Representatie, homotopie en ijktheorie van SU2(R)

Javan Peykar, A.

A. Javanpeykar

Representatie, homotopie en ijktheorie van $\mathrm{SU}_2(\mathbf{R})$

Bachelorscriptie, 11 juli 2008

Scriptiebegeleider (wiskunde): dr. F. Doray

Scriptie begeleider (natuurkunde): dr. K. Schalm

Mathematisch Instituut, Universiteit Leiden

Introduction

Group theory is a very interesting branch of mathematics whereas differential geometry is also very interesting. Therefore it is natural to ask where the combination of these two branches leads us. This question is answered easily: We get Lie groups!

Lie groups are not only mathematically rich of structure, but their applications reach from modern theoretical physics to algebraic geometry. For example, in theoretical physics one can define gauge theories with the aid of the well known representation theory of compact Lie groups. The standard model is an example of a $U_1(\mathbf{R}) \times \mathrm{SU}_2(\mathbf{R}) \times \mathrm{SU}_3(\mathbf{R})$ gauge theory[1]. In algebraic geometry one can look at the generalization of Lie groups which are algebraic groups.

After a brief review of categories we will begin by showing that Lie groups and so called Lie algebras are closely related by basically looking close enough to the identity element, i.e., linearizing the Lie group close to the identity element. This is done by looking at the tangent space at the identity element. This precise relation is actually best expressed using category theory. Using this relation between Lie groups and Lie algebras we will be able to completely capture all finite dimensional representations of Lie groups such as the special linear group $\mathrm{SL}_2(\mathbf{C})$ and $\mathrm{SU}_2(\mathbf{R})$. We finish the mathematical part of this thesis by asking ourselves how homotopy theory of Lie groups can tell us something about the structure and even about the representations of certain Lie groups. For example, the fundamental group of the rotation group $\mathrm{SO}_3(\mathbf{R})$ is closely related to its representations. Now, in the physical part of this thesis we give an exposition of Witten's global anomaly in $\mathrm{SU}_2(\mathbf{R})$ gauge theory. To this extent we shall briefly review some basic aspects of gauge theories which are also known as Yang-Mills theories. The representation theory of $\mathrm{SU}_2(\mathbf{R})$ will seem very useful here.

The reader will find that the bridge between mathematics and physics is captured in the chapter on the fourth homotopy group of $\mathrm{SU}_2(\mathbf{R})$. The result of this chapter was used by E. Witten to prove that certain $\mathrm{SU}_2(\mathbf{R})$ gauge theories contain global anomalies. The reasoning used by E. Witten uses some advanced mathematical theorems which we shall only quote. One of them is the celebrated Atiyah-Singer index theorem.

This thesis is organized as follows.

Chapter 1: This chapter gives a very brief introduction to category theory. The reader will find Yoneda's Lemma and a characterization of group objects with the aid of Yoneda's Lemma.

Chapter 2: We introduce the notion of homotopy and study the fundamental group as a functor. The tangent space of a real manifold is defined as the vector space of derivations from the local ring of the manifold to $\mathbf{R}$. We give the connection with the tangent space of a submanifold of $\mathbf{R}^n$ so that it will become easier to compute Lie algebras of sub Lie groups of $\mathrm{GL}_n(\mathbf{R})$. Eventually we will be capable of stating the equivalence of categories between (some subcategory of) the category of Lie groups and the category of Lie algebras.

Chapter 3: Representations of $\mathrm{SL}_2(\mathbf{C})$, $\mathrm{SU}_2(\mathbf{R})$ and $\mathrm{SO}_3(\mathbf{R})$ are studied. The complex representations of $\mathrm{SU}_2(\mathbf{R})$ are deduced by using those of $\mathrm{SL}_2(\mathbf{C})$ and the equivalence of categories mentioned above. The quaternion algebra of Hamilton will be used when studying the complex representations of $\mathrm{SO}_3(\mathbf{R})$.

[1] In Dutch we say "ijktheorie"

Chapter 4: We devote this chapter to the Hopf map which we obtain as a morphism of $\mathbf{S}^3$ to the sphere $\mathbf{S}^2$. This map is the generator of $\pi_3(\mathbf{S}^2) = \mathbf{Z}$ and its image under the suspension map generates $\pi_4(\mathbf{S}^3) = \mathbf{Z}/2\mathbf{Z}$.

Chapter 5: Some basics of quantum electrodynamics (QED) as a $U_1(\mathbf{R})$ gauge theory and its generalizations are discussed in detail. We will introduce the Grassmann algebra in order to make up for the noncommutative nature of fermions. Integration of Grassmann numbers is introduced and there we shall deduce an important identity used by Witten in his original article. The reader will also find a description of Weyl and Dirac fermions.

Chapter 6: We study the original article by Edward Witten which concerns a global anomaly in $\mathrm{SU}_2(\mathbf{R})$ gauge theory. We fill in some technical details in the original proofs and note on some generalizations of Witten's anomaly.

