F_{\mu}{}^{\lambda} F_{\lambda\nu} + \frac{\epsilon_0}{4} g_{\mu\nu} F^{\lambda\omega} F_{\lambda\omega}$.

Another example is the energy-momentum tensor of a *perfect fluid* .A perfect fluid is characterized by a 4-velocity field u, a scalar density field ρ_0 and a scalar pressure field p. The energy-momentum tensor is defines as

$$T_{\mu\nu} = (\rho_0 + p) u_\mu u_\nu - p g_{\mu\nu}.$$

A special case of this is $p = 0$ which can be viewed as the energy momentum tensor of a flow of noninteracting dust particles. Normally p and ρ_0 are not independent but they are related by the equation of state of the form $p = p(\rho_0, T)$, where T

is the temperature. The requirement that the covariant divergence of the energy-momentum tensor vanishes leads to equations of motion for the perfect fluid. In fact, in case of Minkowski space-time and in a certain limit one gets the classical *Navier-Stokes equations* (from $\partial^\mu T_{\mu k} = 0$ for $k = 1, 2, 3$),

$$\rho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} \right] = -\nabla p$$

and the continuity equation (from $\partial^\mu T_{\mu 0} = 0$),

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0.$$

Here $\rho = \rho_0(1 - \mathbf{u}^2)$.

Let S be some space-like surface with a time-like unit normal vector field n^μ, $n_0 > 0$. Then,

$$\int_S (-\det g)^{1/2} T^{\mu\nu} n_\nu \, d^3x$$

gives the energy and momentum contained in S. Equation (24) leads to the following conservation law of energy and momentum. Suppose that the metric $g_{\alpha\beta}$ does not depend on a particular coordinate x_μ. Then,

$$0 = \partial_\mu g_{\alpha\beta} = \Gamma_{\mu\beta\alpha} + \Gamma_{\mu\alpha\beta} = \Gamma_{\alpha\beta\mu} + \Gamma_{\beta\alpha\mu}. \tag{37}$$

Thus, $\Gamma_{\alpha\beta\mu}$ is antisymmetric in the first two indices. Now,

$$(\nabla_\nu T)^\nu_{\ \mu} = \partial_\nu T^\nu_{\ \mu} + \Gamma^\nu_{\nu\lambda} T^\lambda_{\ \mu} - \Gamma^\lambda_{\nu\mu} T^\nu_{\ \lambda}. \tag{38}$$

The third term on the right-hand side is equal to $-\Gamma_{\nu\lambda\mu} T^{\nu\lambda}$ and it vanishes because the second factor is symmetric in its indices, whereas the first factor is antisymmetric in λ and ν by the remark above. On the other hand, the sum of the first two terms is $(-g)^{-1/2} \partial_\nu [(-g)^{1/2} T^\nu_{\ \mu}]$, according to the result of Exercise 2. Thus, for fixed μ, $j^\nu = (-g)^{1/2} T^\nu_{\ \mu}$ is conserved in the usual sense,

$$\partial_\nu j^\nu = 0. \tag{39}$$

In order to avoid convergence problems with the infinite integrals, we assume that all energy and momentum are contained in a compact region K in space-time. Consider a surface S, consisting of two space-like components S_1 and S_2 and some surface S_3 'far away' such that T vanishes on S_3. Using Gauss' law and the current conservation, we conclude that the surface integral of $(-\det g)^{1/2} T^\nu_{\ \mu} n_\nu$ over S vanishes. In other words,

$$\int_{S_1} (-\det g)^{1/2} T^\nu_{\ \mu} n_\nu \, d^3x = \int_{S_2} (-\det g)^{1/2} T^\nu_{\ \mu} n_\nu \, d^3x. \tag{40}$$

We have taken into account that, since n is future pointing, one of the normal vector fields on S_1 and S_2 is outward directed and the second inward directed. Equation (40) tells us that the stress-energy, in the μ-direction, on S_1 is the same as the corresponding quantity on S_2; one could think of S_i as a fixed time slice at time t_i and one obtains the usual law of conservation of energy or momentum.

Often one uses units in which $G = 1$ and $c = 1$ so that one does not need to write explicitly the coefficient G/c^4 in Einstein's equations.

10.14.1 The Newtonian Limit

It is known that the Newtonian gravitational theory is valid for fields, which can produce only velocities much smaller than the velocity of light. Since the components T^{0i} and T^{ij} are related to spatial momenta and T^{00} is related to energy, this condition says that $|T^{00}|$ is much larger than the other components. Because of Einstein's equations, the same is true for the components of the Einstein tensor. Furthermore, we expect that for weak gravitational fields the metric $g^{\mu\nu}$ differs slightly from the Minkowski metric $\eta^{\mu\nu}$,

$$g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu} \tag{41}$$

for a small perturbation $h^{\mu\nu}$. Next, we compute the connection, curvature, and finally the Ricci tensor to first order in the perturbation $h^{\mu\nu}$. A straight-forward computation, starting from the definitions of the various tensors, gives $G^{\mu\nu} = -\frac{1}{2}\Box(h^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}h)$, where $h = \eta_{\mu\nu}h^{\mu\nu}$. Thus, Einstein's equations, in this approximation, are linear,

$$-\frac{1}{2}\Box\left(h^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}h\right) = 8\pi\frac{G}{c^4}T^{\mu\nu}. \tag{42}$$

Taking into account the remark in the beginning of this section, only the 00-component is relevant,

$$\Box\left(h^{00} - \frac{1}{2}h\right) = -16\pi\frac{G}{c^2}\rho, \tag{43}$$

where $\rho = T^{00}/c^2$ is the matter density in the rest system of the source. We can also drop the time derivatives (in the system of coordinates, where the source is slowly moving, because $\partial_0 = \frac{1}{c}\partial_t$) and so the only relevant equation becomes

$$\nabla^2\left(h^{00} - \frac{1}{2}h\right) = 16\pi\frac{G}{c^2}\rho. \tag{44}$$

This means that,

$$h^{00} - \frac{1}{2}h = \frac{4}{c^2}\phi, \tag{45}$$

where ϕ is the gravitational potential for the matter distribution ρ. (Compare Eq. (62) with the Newtonian equation $\nabla^2\phi = 4\pi G\rho$, where $\phi = -GM/r$!)

Since all the other components of $h^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu} h$ vanish at this order of approximation, we finally get

$$h^{\mu\mu} = \frac{2}{c^2}\phi = -\frac{2GM}{c^2 r} \text{ (no summation!)} \tag{46}$$

for all $\mu = 0, 1, 2, 3$.

Next, we shall compute the geodesics for the metric $g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu}$ in the linear approximation (we neglect higher order terms in $h^{\mu\nu}$). For small velocities, the time component $\dot{x}_0(s)$ of the 4-velocity is much larger than the spatial components. For this reason, we can approximate the geodesic equations of motion as

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{00} \left(\frac{dx^0}{ds} \right)^2 = 0. \tag{47}$$

In the linear approximation,

$$\Gamma^0_{00} = \partial_0 \phi, \;\; \Gamma^i_{00} = \partial_i \phi. \tag{48}$$

Thus, the geodesic equations become

$$\ddot{x}_0 + \partial_0 \phi (\dot{x}_0)^2 = 0, \;\; \ddot{x}_i + \partial_i \phi (\dot{x}_0)^2 = 0. \tag{49}$$

In the coordinate system, where the source is at rest, the first equation says that we can choose the time t as the geodesic parameter, $x_0(s) = s = ct$, and then the second equation becomes

$$\ddot{x}_i = -\partial_i \phi. \tag{50}$$

The right-hand side (after multiplication by the mass m of the test particle) is the gravitational force of the source on m, so this equation is just Newton's second law, $m\mathbf{a} = \mathbf{F}$, where $\mathbf{F} = -\nabla \Phi$ and $\Phi = m\phi$.

10.14.2 The Schwarzschild Metric

The basic problem in Newtonian celestial mechanics is to solve the equations of motions outside of a spherically symmetric mass distribution (orbits of the planets around the Sun, orbits of satellites around the Earth). In general relativity the first natural problem is to search for spherically symmetric solutions of Einstein's equations.

Actually, there is a unique 1-parameter family of spherically symmetric solutions, which are *asymptotically flat*, meaning that at large distances from the source the metric tends to the flat Minkowski metric $ds^2 = dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2$. This is the content of Birkhoff's theorem (which we are not going to prove). The line element of the metric is given as

$$ds^2 = \left(1 - \frac{2GM}{c^2 r}\right) dx_0^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2 d\Omega^2, \tag{51}$$

where $d\Omega^2$ is the angular part of the Euclidean metric in $\mathbb{R}^3$, $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$. It is clear from Eq. (51) that for large distances r the metric approaches the Minkowski metric. The line element (51) is called the *Schwarzschild metric.*

When $r > 2GM/c^2$ the Schwarzschild metric is supposed to describe the gravitational field outside of a spherically symmetric star. The other disconnect region $r < 2GM/c^2$ is the *Schwarzschild black hole.* The singularity at $r = 2GM/c^2$ is actually due to a bad choice of coordinates. There is a way to glue the inside solution in a smooth way to the outside solution by a suitable choice of coordinates; the complete discussion of this was first given by Kruskal and Szekeres in 1960. The Kruskal–Szekeres metric is given as follows. The coordinates are denoted by (u, v, θ, ϕ). The latter two are the ordinary spherical coordinates on a unit sphere. The coordinates (u, v) are restricted to the region $L \subset \mathbb{R}^2$ defined by

$$uv < \frac{2GM}{c^2 e}.$$

The metric is then

$$ds^2 = \frac{16\mu^2}{r} e^{(2\mu - r)/2\mu} du dv - r^2 d\Omega^2, \tag{52}$$

where $\mu = MG/c^2$ and r is a function of u, v defined by

$$uv = (2\mu - r) e^{(r - 2\mu)/2\mu}. \tag{53}$$

Note that $f(x) = xe^{x/a}$ is monotonically increasing when $x > -a$ (and $f(x) > -a/e$) and therefore $y = f(x)$ has a unique solution x for any $y > -a/e$. We treat u as a kind of universal time; a time-like vector is future directed if its projection to ∂_u is positive. The orientation (needed in integration!) is defined by the ordering (v, u, θ, ϕ) of coordinates. Note that the radial null lines (radial light rays) are given by $du = 0$ or $dv = 0$.

The Kruskal–Szekeres space-time can be divided into four regions: K_1 consists of points $v > 0$, $u < 0$, region K_2 of points $u, v > 0$, in region K_4 we have $u, v < 0$, and finally region K_3 is characterized by $u > 0$, $v < 0$. The boundaries between these regions are non-singular points for the metric. The only singularities are at the boundary $uv = 2\mu/e$.

The region K_1 is equivalent with the outer region of a Schwarzschild space-time. This is seen by performing the coordinate transformation $(v, u, \theta, \phi) \mapsto (t, r, \theta, \phi)$, where $r = r(u, v)$ as above and the Schwarzschild time is $t = 2\mu \ln(-v/u)$. With a similar coordinate transformation the region K_3 is seen to be equivalent with the outer Schwarzschild solution. The region K_2 is equivalent with the Schwarzschild black hole. The equivalence is obtained through the coordinate transformation $(v, u, \theta, \phi) \mapsto (t, r, \theta, \phi)$, where $r = r(u, v)$ is the same as before but now $t = 2\mu \ln(v/u)$.

It is easy to construct smooth time-like curves which go from either K_1 or K_3 to the black hole K_2. However, we shall prove that once an observer falls to the black hole there is no way to go back to the 'normal' regions K_1 and K_3.

Let $x(t)$ be the time-like path of the observer. Then along the path

$$\frac{dr}{dt} = \frac{\partial r}{\partial u}\frac{du}{dt} + \frac{\partial r}{\partial v}\frac{dv}{dt} = \frac{r}{8\mu^2}e^{(r-2\mu)/2\mu}\left[\frac{\partial r}{\partial u}g(\partial_v, x'(t)) + \frac{\partial r}{\partial v}g(\partial_u, x'(t))\right] < 0,$$

since $x(t)$ is time-like and in K_2 holds $r\frac{\partial r}{\partial u} = -2\mu v e^{(2\mu-r)/2\mu} < 0$ and similarly for the v-coordinate.

The boundary between K_2 and the normal regions is $r = 2\mu$ (*i.e.*, $u = 0$ or $v = 0$). The function $r(x(t))$ was seen to be decreasing, and therefore the path $x(t)$ can never hit the boundary $r = 2\mu$. But the observer entering K_2 has a deplorable future, since it will eventually hit the true singularity $r = 0$, again using the monotonicity of the function $r(x(t))$.

There is also another singularity, the outer boundary of region K_3. But this is of no great concern because it is in the past; no future directed time-like curve can enter that singularity.

11 Principal bundles and Yang-Mills systems

11.1 Cartan's structural equations

Let $X_1, \ldots, X_n$ be a basis of $Lie(G)$. Then

$$[X_i, X_j] = c_{ij}^k X_k$$

where the c_{ij}^k are the structural constants. Since the Lie bracket is antisymmetric we have $c_{ij}^k = -c_{ji}^k$ and by the Jacobi identity we have

$$c_{ij}^k c_{kl}^m + c_{li}^k c_{kj}^m + c_{jl}^k c_{ki}^m = 0$$

for all i, j, l, m. In terms of the left invariant vector fields X_i, any tangent vector v at $g \in G$ can be written as $v = v^i X_i(g)$. Let us define $\theta^i \in \Omega^1(G)$ as $\theta^i(g)v = v^i$. We compute the exterior derivative $d\theta^i$:

$$\begin{aligned} d\theta^i(g)(X_j, X_k) &= X_j\theta^i(X_k) - X_k\theta^i(X_j) - \theta^i([X_j, X_k]) \\ &= X_j\delta_{ik} - X_k\delta_{ij} - \theta^i(c_{jk}^l X_l) = -c_{jk}^i. \end{aligned}$$

On the other hand,

$$(\theta^i \wedge \theta^j)(X_k, X_l) = \theta^i(X_k)\theta^j(X_l) - \theta^i(X_l)\theta^j(X_k) = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}.$$

Thus we obtain *Cartan's structural equations,*

$$d\theta^i = -\frac{1}{2}c_{kl}^i \theta^k \wedge \theta^l.$$

Denote $X_i\theta^i = g^{-1}dg$. This is a $Lie(G)$-valued 1-form on G. It is tautological at the identity: $(g^{-1}dg)(v) = v$ for $v \in T_1G$. For $\theta = g^{-1}dg$ the structural equations can be written as

$$d\theta + \frac{1}{2}[\theta \wedge \theta] = 0,$$

where $[\theta \wedge \theta] = [X_i, X_j]\theta^i \wedge \theta^j$.

11.2 Principal bundles

Let G be a Lie group and M a smooth manifold. *A principal G bundle over M* is a manifold which locally looks like $M \times G$.