Contents

CHAPTER 1

Categories

In this chapter we briefly recall some basics of category theory without paying attention to logical problems one might encounter. For a more detailed treatment of categories we refer to [**Mac**].

DEFINITION 1.1. We define a category $\mathcal{C}$ as a set of *objects* $\mathrm{Ob}(\mathcal{C})$, a set of *morphisms* $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ for any two objects $X, Y \in \mathrm{Ob}(\mathcal{C})$, and for any $X_1, X_2, X_3 \in \mathrm{Ob}(\mathcal{C})$ a map

$$\circ : \mathrm{Hom}_{\mathcal{C}}(X_1,X_2) \times \mathrm{Hom}_{\mathcal{C}}(X_2,X_3) \to \mathrm{Hom}_{\mathcal{C}}(X_1,X_3), \text{ denoted } (g,f) \mapsto g \circ f$$

which satisfies the following two conditions:

Associativity: $h \circ (g \circ f) = (h \circ g) \circ f$ whenever it makes sense,
Identity morphism: For every object $X \in \mathrm{Ob}(\mathcal{C})$ there is a morphism $\mathrm{id}_X \in \mathrm{Hom}_{\mathcal{C}}(X,X)$ such that $\mathrm{id}_X \circ f = f$ and $g \circ \mathrm{id}_X = g$ whenever it makes sense.

DEFINITION 1.2. Given a category $\mathcal{C}$ we define its *opposite* category $\mathcal{C}^{op}$ as follows:

(1) $\mathrm{Ob}(\mathcal{C}^{op}) = \mathrm{Ob}(\mathcal{C})$,
(2) $\forall X_1, X_2 \in \mathrm{Ob}(\mathcal{C}^{op})$, $\mathrm{Hom}_{\mathcal{C}^{op}}(X_1,X_2) = \mathrm{Hom}_{\mathcal{C}}(X_2,X_1)$.

NOTATION 1.3. We fix the notations for some common categories we will use throughout this thesis.

The category of sets is denoted by $\mathfrak{Set}$, the category of groups is denoted by $\mathfrak{Grp}$, the category of topological spaces is denoted by $\mathfrak{Top}$ and the category of pointed topological spaces is denoted by $\mathfrak{Top}_*$.

DEFINITION 1.4. Let $\mathcal{C}$ be a category and let $f : X \to Y$ a morphism in $\mathcal{C}$. We say that f is an *isomorphism (in $\mathcal{C}$)*, often denoted by $X \simeq Y$, when there exists a morphism $g \in \mathrm{Hom}_{\mathcal{C}}(Y,X)$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. In case f is an isomorphism the morphism g is unique and usually denoted by f^{-1}. We say that f is an *automorphism* of X if f is an isomorphism from X to X.

DEFINITION 1.5. **(Functor)** A *functor*, or *morphism of categories*, F from a category $\mathcal{C}$ to a category $\mathcal{D}$, denoted by $F : \mathcal{C} \to \mathcal{D}$, consists of a map $\mathrm{Ob}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{D})$, denoted by $X \mapsto F(X)$, and for all $X, Y \in \mathrm{Ob}(\mathcal{C})$ a map $\mathrm{Hom}_{\mathcal{C}}(X,Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X),F(Y))$, still denoted by $f \mapsto F(f)$, satisfying $F(g \circ f) = F(g) \circ F(f)$.

EXAMPLE 1.6. Let $\mathcal{C}$ be a category and let $X \in \mathrm{Ob}(\mathcal{C})$ be an object of $\mathcal{C}$. Then we define the functor $\mathrm{Hom}_{\mathcal{C}}(X,-) : \mathcal{C} \to \mathfrak{Set}$ which takes any object $Y \in \mathrm{Ob}(\mathcal{C})$ to $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ and any morphisms $f : Z \to Z'$ between objects $Z, Z' \in \mathrm{Ob}(\mathcal{C})$ to the morphism of sets $\mathrm{Hom}_{\mathcal{C}}(X,f) : \mathrm{Hom}_{\mathcal{C}}(X,Z) \to \mathrm{Hom}_{\mathcal{C}}(X,Z')$ given by $g \mapsto f \circ g$. Another example is given by the functor $U : \mathfrak{Grp} \to \mathfrak{Set}$ which assigns to each group G its underlying set UG and to each morphism $f : G \to H$ of groups the same morphism, regarded just as a morphism of sets. We call U a *forgetful functor* since it forgets some of the structure on the objects.

DEFINITION 1.7. **(Natural Transformations)** Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. Given two functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{C} \to \mathcal{D}$ a *natural transformation*, or *morphism of functors*, $\alpha : F \to G$ consists for all objects $X \in \mathcal{C}$ of a morphism $\alpha(X) : F(X) \to G(X)$ in $\mathcal{D}$ such that for any morphism $f : X_1 \to X_2$ in $\mathcal{C}$, $\alpha(X_2) \circ F(f) = G(f) \circ \alpha(X_1)$.