Definition *A smooth manifold P is a principal G bundle over the manifold M, if a smooth right action of G on P is given, i. e., a map $P \times G \to P$, $(p, g) \mapsto pg$, such that $p(gg') = (pg)g' \forall p \in P$ and g, g' in G, and if there is given a smooth map $\pi : P \to M$ such that*

- *$\pi(pg) = \pi(p)$ for all g in G.*
- *$\forall x \in M$ there exists an open neighborhood U of x and a diffeomorphism (local trivialization) $f : \pi^{-1}(U) \to U \times G$ of the form $f(p) = (\pi(p), \phi(p))$ such that $\phi(pg) = \phi(p)g \, \forall p \in \pi^{-1}(U), g \in G$.*

The manifold P is the total space of the bundle, M is the *the base space,* and π is the bundle projection. The *trivial bundle* $P = M \times G$ is defined by the projection $\pi(x, g) = x$ and by the natural right action of G on itself.

Consider two bundles $P_i = (P_i, \pi_i, M_i; G)$ with the same structure group G. A smooth map $\phi : P_1 \to P_2$ is a *G bundle map* , if $\phi(pg) = \phi(p)g$ for all p and g. Two bundles P_1 and P_2 are *isomorphic* if there is a bijective bundle map $P_1 \to P_2$. An isomorphism of a bundle onto itself is an *automorphism* .

If $H \subset G$ is a closed subgroup then G is a principal H bundle over the homogeneous space G/H. The right action of H on G is just the right multiplication in G and the projection is the canonical projection on the quotient.

Example Take $G = SU(2)$ and $H = U(1)$

$$H : \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix}, \varphi \in \mathbb{R}.$$

A general element g of G is

$$g = \begin{pmatrix} z_1 & -\overline{z}_2 \\ z_2 & \overline{z}_1 \end{pmatrix},$$

with $|z_1|^2 + |z_2|^2 = 1$. Writing z_1 and z_2 in terms of their real and imaginary parts we see that the group G can be identified with the unit sphere S^3 in $\mathbb{R}^4$. We can define a map $\pi : G \to S^2$ by $\pi(g) = g\sigma_3 g^{-1}$, where σ_3 is the matrix $diag(1, -1)$; elements of $\mathbb{R}^3$ are represented by Hermitian traceless 2×2 matrices. The Euclidean metric is given by $\|x\|^2 = -\det x$. The kernel of the map π is precisely $U(1)$; thus we have a $U(1)$ fibration over $S^2 = SU(2)/U(1)$ in S^3.

Exercise Let $S_+ = \{x \in S^2 | x_3 \neq -1\}$ and $S_- = \{x \in S^2 | x_3 \neq +1\}$. Construct local trivializations $f_\pm : \pi^{-1}(S_\pm) \to S_\pm \times U(1)$.

The bundle $S^3 \to S^2$ is nontrivial; it is not isomorphic to $S^2 \times S^1$ for topological reasons. Namely, S^3 is a simply connected manifold whereas the fundamental group of $S^2 \times S^1$ is equal to $\pi_1(S^1) = \mathbb{Z}$ [M. Greenberg: Lectures on Algebraic Topology].

Let $\{U_\alpha\}_{\alpha\in\Lambda}$ be an open cover of the base space M of a principal bundle P and let $p \mapsto (\pi(p), \phi_\alpha(p)) \in U_\alpha \times G$ be a set of local trivializations. If $p \in \pi^{-1}(U_\alpha \cap U_\beta)$, we can write

$$\phi_\alpha(p) = \xi_{\alpha\beta}(p)\phi_\beta(p),$$

where $\xi_{\alpha\beta}(p) \in G$. Now $\phi_\alpha(pg) = \phi_\alpha(p)g$ and $\phi_\beta(pg) = \phi_\beta(p)g$ from which follows that $\xi_{\alpha\beta}(pg) = \xi_{\alpha\beta}(p)$ and thus $\xi_{\alpha\beta}$ can be thought of as a function on the base space $U_\alpha \cap U_\beta$. If $p \in \pi^{-1}(U_\alpha \cap U_\beta \cap U_\gamma)$ and $x = \pi(p)$, then $\phi_\alpha(p) = \xi_{\alpha\beta}(x)\phi_\beta(p) = \xi_{\alpha\beta}(x)\xi_{\beta\gamma}(x)\phi_\gamma(p)$ so that

$$\xi_{\alpha\beta}(x)\xi_{\beta\gamma}(x) = \xi_{\alpha\gamma}(x).$$

In general, a collection of G-valued functions $\{\xi_{\alpha\beta}\}$ for the covering $\{U_\alpha\}$ is a *one-cocycle (with values in G)* if the above equation holds for all x in $U_\alpha \cap U_\beta \cap U_\gamma$ and for all triples of indices.

If we make the transformations $\phi'_\alpha = \eta_\alpha \phi_\alpha$ for some functions $\eta_\alpha : U_\alpha \to G$, then

$$\xi_{\alpha\beta} \mapsto \xi'_{\alpha\beta} = \eta_\alpha^{-1}\xi_{\alpha\beta}\eta_\beta.$$

If we can find the maps η_α such that $\xi'_{\alpha\beta} = 1 \forall \alpha, \beta$, then $\xi_{\alpha\beta} = \eta_\alpha \eta_\beta^{-1}$ and we say that the one-cocycle ξ is a *coboundary*.

Let $(P, \pi, M), (P', \pi', M')$ be a pair of principal G bundles and let $f : P \to P'$ be a bundle map. We define the induced map $\hat{f} : M \to M'$ by $\hat{f}(x) = \pi'(f(p))$, where p is an arbitrary element in the fiber $\pi^{-1}(x)$.

Theorem 11.1 *Let P and P' be a pair of principal G bundles over M. Let $\{U_\alpha, \phi_\alpha\}_{\alpha\in\Lambda}$ (respectively, $\{U_\alpha, \phi'_\alpha\}_{\alpha\in\Lambda}$) be a system of local trivializations for P (respectively, for P'). Let $\xi_{\alpha\beta}$ and $\xi'_{\alpha\beta}$ be the corresponding transition functions. Then there exists an isomorphism $f : P \to P'$ such that $\hat{f} = id_M$ if and only if the transition functions differ by a coboundary, that is, $\xi'_{\alpha\beta}(x) = \eta_\alpha(x)^{-1}\xi_{\alpha\beta}(x)\eta_\beta(x)$ in $U_\alpha \cap U_\beta$ for some functions $\eta_\alpha : U_\alpha \to G$.*

Proof. 1) Suppose first that $\xi'_{\alpha\beta} = \eta_\alpha^{-1}\xi_{\alpha\beta}\eta_\beta$ for all $\alpha, \beta \in \Lambda$. Define $f : P \to P'$ as follows. Let $p \in P$ and $x = \pi(p)$. Choose $\alpha \in \Lambda$ such that $x \in U_\alpha$. Using a local trivialization $(U_\alpha, \phi'_\alpha$ at x we set $f(p) = (x, f_\alpha(p))$, where $f_\alpha(p) = \eta_\alpha(x)^{-1}\phi_\alpha(p)$. We have to show that the map is well-defined: If $x \in U_\alpha \cap U_\beta$ then $\phi_\beta(p) = \xi_{\beta\alpha}(x)\phi_\alpha(p)$ and thus

$$\begin{aligned} f_\beta(p) &= \eta_\beta(x)^{-1}\phi_\beta(p) = \eta_\beta(x)^{-1}\xi_{\beta\alpha}(x)\phi_\alpha(p) \\ &= \xi'_{\beta\alpha}(x)[\eta_\alpha(x)^{-1}\phi_\alpha(p)] = \xi'_{\beta\alpha}r(x)f_\alpha(p). \end{aligned}$$

We conclude that $(x, f_\alpha(p))$ and $(x, f_\beta(p))$ represent the same element in P'. The equation $f(pg) = f(p)g$ follows from $\phi_\alpha(pg) = \phi_\alpha(p)g$.