PROPOSITION 1.8. For any two categories $\mathcal{C}$ and $\mathcal{D}$ the set of functors from $\mathcal{C}$ to $\mathcal{D}$ forms a category, denoted by $\mathcal{D}^{\mathcal{C}}$, whose morphisms are natural transformations. $\square$

DEFINITION 1.9. Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. Let $F : \mathcal{C} \to \mathcal{D}$ be a functor. We say that F *is an equivalence of categories* if there exists a functor $G : \mathcal{D} \to \mathcal{C}$ such that the composition FG is isomorphic to $\mathrm{id}_{\mathcal{D}}$ in $\mathcal{D}^{\mathcal{D}}$ and the composition GF is isomorphic to $\mathrm{id}_{\mathcal{C}}$ in $\mathcal{C}^{\mathcal{C}}$.

NOTATION 1.10. Let $\mathcal{C}$ be a category. We denote the category $(\mathfrak{Set})^{\mathcal{C}^{op}}$ of functors $\mathcal{C}^{op} \to \mathfrak{Set}$ by $\mathcal{C}^{\vee}$. The category $(\mathfrak{Set}^{op})^{\mathcal{C}^{op}}$ is denoted by $\mathcal{C}^{\wedge}$. Note that $\mathcal{C}^{\wedge}$ is the opposite category of $\mathfrak{Set}^{\mathcal{C}}$.

NOTATION 1.11. Given a category $\mathcal{C}$ let $h^{\vee}_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}^{\vee}$ be the functor given by $h^{\vee}_{\mathcal{C}}(X) = \mathrm{Hom}_{\mathcal{C}}(-, X)$ and let $h^{\wedge}_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}^{\wedge}$ be the functor given by $h^{\wedge}_{\mathcal{C}}(X) = \mathrm{Hom}_{\mathcal{C}}(X, -)$.

PROPOSITION 1.12. **(Yoneda's Lemma)** Let $\mathcal{C}$ be a category. Then for any $F \in \mathcal{C}^{\vee}$ and $X \in \mathcal{C}$, $\mathrm{Hom}_{\mathcal{C}^{\vee}}(h^{\vee}_{\mathcal{C}}(X), F) \simeq F(X)$. Also, for any $F \in \mathcal{C}^{\wedge}$ and $X \in \mathcal{C}$, $\mathrm{Hom}_{\mathcal{C}^{\wedge}}(F, h^{\wedge}_{\mathcal{C}}(X)) \simeq F(X)$. Consequently, for any two objects X and Y in $\mathcal{C}$, $\mathrm{Hom}_{\mathcal{C}^{\vee}}(h^{\vee}_{\mathcal{C}}(X), h^{\vee}_{\mathcal{C}}(Y) \simeq h^{\vee}_{\mathcal{C}}(Y)(X) = \mathrm{Hom}_{\mathcal{C}}(X, Y)$, and $\mathrm{Hom}_{\mathcal{C}^{\wedge}}(h^{\wedge}_{\mathcal{C}}(X), h^{\wedge}_{\mathcal{C}}(Y) \simeq h^{\wedge}_{\mathcal{C}}(Y)(X) = \mathrm{Hom}_{\mathcal{C}}(X, Y)$.

PROOF. [**Dor**, Proposition 2.4.9, pages 64-65]. $\square$

DEFINITIONS 1.13. Let $\mathcal{C}$ be a category. The *product* of two objects X_1 and X_2 in $\mathcal{C}$ is a triple (X_3, p_1, p_2) where X_3 is an object of $\mathcal{C}$ and $p_1 : X_3 \to X_1$ and $p_2 : X_3 \to X_2$ are two morphisms such that for any object Z and any two morphisms $f_1 : Z \to X_1$ and $f_2 : Z \to X_2$ there exists a unique morphism $f : Z \to X_3$ such that $p_1 \circ f = f_1$ and $p_2 \circ f = f_2$. Note that if the product of two objects exists it is unique up to a unique isomorphism. The *coproduct* of two objects Y_1 and Y_2 in $\mathcal{C}$ is the product of Y_1 and Y_2 in the opposite category $\mathcal{C}^{op}$. An object $X \in \mathcal{C}$ is called a *terminal object* if $\forall Y \in \mathcal{C}$, $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a set with one element. A *group object* of $\mathcal{C}$ is a pair (G, π) where G is an object in $\mathcal{C}$ and $\pi : \mathcal{C}^{op} \to \mathfrak{Grp}$ is a functor whose composite with the forgetful functor $\mathfrak{Grp} \to \mathfrak{Set}$ equals $h^{\vee}_{\mathcal{C}}(G)$. A *cogroup object* of $\mathcal{C}$ is a pair (G, ϖ) such that (G, ϖ) is a group object of $\mathcal{C}^{op}$.