2) Let $f : P \to P'$ be an isomorphism. We can define

$$\eta_\alpha(x) = \phi_\alpha(p)\phi'_\alpha(f(p))^{-1},$$

where $p \in \pi^{-1}(x)$ is arbitrary. It follows at once from the definition of the transition functions that the collection $\{\eta_\alpha\}_{\alpha\in\Lambda}$ satisfies the requirements.

Let $\{\xi_{\alpha\beta}\}_{\alpha,\beta\in\Lambda}$ be a one-cocycle with values in G, subordinate to an open cover $\{U_\alpha\}$ on a manifold M. We can construct a principal G bundle P from this data. Let $C = \amalg(\alpha, U_\alpha \times G)$ be the disjoint union of all the sets $U_\alpha \times G$. Define an equivalence relation in C by $(\alpha, x, g) \sim (\alpha', x', g')$ if and only if $x = x'$ and $g' = \xi_{\alpha'\alpha}(x)g$. Set $P = C/\sim$. The action of G in P is given by $(\alpha, x, g)g_0 = (\alpha, x, gg_0)$. The smooth structure on P is defined such that the sets $U_\alpha \times G$ are smooth coordinate charts for P.

Exercise Complete the construction of P above.

Let (P, π, M) be a principal G bundle. A *(global) section* of P is a map $\psi : M \to P$ such that $\pi \circ \psi = id_M$.

Exercise Show that a principal bundle is trivial if and only if it has a global section.

A local section consists of an open set $U \subset M$ and a map $\psi : U \to P$ such that $\pi \circ \psi = id_U$. If $f : \pi^{-1}(U) \to U \times G$ is a local trivialization we can define a local section by $\psi(x) = f^{-1}(x, h(x))$, where $h : U \to G$ is an arbitrary (smooth) function.

Let $H \subset G$ be a closed subgroup. We say that the bundle P has been *reduced* to a principal H *subbundle* Q, if $Q \subset P$ is a submanifold such that $qh \in Q$ for all $q \in Q, h \in H, \pi(Q) = M$ and H acts transitively in each fiber $Q_x = \pi^{-1}(x) \cap Q$.

Any manifold M of dimension n carries a natural principal $GL(n, \mathbb{R})$ bundle, namely, the bundle FM of linear frames. The fiber F_xM at a point $x \in M$ consists of all frames (ordered basis) of the tangent space T_xM. The group $GL(n, \mathbb{R})$ acts in F_xM by $(f_1, f_2, ..., f_n)A = (\sum_{i=1}^n A_{i1}f_i, \sum_{i=1}^n A_{i2}f_i, ..., \sum_{i=1}^n A_{in}f_i)$, where the f_i's are tangent
vectors at x and $A = (A_{ij}) \in GL(n, \mathbb{R})$. One can construct a local trivialization by choosing a local coordinate system $(x_1, x_2, ..., x_n)$ in M. In local coordinates the vectors of a frame f can be written as $f_i = \sum f_{ij}\partial_j$. This defines a mapping $f \mapsto (f_{ij}) \in GL(n, \mathbb{R})$. The collection $(\partial_1, ..., \partial_n)$ of vector fields defines a local section of FM.

If the manifold M has some additional structure the bundle FM can generally be reduced to a subbundle. For example, if M is a Riemannian manifold with metric g, then we can define the subbundle $OFM \subset FM$ consisting of *orthonormal* frames with respect to the metric g. If in addition M is oriented, then it makes sense to speak of the bundle *SOFM of oriented orthonormal* frames: A frame $(f_1, \ldots, f_n)$ at a point x is oriented if $\o(x; f_1, \ldots, f_n)$ is positive, where $\o$ is a n form defining the orientation. The structure group of OFM is the orthogonal group $O(n)$ and of $SOFM$ the special orthogonal group $SO(n)$ consisting of orthogonal matrices with determinant=1.

Let $\mathbf{g}$ be the Lie algebra of the Lie group G. To any $A \in \mathbf{g}$ there corresponds canonically a one-parameter subgroup $h_A(t) = \exp tA$. We define a vector field $\hat{A}$ on the G bundle P such that the tangent vector $\hat{A}(p)$ at $p \in P$ is equal to $\frac{d}{dt}[p \cdot h_A(t)]|_{t=0}$. Let $g \in G$ be any fixed element. The right translation $r_g(p) = pg$ on P determines canonically a transformation $X \mapsto (r_g)_* X$ on vector fields: The tangent vector of the transformed field at a point p is simply obtained by applying the derivative of the mapping r_g to the tangent vector $X(pg^{-1})$.

Theorem 11.2 *For any $A \in \mathbf{g}$ the vector field $\hat{A}$ is equivariant, that is, $(r_g)_* \hat{A} = \widehat{ad_g^{-1} A} \forall g \in G$.*

Proof. Using a local trivialization,

$$\hat{A}(p) = \frac{d}{dt}(\pi(p), \phi(pe^{tA}))\bigg|_{t=0}$$

and therefore

$$\begin{aligned}
((r_g)_* \hat{A})(p) &= T_{pg^{-1}} r_g \cdot \frac{d}{dt}(\pi(pg^{-1}), \phi(pg^{-1}e^{tA}))|_{t=0} \\
&= \frac{d}{dt}(\pi(pg^{-1}), \phi(pg^{-1}e^{tA}g))|_{t=0} \\
&= \frac{d}{dt}(0, \phi(pe^{tad_g^{-1}A}))|_{t=0} \\
&= \widehat{ad_g^{-1} A}(p).
\end{aligned}$$

11.3 Connection and curvature in a principal bundle

Let E and M be a pair of manifolds, V a vector space and $\pi : E \to M$ a smooth surjective map.

Definition *The manifold E is a vector bundle over M with fiber V, if*

- *$E_x = \pi^{-1}(x)$ is isomorphic with the vector space V for each $x \in M$*
- *$\pi : E \to M$ is locally trivial: Any $x \in M$ has an open neighborhood U with a diffeomorphism $\phi : \pi^{-1}(U) \to U \times V$, $\phi(z) = (\pi(z), \xi(z))$, where the restriction of ξ to a fiber E_x is a linear isomorphism onto V.*

The product $M \times V$ is the *trivial vector bundle* over M, with fiber V. In this case the projection map $M \times V \to M$ is simply the projection onto the first factor.

A direct sum of two vector bundles E and F over the same manifold M is the bundle $E \oplus F$ with fiber $E_x \oplus F_x$ at a point $x \in M$. The *tensor product bundle* $E \otimes F$ is the vector bundle with fiber $E_x \otimes F_x$ at $x \in M$.

Example The tangent bundle TM of a manifold M is a vector bundle over M with fiber $T_xM \simeq \mathbb{R}^n$, where $n = dimM$. The local trivializations are given by local coordinates: If $(x_1, x_2, ..., x_n)$ are local coordinates on $U \subset M$, then the value of ξ for a tangent vector $w \in T_xM$, $x \in M$, is obtained by expanding w in the basis defined by the vector fields $(\partial_1, ..., \partial_n)$.

A section of a vector bundle E is a map $\psi : M \to E$ such that $\pi \circ \psi = id_M$. The space $\Gamma(E)$ of sections of E is a linear vector space; the addition and multiplication by scalars is defined pointwise. A principal bundle may or may not have global sections but a vector bundle always has nonzero sections. A section can be multiplied by a smooth function $f \in C^\infty(M)$ pointwise, $(f\psi)(x) = f(x)\psi(x)$.