Let $\mathcal{C}$ be a category with a terminal object, denoted by $\{pt\}$, and suppose that for any two objects X_1 and X_2 in $\mathcal{C}$ their product exists. By Yoneda's Lemma we have the following Proposition.

PROPOSITION 1.14. To give a group object (G, π) in $\mathcal{C}$ is to give three morphisms in $\mathcal{C}$, $m_G : G \times G \to G$ (multiplication), $i_G : G \to G$ (inversion) and $e : \{pt\} \to G$ (identity). Denoting the product of G with G by $(G \times G, p_1, p_2)$ these maps are required to satisfy $m_G \circ (e_G \times \mathrm{id}_G) = \mathrm{id}_G$ (left identity), $(\mathrm{id}_G \times e_G) \circ m_G = \mathrm{id}_G$ (right identity), $m_G \circ (m_G \times \mathrm{id}_G) = (\mathrm{id}_G \times m_G) \times m_G$ (group associativity) and such that the following diagrams (left and right inverse)

$$\begin{tikzcd}
G \arrow[r, "\Delta"] \arrow[d, "\exists!"'] & G \times G \arrow[r, "i_G \times \mathrm{id}_G"] & G \times G \arrow[d, "m_G"] \\
\{pt\} \arrow[rr, "e_G"'] & & G
\end{tikzcd}$$

and

$$\begin{tikzcd}
G \arrow[r, "\Delta"] \arrow[d, "\exists!"'] & G \times G \arrow[r, "\mathrm{id}_G \times i_G"] & G \times G \arrow[d, "m_G"] \\
\{pt\} \arrow[rr, "e_G"'] & & G
\end{tikzcd}$$

are commutative. Here $\Delta : G \to G \times G$ is the unique morphism such that $p_i \circ \Delta = \mathrm{id}_G$ for $i = 1$ and $i = 2$. $\square$

DEFINITION 1.15. A morphism of two group objects (G, m_G, i_G, e_G) and (H, m_H, i_H, e_H) of $\mathcal{C}$ is a morphism $f : G \to H$ in $\mathcal{C}$ such that $f \circ m_G = m_H \circ (f \times f)$.

CHAPTER 2

Lie groups and Lie algebras

1. Sub Lie groups of $\mathrm{GL}_n(\mathbf{R})$

We shall always assume n to be some positive integer.

By a (real) differentiable map we shall always mean a $\mathcal{C}^\infty$ map, i.e all derivatives exist and are continuous. By a complex analytic map we shall always mean a map which is locally given by a convergent power series.

Notation 2.1. Let F be a field and m be some integer. The F-vector space of all $(m \times n)$-matrices with coordinates in F is denoted by $textrmM_{m\times n}(F)$. The F-algebra of all $(n \times n)$-matrices with coordinates in F is denoted by $\mathrm{M}_n(F)$. The group of invertible elements in $\mathrm{M}_n(F)$ is denoted by $\mathrm{GL}_n(F)$. Let $U \subset \mathbf{R}^n$ be an open subset and $f = (f_1, \ldots, f_m) : U \to \mathbf{R}^m$ be a differentiable map. Then the *Jacobian (matrix)* of f at $x \in U$ is

$$Df(x) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} \in \mathrm{M}_{m\times n}(\mathbf{R}).$$

Definition 2.2. Let k be a nonnegative integer. A subset $X \subset \mathbf{R}^n$ is called a *k-dimensional submanifold* of $\mathbf{R}^n$ if for every $p \in X$ there exists an open subset $U_p \subset \mathbf{R}^n$ with $p \in U_p$, and a differentiable map $f = (f_1, \ldots, f_{n-k}) : U_p \to \mathbf{R}^{n-k}$ such that $f^{-1}(\{0\}) = X \cap U_p$ and $\mathrm{rk}\, Df(p) = n - k$. Here $\mathrm{rk}\, Df(p)$ denotes the rank of the Jacobian of f at p. Note that a subset $X \subset \mathbf{R}^n$ is a k-dimensional submanifold of $\mathbf{R}^n$ if and only if for every $p \in X$ there exists an open neighborhood U of p and an open subset $V \subset \mathbf{R}^{n-k}$ such that U and V are diffeomorphic.

Examples 2.3. The 0-dimensional submanifolds of $\mathbf{R}^n$ are precisely the discrete subsets and the n-dimensional submanifolds of $\mathbf{R}^n$ are precisely the open subsets. The unit sphere $\mathbf{S}^{n-1} = \{(x_1, \ldots, x_n) \in \mathbf{R}^n \mid x_1^2 + \ldots + x_{n+1}^2 = 1\} \subset \mathbf{R}^n$ is an $(n-1)$-dimensional submanifold of $\mathbf{R}^n$ since it is the zero locus of the differentiable map $f : \mathbf{R}^n \to \mathbf{R}$ defined by $f(x_1, \ldots, x_n) = x_1^2 + x_2^2 + \ldots + x_n^2 - 1$.

Let $X \subset \mathbf{R}^n$ be a k-dimensional submanifold, and $p \in X$.

Definition 2.4. A *curve* in X through p is a differentiable map $\gamma : U \to \mathbf{R}^n$, with $U \subset \mathbf{R}$ an open subset containing 0, $\gamma(U) \subset X$ and $\gamma(0) = p$.

Remark 2.5. Since a curve γ in X through p is differentiable as a map to $\mathbf{R}^n$, $\dot\gamma(0) = \frac{d\gamma}{dt}(0) \in \mathbf{R}^n$ is well-defined.

Definition 2.6. The set $T_pX := \{\ \dot\gamma(0) \mid \gamma \text{ a curve in } X \text{ through } p\ \}$ is called the *tangent space (to X at p)*.

Proposition 2.7. Let f be as in Definition 2.2. Then $T_pX = \ker Df(p)$. In particular, T_pX is a k-dimensional linear subspace of $\mathbf{R}^n$ and $\ker Df(p)$ is independent of the choice of f.

Proof. Note that the rank condition on f implies that $\ker Df(p)$ is a k-dimensional linear subspace of $\mathbf{R}^n$. We show that $\mathbf{R}^k \subset T_pX$. For any w in $\mathbf{R}^k$ we define a curve γ in X through p by $\gamma(t) = p + tw$ with t in some small enough open neighborhood of 0. It is clear that $\dot\gamma(0) = w$.

We show that $T_pX \subset \ker Df(p)$. Let $\gamma : U \to \mathbf{R}^n$ be a curve in X through p and let $\dot\gamma(0)$ be in T_pX. Let U_p be as in Definition 2.2. By choosing U small enough we may assume $\gamma(U) \subset X \cap U_p$. Then $f \circ \gamma \equiv 0$, and hence, by the chain rule, $0 = \frac{d}{dt}(f \circ \gamma(t))|_{t=0} = Df(p)(\dot\gamma(0))$. This implies that $\dot\gamma(0) \in \ker Df(p)$. This proves that $T_pX = \ker Df(p)$. □

Examples 2.8. By Example 2.3, $\mathbf{S}^n$ is an n-dimensional submanifold of $\mathbf{R}^{n+1}$. The tangent space to $\mathbf{S}^n$ at $x = (x_1, \ldots, x_{n+1}) \in \mathbf{S}^n$ is given by $\{ y = (y_1, \ldots, y_{n+1}) \in \mathbf{R}^{n+1} \mid x_1y_1 + \ldots + x_{n+1}y_{n+1} = 0 \}$. The tangent space to an open subset $U \subset \mathbf{R}^n$ at any element of U is easily seen to be equal to $\mathbf{R}^n$.

Notation 2.9. Let $G = \mathrm{GL}_n(\mathbf{R})$. Let I_n denote the $(n \times n)$-matrix. For a matrix A we denote the transpose of A by A^t, the complex conjugate of A by $\overline{A}$ and the hermitian conjugate of A by $A^* = \overline{A}^t = \overline{A^t}$. A matrix is called anti-symmetric when $A^t = -A$ and anti-hermitian when $A^* = -A$.

Remark 2.10. Note that the determinant map is polynomial and hence continuous. Hence $G = M_n(\mathbf{R}) \setminus \det^{-1}\{0\}$ is open in $\mathrm{M}_n(\mathbf{R})$ and the tangent space to G at any element of G may be identified with $\mathrm{M}_n(\mathbf{R})$.

Definition 2.11. A subgroup of G which is also closed in G and a submanifold of G is called a sub Lie group of G.

Proposition 2.12. The orthogonal group $\mathrm{O}_n(\mathbf{R}) = \{ A \in G \mid AA^t = I_n \}$ is an $\frac{n^2-n}{2}$-dimensional sub Lie group of G and its tangent space at the identity element is the $\mathbf{R}$-vector space of anti-symmetric matrices in $\mathrm{M}_n(\mathbf{R})$.