Let (P, π, M) be a principal G bundle. *The space V of vertical vectors* in the tangent bundle TP is the subbundle of TP with fiber $\{v \in T_pP | \pi(v) = 0\}$ at $p \in P$. If P is trivial, $P = M \times G$, then the vertical subspace at $p = (x, g)$ consists of vectors tangential to G at g. In general, the dimension of the fiber V_p is equal to $\dim G$.

Definition *A connection in the principal bundle P is a smooth distribution $p \mapsto H_p$ of subspaces of T_p such that*

- *The tangent space T_p is a direct sum of V_p and H_p $\forall p \in P$*
- *The distribution is equivariant, i.e., $r_g H_p = H_{pg}$ $\forall p \in P, g \in G$.*

Smoothness means that the distribution can be locally spanned by smooth vector fields. We shall denote by pr_h (respectively, pr_v) the projection in T_p to the *horizontal* subspace H_p (respectively, vertical subspace V_p).

Let $A \in \mathbf{g}$ and let $\hat{A}$ be the corresponding equivariant vector field on P. The field $\hat{A}$ is vertical at each point. Since the group G acts freely and transitively on P, the mapping $A \mapsto \hat{A}(p)$ is a linear isomorphism onto V_p for all $p \in P$. Thus for each $X \in T_pP$ there is a uniquely defined element $\o_p(X) \in \mathbf{g}$ such that

$$\widehat{\omega_p(X)} = pr_v X$$

at p. The mapping $\omega_p : T_pP \to \mathbf{g}$ is linear, thus defining a differential form of degree one on P, with values in the Lie algebra $\mathfrak{g}$. The form ω is the *connection form* of the connection H.

Theorem 11.3 *The connection form satisfies*

- $\omega_p(\hat{A}(p)) = A \, \forall A \in \mathbf{g}$,

- $r_a^*\omega = ad_a\omega\,\forall a \in G$.

Furthermore, each **g***-valued differential form on* P *which satisfies the above conditions is a connection form of a uniquely defined connection in* P.

Proof. The first equation follows immediately from the definition of ω. To prove the second, we first note that

$$(\widehat{ad_a^{-1}A})(p) = \frac{d}{dt}pe^{tad_a^{-1}A}|_{t=0} = \frac{d}{dt}pa^{-1}e^{tA}a|_{t=0} = r_a\hat{A}(pa^{-1}).$$

By the equivariantness property of the distribution H_p, the right translations r_a commute with the horizontal and vertical projection operators. Thus

$$\begin{aligned}(ad_a\omega_p(X))\hat{}(p) &= r_a^{-1}\cdot\widehat{\omega_p(X)}(pa)\\ &= r_a^{-1}(pr_vX)(pa) = pr_v(r_a^{-1}X)(pa)\\ &= (\omega_p(r_a^{-1}X))\hat{}(pa).\end{aligned}$$

Taking account that $(r_a^*\omega)_p(X) = \omega_{pa}(r_aX)$ we get the second relation.

Let then ω be any form satisfying both equations. We define the horizontal subspaces $H_p = \{X \in T_p | \omega_p(X) = 0\}$. If $X \in H_p \cap V_p$, then $X = \hat{A}(p)$ for some $A \in \mathbf{g}$ and $\o_p(\hat{A}(p)) = A = \omega_p(X) = 0$, from which follows $H_p \cap V_p = 0$. By (1) and a simple dimensional argument we get $T_p = H_p + V_p$. For $X \in H_p$ and $a \in G$ we obtain

$$\omega_{pa}(r_aX) = (r_a^*\omega)_p(X) = ad_a\omega_p(X) = 0,$$

and therefore $r_aX \in H_{pa}$, which shows that the distribution H_p is equivariant and indeed defines a connection in P.

Let ω be a connection form in (P, π, M). Let $U \subset M$ be open and $\psi : U \to P$ a local section. The pull-back $A = \psi^*\omega$ is a one-form on U. Consider another local section $\phi : V \to P$ and set $A' = \phi^*\omega$. We can write $\psi(x) = \phi(x)g(x)$ for $g : U \cap V \to G$, where $g(x)$ is a smooth G valued function. We want to relate A to A'. Noting that

$$T_x\psi = r_{g(x)}T_x\phi + (g^{-1}T_xg)\hat{}(\phi(x))$$

by the Leibnitz rule, we get

$$\begin{aligned}A_x(u) &= \omega_{\psi(x)}(T_x\psi\cdot u) = \omega_{\psi(x)}(r_{g(x)}T_x\phi\cdot u + (g^{-1}T_xg\cdot u)\hat{}(\phi(x)))\\ &= ad_{g(x)}^{-1}\omega_{\phi(x)}(T_x\phi\cdot u) + g^{-1}T_xg\cdot u.\end{aligned}$$

For a matrix group G we can simply write

$$A = g^{-1}A'g + g^{-1}dg.$$

The transformation relating A to A' is called *a gauge transformation* . Next we define the two-form

$$F = dA + \frac{1}{2}[A, A]$$

on U. The commutator of Lie algebra valued one-forms is defined by

$$[A, B](u, v) = [A(u), B(v)] - [A(v), B(u)]$$

for a pair u, v of tangent vectors. We shall compute the effect of a gauge transformation $(U, \psi) \to (V, \phi)$ on F:

$$\begin{aligned} F &= dA + \tfrac{1}{2}[A, A] \\ &= g^{-1} dA' g - [g^{-1}dg, g^{-1}A'g] - \tfrac{1}{2}[g^{-1}dg, g^{-1}dg] \\ &\qquad + \tfrac{1}{2}[g^{-1}A'g + g^{-1}dg, g^{-1}A'g + g^{-1}dg] \\ &= g^{-1}(dA' + \tfrac{1}{2}[A', A'])g = g^{-1}F'g. \end{aligned}$$

The *curvature form* F is a pull-back under ψ of a gobally defined two-form Ω on P. The latter is defined by

$$\Omega_p(u, v) = a^{-1} F_x(\pi u, \pi v) a,$$

where $p \in \pi^{-1}(x)$, u, v tangent vectors at p and $a \in G$ is an element such that $p = \psi(x)a$. The left-hand side does not depend on the local section. Writing $p = \phi(x)a' = \psi(x)g(x)a'$ we get

$$a'^{-1} F'_x(\pi u, \pi v)a' = a'^{-1} g(x)^{-1} F_x(\pi u, \pi v) g(x) a' = a^{-1} F_x(\pi u, \pi v) a.$$

Since A is the pull-back of ω and F is the pull-back of Ω we obtain from 4.3.5

$$\Omega = d\omega + \frac{1}{2}[\omega, \omega]$$

Exercise Prove the *Bianchi identity* $dF + [A, F] = 0$. (The 3-form $[A, F]$ is defined by an antisymmetrization of $[A(u), F(v, w)]$ with respect to the triplet (u, v, w) of tangent vectors.)