Proof. Let $S_n(\mathbf{R}) = \{ A \in M_n(\mathbf{R}) \mid A = A^t \}$ be the vector space of symmetric $(n \times n)$-matrices with coordinates in $\mathbf{R}$. This is identified with $\mathbf{R}^{\frac{n^2+n}{2}}$ by taking the coefficients on and above the diagonal as coordinates. Note that $\mathrm{O}_n(\mathbf{R})$ is the zero set of the differentiable map $f : G \to S_n(\mathbf{R})$ given by $A \mapsto AA^t - I_n$. By looking at the defining equations for elements of $\mathrm{O}_n(\mathbf{R})$ it is clear that $\mathrm{O}_n(\mathbf{R})$ is closed and even compact in G. For every A in G, the Jacobian $Df(A) : \mathrm{M}_n(\mathbf{R}) \to T_{f(A)}S_n(\mathbf{R}) = S_n(\mathbf{R})$ is given by $Df(A)(B) = AB^t + BA^t$. To prove this we take a real number $t > 0$ small enough such that $A + tB$ is in G. This is possible since we may compute the product $(A+tB)(I_n - tA^{-1}B) = A + O(t^2)$ in $M_n(\mathbf{R})$ to see that $A + tB$ is invertible for t small enough. We see that

$$
\begin{aligned}
\frac{d}{dt}(f(A+tB))\,|_{t=0} &= \frac{d}{dt}(AA^t + t(AB^t + BA^t) + t^2(BB^t - I_n)\,|_{t=0} \\
&= (AB^t + BA^t + 2tBB^t)\,|_{t=0} \\
&= AB^t + BA^t.
\end{aligned}
$$

Take now A in $\mathrm{O}_n(\mathbf{R})$ and C in $S_n(\mathbf{R})$, and define $B := \frac{1}{2}CA \in M_n(\mathbf{R})$. Then $B^t = \frac{1}{2}A^tC$, and we get $Df(A)(B) = \frac{1}{2}(AA^tC + CAA^t) = C$. This shows that $Df(A)$ is surjective for all A in $\mathrm{O}_n(\mathbf{R})$. This implies that $\mathrm{rk}\, Df(A) = \dim(S_n(\mathbf{R})) = \frac{n^2+n}{2}$. By the definition of a submanifold (Definition 2.2) we may conclude that $\mathrm{O}_n(\mathbf{R})$ is a submanifold of $M_n(\mathbf{R})$ of dimension $n^2 - \frac{n^2+n}{2} = \frac{n^2-n}{2}$. The last part follows by $\ker Df(e) = \{ X \in \mathrm{M}_n(\mathbf{R}) \mid X^t = -X \}$ and Proposition 2.7. □

Proposition 2.13. Let k be $\mathbf{R}$ or $\mathbf{C}$.

(1) The special linear group $\mathrm{SL}_n(k) = \{ A \in \mathrm{GL}_n(k) \mid \det(A) = 1 \}$ is an (n^2-1)-dimensional sub Lie group of $\mathrm{GL}_n(k)$ and the tangent space to $\mathrm{SL}_n(k)$ at the identity element is the k-vector space of $(n \times n)$-matrices with coordinates in k and trace zero.
(2) The special orthogonal group $SO_n(\mathbf{R}) = \{ A \in \mathrm{O}_n(\mathbf{R}) \mid \det(A) = 1 \}$ is a compact $\frac{n^2-n}{2}$-dimensional sub Lie group of G and the tangent space to $SO_n(\mathbf{R})$ at the identity element is isomorphic to the $\mathbf{R}$-vector space of antisymmetric $(n \times n)$-matrices with trace zero.

(3) The unitary group $\mathrm{U}_n(\mathbf{R}) = \{ A \in \mathrm{GL}_n(\mathbf{C}) \mid AA^* = I_n \}$ is a n^2-dimensional real sub Lie group of $\mathrm{GL}_n(\mathbf{C})$ and the tangent space to $\mathrm{U}_n(\mathbf{R})$ at the identity element is the $\mathbf{R}$-vector space of complex anti-hermitian $(n \times n)$-matrices.
(4) The special unitary group $SU_n(\mathbf{R}) = \{ A \in U_n(\mathbf{R}) \mid \det(A) = 1 \}$ is a compact real (n^2-1)-dimensional sub Lie group of $\mathrm{GL}_n(\mathbf{C})$ and the tangent space to $\mathrm{SU}_n(\mathbf{R})$ at the identity element is isomorphic to the $\mathbf{R}$-vector space of complex anti-hermitian $(n \times n)$-matrices with trace zero.

PROOF. The proofs are similar to the proof of Proposition 2.12. □

2. Definitions of Lie groups and Lie algebras

DEFINITIONS 2.14. Let X be a topological space and let n be a nonnegative integer. An n-dimensional real (resp. complex) *atlas* for X consists of the following data: a set I, for each i an open subset X_i of X, an open subset U_i of $\mathbf{R}^n$ (resp. $\mathbf{C}^n$), and a homeomorphism $\phi_i : U_i \to X_i$. These data are required to satisfy the following conditions: First of all, the X_i cover X, i.e. $X = \bigcup_i X_i$. Secondly, the *charts* ϕ_i are compatible in the following sense. For i and j let $X_{i,j}$ be $X_i \cap X_j$, and let $U_{i,j}$ be $\phi_i^{-1}(X_{i,j})$. Then we require $U_{i,j}$ to be open in U_i, $U_{j,i}$ to be open in U_j and the bijection $\phi_j^{-1} \circ \phi_i|_{U_{i,j}} : U_{i,j} \to U_{j,i}$ to be differentiable (resp. complex analytic).