Let (P, π, M) be a principal G bundle and $\rho : G \to AutV$ a linear representation of G in a vector space V. We define the manifold $P \times_G V$ to be the set of equivalence classes $P \times V / \sim$, where the equivalence relation is defined by $(p, v) \sim (pg^{-1}, \rho(g)v)$, for $g \in G$. There is a natural projection $\theta : P \times_G V \to M$, $[(p, v)] \mapsto \pi(p)$. The inverse image $\theta^{-1}(x) \cong V$, since G acts freely and transitively in the fibers of P. The linear structure in a fiber $\theta^{-1}(x)$ is defined by $[(p, v)] + [(p, w)] = [(p, v + w)]$, $l[(p, v)] = [(p, lv)]$. Local trivializations of $P \times_G V$ are obtained from local trivializations $p \mapsto (\pi(p), \phi(p)) \in M \times G$ of P by $[(p, v)] \mapsto (\pi(p), \rho(\phi(p))v)$. Thus $P \times_G V$ is a vector bundle over M, *the vector bundle associated to* P via the representation ρ of G.

Example Let $P = SU(2), M = S^2 = SU(2)/U(1), G = U(1), V = \mathbb{C}$ and $\rho(\lambda) = \lambda^2$ for $\lambda \in U(1)$. The associated vector bundle $E = SU(2) \times_{U(1)} \mathbb{C}$ is in fact the tangent bundle of the sphere S^2. The isomorphism is obtained as follows. Fix a linear isomorphism of $\mathbb{C} \cong \mathbb{R}^2$ with the tangent space of S^2 at the point x, which has as its isotropy group the given $U(1)$. The map $E \to TS^2$ is defined by $(g, v) \mapsto D(g)v$, where $D(g)$ is the 2-1 representation of $SU(2)$ in $\mathbb{R}^3$. The tangent vectors of S^2 are represented by vectors in $\mathbb{R}^3$ by the natural embedding $S^2 \subset \mathbb{R}^3$.

11.4 Parallel transport

Let H be a connection in a principal G bundle (P, π, M). *A horizontal lift* of a smooth curve $\gamma(t)$ on the base manifold M is a smooth curve $\gamma^*(t)$ on P such that the tangent vector $\dot{\gamma}^*(t)$ is *horizontal* at each point on the curve and $\pi(\gamma^*(t)) = \gamma(t)$.
Lemma *Let $X(t)$ be a smooth curve on the Lie algebra* $\mathbf{g}$ *of G, defined on an interval $[t_0, t_1]$. Then there exists a unique smooth curve $a(t)$ on G such that $\dot{a}(t)a(t)^{-1} = X(t) \forall t \in [t_0, t_1]$ and such that $a(t_0) = e$.*

Proof. See Kobayashi and Nomizu, vol. I, p. 69.

Theorem 11.4 *Let $\gamma(t)$ be a smooth curve on M and p an element in the fiber over $\gamma(t_0)$. Then there exists a unique horizontal lift $\gamma^*(t)$ of $\gamma(t)$ such that $\gamma^*(t_0) = p$.*

Proof. Choose first any (smooth) curve $\phi(t)$ on P such that $\pi(\phi) = \gamma$ and $\phi(t_0) = p$. We are looking for the solution in the form $\gamma^*(t) = \phi(t)g(t)$, where $g(t)$ is a curve on G such that $g(t_0) = e$. Now $\gamma^*(t)$ is a solution if

$$\dot{\gamma}^*(t) = r_{g(t)} \cdot \dot{\phi}(t) + (g(t)^{-1}\dot{g}(t))\hat{}[\phi(t)g(t)]$$

is horizontal. Let ø be the connection form of the connection H. A tangent vector on P is horizontal if and only if it is in the kernel of ω. We get the differential equation

$$\begin{aligned} 0 = ø(\dot{\gamma}^*(t)) &= ø(r_{g(t)}\dot{\phi}(t)) + \omega([g(t)^{-1}\dot{g}(t)]\hat{}[\phi(t)g(t)]) \\ &= ad^{-1}_{g(t)}\omega(\dot{\phi}(t)) + g(t)^{-1}\dot{g}(t). \end{aligned}$$

Applying ad_g to this equation we get

$$\dot{g}(t)g(t)^{-1} = -\omega(\dot{\phi}(t)).$$

The solution $g(t)$ exists and is unique by the previous lemma.

Example Let $P = M \times U(1)$, M simply connected. A connection form ø can be written as $\omega_{(x,g)}(u, a) = A_x(u) + g^{-1} \cdot a$, where u is a tangent vector at $x \in M$ and a is a tangent vector at $g \in U(1)$; the Lie algebra of $U(1)$ is identified by the set of purely imaginary complex numbers. Let $\gamma(t)$ be a curve on M. The horizontal lift of $\gamma(t)$ which goes through $(\gamma(t), g)$ at time $t = 0$ is $\gamma^*(t) = (\gamma(t), g(t))$ with

$$g(t) = g \cdot \exp\left(\int_0^t -A_{\gamma(s)}(\dot{\gamma}(s))ds\right).$$

In particular, for a *closed contractible curve*, $\gamma(0) = \gamma(1)$, we get by Stokes's theorem

$$g(1) = g \cdot \exp(-\int_S F),$$

where $F = dA$ is the curvature two-form and the integration is taken over any surface on M bounded by the closed curve γ.

We define *the parallel transport* along a curve $\gamma(t)$ on M as a mapping $\tau : \pi^{-1}(x_0) \to \pi^{-1}(x_1)$ ($x_0 = \gamma(t_0), x_1 = \gamma(t_1)$ points on the curve). The value $\tau(p_0)$ for $p_0 \in \pi^{-1}(x_0)$ is given as follows: Let $\gamma^*(t)$ be a horizontal lift of $\gamma(t)$ such that $\gamma^*(t_0) = p_0$. Then $\tau(p_0) = \gamma^*(t_1)$.

Exercise Prove the following properties of the parallel transport.

- $\tau \circ r_g = r_g \circ \tau \forall g \in G$
- If γ_1 is a path from x_0 to x_1 and γ_2 is a path from x_1 to x_2 then the parallel transport along the composed path $\gamma_2 * \gamma_1$ is equal to the product of parallel transport along γ_1 followed by a parallel transport along γ_2.
- The parallel transport is a one-to-one mapping between the fibers $\pi^{-1}(x_0)$ and $\pi^{-1}(x_1)$.

11.5 Covariant differentiation in vector bundles

Let E be a vector bundle over a manifold M with fiber V, dim $V = n$. The vector space V is defined over the field $\mathbb{K} = \mathbb{R}$ or $K = \mathbb{C}$. A vector bundle can always be thought of as an associated bundle to a principal bundle. Namely, let P_x denote the space of all linear frames in the fiber E_x for $x \in M$. Using the local trivializations of E it is not difficult to see that the spaces P_x fit together and form naturally a smooth manifold P. Fix a basis $w = \{w_1, \dots, w_n\}$ in E_x. Then any other basis of E_x can be obtained from w by a linear tranformation $w'_i = \sum A_{ji} w_j$ and therefore P_x can be identified with the group $GL(n, \mathbb{K})$ of all linear transformations in $\mathbb{K}^n$; it should be stressed that this identification depends on the choice of w. However, we have a well-defined mapping $P \times GL(n, \mathbb{K}) \to P$ given by the basis transformations and this shows that P can be thought of as a principal $GL(n, \mathbb{K})$ bundle over M.

The vector bundle E is now isomorphic with the associated bundle $P \times_\rho \mathbb{K}^n$, where ρ is the natural representation of $GL(n, \mathbb{K})$ in $\mathbb{K}^n$. The isomorphism is defined as follows. Let $w \in P_x$ and $a \in \mathbb{K}^n$. We set $\phi(w, a) = \sum a_i w_i$. This gives a mapping from $P \times \mathbb{K}^n$ to E which is obviously linear in a. For a fixed w the mapping $a \mapsto \phi(w, a)$ gives an isomorphism between $\mathbb{K}^n$ and E_x. Let $w' = w \cdot g$ and $a' = \rho(g^{-1})a$ for some $g \in GL(n, \mathbb{K})$. We have to show that $\phi(w', a') = \phi(w, a)$; but this follows immediately from the definitions.

Often the bundle E can be thought of as an associated bundle to a principal bundle with a smaller structure group than the group $GL(n, \mathbb{K})$. This happens when there is some extra structure in E. For example, assume there is a *fiber metric* in E: This means that there is an inner product $< \cdot, \cdot >_x$ in each fiber E_x such that $x \mapsto < \psi(x), \psi(x) >_x$ is a smooth function for any (local) section ψ. We can then define the bundle of orthonormal frames in E with structure group $U(n)$ in the complex case and $O(n)$ in the real case. The vector bundle E is now an associated bundle to the bundle of orthonormal frames.

We shall now assume that E is given as an associated vector bundle $P \times_\rho V$ to some principal bundle P, with a connection H, over M. Let G be the structure group of P. For each vector field X on M we can define a linear map ∇_X of the space $\Gamma(E)$ of sections into itself such that

- $\nabla_{X+Y} = \nabla_X + \nabla_Y$
- $\nabla_{fX} = f\nabla_X$
- $\nabla_X(f\psi) = (Xf)\psi + f\nabla_X\psi$

for all vector fields X, Y, smooth functions f and sections ψ. We shall give the definition in terms of a local trivialization $\xi : U \to P$, where $U \subset M$ is open. Locally, a section $\psi : M \to E$ can be written as

$$\psi(x) = (\xi(x), \phi(x)),$$

where $\phi : U \to V$ is some smooth function. Let A denote the pull-back $\xi^* \o$ of the connection form ω in P. The representation ρ of G in V defines also an action of the Lie algebra $\mathfrak{g}$ in V. We set

$$\nabla_X\psi = (\xi, X\phi + A(X)\phi),$$

where $A(X)$ is the Lie algebra valued function giving the value of the one-form A in the direction of the vector field X.

We have to check that our definition does not depend on the choice of the local trivialization. So let $\xi'(x) = \xi(x) \cdot g(x)$ be another local trivialization, where $g : U \to G$ is a smooth function. The vector potential with respect to the trivialization ξ' is $A' = g^{-1}Ag + g^{-1}dg$. Now $(\xi, \phi) \sim (\xi', \phi')$, where $\phi' = g^{-1}\phi$ (we simplify the notation by dropping ρ) and therefore $(\xi', X\phi' + A'(X)\phi')$ is equal to

$$\begin{aligned} & (\xi', -g^{-1}(Xg)g^{-1}\phi + g^{-1}X\phi + (g^{-1}Ag + g^{-1}Xg)g^{-1}\phi) \\ = \ & (\xi', g^{-1}(X\phi + A(X)\phi)) \sim (\xi, X\phi + A(X)\phi) \end{aligned}$$

which shows that ∇_X is well-defined.

Exercise Prove that ∇_X defined above satisfies (1)-(3).

The commutator of the *covariant derivatives* ∇_X is related to the curvature of the connection in the following way:

$$\begin{aligned}
[\nabla_X, \nabla_Y]\psi &= (\xi, [X + A(X), Y + A(Y)]\phi) \\
&= (\xi, ([X, Y] + X \cdot A(Y) - Y \cdot A(X) + [A(X), A(Y)])\phi) \\
&= (\xi, (F(X, Y) + [X, Y] + A([X, Y]))\phi)
\end{aligned}$$

where $F = dA + \frac{1}{2}[A, A]$. Thus we can write

$$[\nabla_X, \nabla_Y] - \nabla_{[X,Y]} = F(X, Y)$$

when acting on the functions ϕ.

A section ψ is *covariantly constant* if $\nabla_X \psi = 0$ for all vector fields. From the above commutator formula we conclude that one can find at each point in the base space a local basis of covariantly constant sections of the vector bundle if and only if the curvature vanishes.

11.6 An example: The monopole line bundle

Let G be a Lie group and $\mathbf{g}$ its Lie algebra. Let us denote by ℓ_g the left translation $\ell_g(a) = ga$ in G. The left invariant Maurer-Cartan form $\theta_L = g^{-1}dg$ is the $\mathbf{g}$-valued one form on G which sends a tangent vector X at $g \in G$ to the element $\ell_g^{-1} X \in T_e G$ in the Lie algebra. Similarly, we can define the right invariant Maurer-Cartan form $\theta_R = dgg^{-1}$, $\theta_R(g; X) = r_g^{-1} X$. By taking commutators, we can define higher order forms. For example, the form $[g^{-1}dg, g^{-1}dg]$ sends the pair (X, Y) of tangent vectors at g to $2[\ell_g^{-1}X, \ell_g^{-1}Y] \in \mathbf{g}$.

Taking projections to one dimensional subspaces of $\mathbf{g}$ we get real valued one-forms on G.

Let $< \cdot, \cdot >$ be a bilinear form on $\mathbf{g}$ and $\sigma \in \mathbf{g}$. Then $\alpha =< \sigma, g^{-1}dg >$ is a well-defined one form. Let us compute the exterior derivative of α. Let X, Y be a pair of left invariant vector fields on G. Now

$$\begin{aligned}
d\alpha(g; X, Y) &= X \cdot \alpha(Y) - Y \cdot \alpha(X) - \alpha([X, Y]) \\
&= -\alpha([X, Y])
\end{aligned}$$

since $\alpha(Y)(g) =< \sigma, \ell_g^{-1} Y >$ is a constant function on G and similarly for $\alpha(X)$. Since the left invariant vector fields on a Lie group span the tangent space at each point, we conclude

$$d\alpha = - < \sigma, \tfrac{1}{2}[g^{-1}dg, g^{-1}dg] > .$$

We have not yet defined the exterior derivative of a Lie algebra valued differential form, but motivated by the computation above we set

$$d(g^{-1}dg) = -\frac{1}{2}[g^{-1}dg, g^{-1}dg].$$

A bilinear form $< \cdot, \cdot >$ on $\mathbf{g}$ is *invariant* if

$$< [X,Y], Z >= - < Y, [X,Z] >$$

for all X, Y, and Z. Given an invariant bilinear form, the group G has a natural **closed** three-form c_3 which is defined by

$$c_3(g; X, Y, Z) =< \ell_g^{-1}X, [\ell_g^{-1}Y, \ell_g^{-1}Z] > .$$

Thus

$$c_3 =< g^{-1}dg, \tfrac{1}{2}[g^{-1}dg, g^{-1}dg] > .$$

Theorem 11.5 $dc_3 = 0$.

Proof. Recall the definition of the exterior differentiation d: If ω is a n-form and $V_1, ..., V_{n+1}$ are vector fields, then

$$\begin{aligned} d\omega(V_1, ..., V_{n+1}) &= \sum_{i=1}^{n+1}(-1)^{i+1}V_i \cdot \omega(V_1, ..., \hat{V}_i, ..., V_{n+1}) \\ &\quad + \sum_{i<j}(-1)^{i+j}\omega([V_i, V_j], V_1, ..., \hat{V}_i, ..., \hat{V}_j, ..., V_{n+1}), \end{aligned}$$

where the caret means that the corresponding variable has been dropped. Let us compute dc_3 for left invariant vector fields $X_1, ..., X_4$. Taking account that $c_3(X_i, X_j, X_k)$ is a constant function we get

$$\begin{aligned} dc_3(X_1, ...X_4) &= -2 < [X_1, X_2], [X_3, X_4] > +2 < [X_1, X_3], [X_2, X_4] > \\ &\quad -2 < [X_1, X_4], [X_2, X_3] > \\ &= 2 < X_1, [[X_3, X_4], X_2] - [[X_2, X_4], X_3] + [[X_2, X_3], X_4] > \\ &= 0 \end{aligned}$$

by Jacobi's identity.