A *real (resp. complex) manifold* is a topological space X, separated and with a countable basis for the topology, equipped with a real (resp. complex) atlas.

Let (X, I, n, U, ϕ) and (Y, J, m, V, ψ) be two real (resp. complex) manifolds. Let f be a continuous map from X to Y. Let x be in X. Then f is called *differentiable at* x if for every (i, j)[1] such that $x \in X_i$ and $f(x) \in Y_j$ the map $\psi_j^{-1} f \phi_i$ from $\phi_i^{-1}((f^{-1}(Y_j)) \cap X_i) \subset \mathbf{R}^n$ (resp. $\mathbf{C}^n$) to $\mathbf{R}^m$ (resp. $\mathbf{C}^m$) is differentiable (resp. complex analytic) at $\phi_i^{-1}(x)$. The map f is called *differentiable*, or a morphism of real (resp. complex) manifolds, if it is differentiable at all x in X.

REMARK 2.15. Since the composition of morphisms of manifolds is associative and for each manifold X the identity map id_X acts as the identity morphism we have the category of manifolds.

DEFINITIONS 2.16. A (real) Lie group is a group G equipped with the structure of a real (resp. complex) manifold, such that the two maps

$$G \times G \to G, \qquad (x, y) \mapsto xy$$

and

$$G \to G, \qquad x \mapsto x^{-1}$$

are morphisms of real manifolds.

A (complex) Lie group is a group G equipped with the structure of a real (resp. complex) manifold, such that the two maps above are morphisms of real manifolds.

A morphism of Lie groups is a morphism of manifolds which is also a morphism of groups.

EXAMPLE 2.17. Let k be $\mathbf{R}$ or $\mathbf{C}$. Let V be a finite dimensional k-vector space. Let $\mathrm{End}_k(V)$, or just $\mathrm{End}(V)$, be the k-algebra of endomorphisms of V endowed with its natural topology and differentiable structure from k (or $k^{\dim V}$). We already noted in Remark 2.10 that the group of invertible endomorphisms of V, denoted by $\mathrm{GL}(V)$, is an open subset of $\mathrm{End}(V)$. This gives $\mathrm{GL}(V)$ the structure of a manifold. Since the maps

$$\mathrm{GL}(V) \times \mathrm{GL}(V) \to \mathrm{GL}(V), \qquad (x, y) \mapsto xy$$

and

$$\mathrm{GL}(V) \to \mathrm{GL}(V), \quad x \mapsto x^{-1}$$

are polynomial they are morphisms of manifolds. We conclude that $\mathrm{GL}(V)$ is a Lie group.

[1]We can also require there to be only one pair (i, j). The definition does not change in that case.

PROPOSITION 2.18. Any submanifold of $\mathbf{R}^n$ has the structure of a real differentiable manifold. In particular, any sub Lie group of $\mathrm{GL}_n(\mathbf{R})$ is a Lie group.

PROOF. [**Lke**, Theorem 1.1.4, pages 7-10]. □

DEFINITIONS 2.19. Let k be a field. Let $\mathfrak{g}$ be a k-vector space. If $\mathfrak{g}$ is equipped with a k-bilinear map $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfying

(1) $[\cdot,\cdot]$ is *alternating*: $[x,x] = 0$ for all x in $\mathfrak{g}$,
(2) $[\cdot,\cdot]$ satisfies the *Jacobi identity*: $[x,[y,z]] + [z,[x,y]] = -[y,[z,x]]$ for all x, y and z in $\mathfrak{g}$,

we call $\mathfrak{g}$ a *k-Lie algebra*. We call the map $[\cdot,\cdot]$ the *Lie bracket* on $\mathfrak{g}$. If $\mathfrak{g}$ and $\mathfrak{h}$ are two Lie algebras, then a morphism of Lie algebras from $\mathfrak{g}$ to $\mathfrak{h}$ is a k-linear map $f : \mathfrak{g} \to \mathfrak{h}$ such that for all x and y in $\mathfrak{g}$ one has $[f(x), f(y)] = f([x,y])$.

EXAMPLE 2.20. Let $(A, +, \cdot)$ be an associative algebra. Then $(A, +)$ has the structure of a Lie algebra when we define $[x,y] = x \cdot y - y \cdot x = xy - yx$ for any x and y in A. It is easy to see that the commutator $[\cdot,\cdot]$ on A is a Lie bracket on A. Since bilinearity and alternativity are easy we only check the Jacobi identity. For any x, y and z in A,

$$\begin{aligned}
[x,[y,z]] + [z,[x,y]] &= x(yz - zy) - (yz - zy)x + z(xy - yx) - (xy - yx)z \\
&= xyz - xzy - yzx + zyx + zxy - zyx - xyz + yxz \\
&= zxy - xzy - yzx + yxz = (zx - xz)y - y(zx - xz) = -[y,[z,x]].
\end{aligned}$$

EXAMPLE 2.21. Let k be a field and V a k-vector space. By Example 2.20, the underlying vector space of the k-algebra $\mathrm{End}_k(V)$ has the structure of a k-Lie algebra. The Lie bracket is the commutator $[\cdot,\cdot]$ given by $[f,g] = f \circ g - g \circ f$.