If G is a group of matrices we can define an invariant form on $\mathfrak{g}$ by $< X, Y >= \operatorname{tr} XY$. Then the form c_3 can be written as

$$c_3 = \operatorname{tr}(g^{-1}dg)^3.$$

As an example we shall consider in detail the case $G = SU(2)$. Let $\sigma_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$ and define the one-form $\alpha = -\frac{1}{2}\operatorname{tr}\sigma_3 g^{-1}dg$. Remember that $SU(2) \to SU(2)/U(1) = S^2$ is a principal $U(1)$ bundle. The form α is invariant with respect to right translations $g \mapsto gh$ by $h \in U(1)$. Thus α is a connection form in the bundle $SU(2)$ [the Lie algebra of the structure group $U(1)$ can be identified with $i\mathbb{R}$]. Let us compute the curvature. The exterior derivative of α is $\frac{1}{4}\mathrm{tr}\,\sigma_3[g^{-1}dg, g^{-1}dg]$. A tangent vector at $x \in S^2$ can be represented by a tangent vector $\ell_g X$ at $g \in \pi^{-1}(x)\,, X \in \mathbf{g}$, such that X is orthogonal to the $U(1)$ direction, $\mathrm{tr}\,\sigma_3 X = 0$. The curvature in the base space S^2 is then $\Omega(X,Y) = \frac{1}{2}\mathrm{tr}\,\sigma_3[X,Y]$. The form Ω is $\frac{1}{2}\times$ the volume form on S^2: If {X,Y} is an orthonormal system at $x \in S^2$, then $[X,Y] = \pm\frac{i}{2}\sigma_3$ (exercise), the sign depending on the orientation. We obtain $\Omega(X,Y) = \pm\frac{i}{4}\mathrm{tr}\,\sigma_3^2 = \pm\frac{i}{2}$.

The basic monopole line bundle is defined as the associated bundle to the bundle $SU(2) \to S^2$, constructed using the natural one dimensional representation of $U(1)$ in $\mathbb{C}$.

Embedding $S^2 \subset \mathbb{R}^3$ and using Cartesian coordinates $\{x_1, x_2, x_3\}$ we can write the curvature form as

$$\Omega = \frac{1}{4r^3}\varepsilon^{ijk} x_i dx_j \wedge dx_k,$$

where $r^2 = x_1^2 + x_2^2 + x_3^2$ is equal to 1 on S^2. However, we can extend Ω to the space $\mathbb{R}^3 \setminus \{0\}$ using the above formula. The coefficients of the linearly independent forms $dx_2 \wedge dx_3$, $dx_3 \wedge dx_1$ and $dx_1 \wedge dx_2$ form a vector $\vec{B} = \frac{1}{2r^3}(x_1, x_2, x_3) = \frac{\vec{x}}{r^3}$. The field $\vec{B}$ satisfies

$$\vec{\nabla} \cdot \vec{B} = 0$$

$$\vec{\nabla} \times \vec{B} = 0,$$

i.e., it satisfies Maxwell's equations in vacuum. On the other hand,

$$\int_{S^2} \vec{B} \cdot d\vec{S} = 2\pi$$

for *any* sphere containing the origin. Because of these properties, the field $\vec{B}$ can be interpreted as the magnetic field of a magnetic monopole located at the origin. The integral (3) multiplied by the dimensional constant $1/e$ (e is the unit electric charge) is called the *monopole strength.*

11.7 Yang-Mills equations

Let M be a Riemann manifold with Riemann metric g. In local coordinates the metric is represented as a symmetric nondegenerate tensor field $g_{ij}(x)$ with $i, j = 1, 2, \ldots, n$, where $n = \dim M$. Let $\pi : P \to M$ be a principal G bundle over M. Let $\rho : G \to Aut(V)$ be a unitary finite-dimensional representation of G in V. This defines an associated vector bundle $E = P \times_\rho V$ and the curvature tensor F of a connection in P is represented (locally) by matrix functions $F_{ij}(x) = \partial_i A_j - \partial_j A_i + [A_i, A_j]$ acting on vectors in V.

We shall define raising and lowering of space-time indices (i.e., coordinate indices in M) as usual, $A^i = g^{ij} A_j, B_i = g_{ij} B^j$, where the matrix (g^{ij}) is the inverse of (g_{ij}).

Recall also that the metric g defines a volume form on M, $d(vol_M) = \sqrt{\det(g)}dx_1 \wedge dx_2 \cdots \wedge dx_n$. We define the *Yang-Mills functional*

$$Y(A) = \frac{1}{4}\int_M \text{tr } F_{\mu\nu}F^{\mu\nu}d(vol_M).$$

The Yang-Mills action is invariant under gauge transformations $F' = g^{-1}Fg$. There is an alternative way to write the YM action as

$$Y(A) = -\frac{1}{2}\int_M \text{tr } F \wedge *F.$$

The action leads to field equations through Euler-Lagrange variational priciple. Let $A + tB$ be a 1-parameter family of vector potentials:

$$\frac{d}{dt}Y(A+tB)|_{t=0} = \frac{1}{2}\int_M \text{tr } F^{\mu\nu}(\partial_\mu B_\nu - \partial_\nu B_\mu + [A_\mu, B_\nu] + [B_\mu, A_\nu])d(vol_M).$$

When M is a manifold without boundary, we can integrate by parts and we get

$$\delta Y(A) = -\int_M \text{tr } B_\nu(\partial_\mu F^{\mu\nu} + [A_\mu, F^{\mu\nu}])d(vol_M).$$

If A is an extremal the YM action then we obtain *the Yang-Mills equations*

$$D_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + [A_\mu, F^{\mu\nu}] = 0.$$

When G is abelian this gives the Maxwell's equations $\partial_\mu F^{\mu\nu} = 0$ in vacuum. In addition, we have the Bianchi identities

$$D_\mu F_{\nu\lambda} + D_\lambda F_{\mu\nu} + D_\nu F_{\lambda\mu} = 0$$

for all indices λ, μ, ν. If there are external sources we have instead

$$D_\mu F^{\mu\nu} = j^\nu$$

for some Lie algebra valued current j^ν.

The Yang-Mills equations is a complicated nonlinear system of second order partial differential equations. Not much is known about the general solutions. However, there is a class of solutions which is well understood. These so-called (anti) instantons are characterized by the (anti) self-duality property $F = *F$ $(F = - * F)$ in the case of a Riemannian 4-manifold M. Recall that

$$* : \Omega^k(M) \to \Omega^{n-k}(M)$$

is a linear map and $** = \pm 1$. When $n = 4$ and $k = 2$ the sign is + (exeercise) For this reason the eigenvalues of $*$ are ± 1, when restricted to 2-forms on a 4-manifold.

In the case of Lorentzian metric $** = -1$ on 2-forms and therefore in this case there are no real eigenvalues (and no real (anti) instantons).

In the case of an instanton we have

$$Y(A) = -\frac{1}{2}\int_M \mathrm{tr}\ F \wedge F$$

and so the value of the YM functional is given by the second Chern class. In particular, when $M = S^4$ we get

$$Y(A) \sim \int_{S^3} \mathrm{tr}\ (g^{-1}dg)^3,$$

where $g : S^3 \to G$ is the transition function on the equator. Thus for self-dual solutions the YM functional is quantized in units $(2\pi)^2 k$ with $k \in \mathbb{Z}$.