EXAMPLE 2.22. Let k be a field. Let A be a k-algebra[2] and let M be an A-module. A *derivation* D (over k) of A to M is a k-linear morphism of A to M such that the *Leibniz formula*

$$D(ab) = D(a)b + a(Db)$$

holds for any a and b in A. The set $\mathrm{Der}_k(A,M)$ of derivations of A to M is clearly a k-linear subspace of the k-vector space $\mathrm{Hom}_k(A,M)$. Let M be A. The commutator $[D_1, D_2] = D_1 \circ D_2 - D_2 \circ D_1$ of two derivations of A (into itself) is a derivation since for any a and b in A

$$(D_1D_2)(ab) = D_1(aD_2b + (D_2a)b) = aD_1D_2b + (D_1a)(D_2b) + (D_2a)(D_1b) + (D_1D_2a)b.$$

Hence

$$\begin{aligned}
[D_1, D_2](ab) &= (D_1D_2 - D_2D_1)(ab) = (D_1D_2)(ab) - (D_2D_1)(ab) \\
&= aD_1D_2b - aD_2D_1b + (D_1D_2a)b - (D_2D_1a)b \\
&= ([D_1, D_2]a)b - a[D_1, D_2]b.
\end{aligned}$$

We conclude that $\mathrm{Der}_k(A,A)$ is a k-Lie algebra; it is called the *derivation algebra* of A and it is usually denoted by $\mathrm{Der}(A)$.

REMARK 2.23. If $\mathfrak{g}$ is any Lie algebra, with basis $(x_1, \ldots, x_n)$ the entire multiplication table[3] of $\mathfrak{g}$ can be recovered from the *structure constants* c_{ij}^k which occur in the expressions $[x_i, x_j] = \sum_{k=1}^n c_{ij}^k x_k$. Those for which $i \geq j$ can even be deduced from the others due to the Jacobi identity and the fact that $[x,x] = 0$ for any x in $\mathfrak{g}$. Turning this remark around, it is possible to define a Lie algebra by specifying a set of structure constants $\{c_{ij}^k\}$ which satisfy

$$c_{ii}^k = 0, \quad \sum_k (c_{ij}^k c_{kl}^m + c_{jl}^k c_{ki}^m + c_{li}^k c_{kj}^m) = 0.$$

In practice we shall never construct a Lie algebra in this way, but it shows us that it is enough to specify the basis of a Lie algebra and the Lie brackets of the basis elements to completely determine the structure of a Lie algebra.

---

[2] A is not necessarily associative.

[3] The multiplication is of course the Lie bracket.

3. Coverings and the fundamental group functor

Let I denote the closed unit interval $[0,1] \subset \mathbf{R}$.

DEFINITIONS 2.24. **(Homotopy)** Two continuous maps $f, g : X \to Y$ between two topological spaces X and Y are said to be *homotopy equivalent*, denoted by $f \sim_H g$, if and only if there exists a continuous map $F : X \times I \to Y$ such that $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$ for all points $x \in X$. We say that F is a *homotopy from f to g.*

We say that two morphisms f and g between pointed topological spaces are *homotopic* if there exists a base point preserving homotopy from f to g.

PROPOSITION 2.25. Let X and Y be two topological spaces. Homotopy equivalence induces an equivalence relation on the set of continuous functions between X and Y. Similarly, homotopy equivalence induces an equivalence relation on the set of continuous maps between two pointed topological spaces (X, x_0) and (Y, y_0). Furthermore, let X, Y and Z be topological spaces. If we have maps $f, g : X \to Y$ and $h : Y \to Z$ such that there is a homotopy F from f to g, then $h \circ F$ is a homotopy from hf to hg. If we have maps $f : X \to Y$ and $g, h : Y \to Z$ such that there is a homotopy G from g to h then $g \circ f$ and $h \circ f$ are homotopic through the homotopy $F(x,t) = G(f(x), t)$. We conclude that homotopy behaves well with respect to composition of maps.

PROOF. [**Arms**, Lemma 5.2, 5.3 and 5.4, pages 100-101]. □

Let $\mathbf{S}^1 = \{\ x = (x_1, x_2) \in \mathbf{R}^2 \mid x_1^2 + x_2^2 = 1\ \}$ be the unit circle in $\mathbf{R}^2$. Choose the base point to be $s = (1,0) \in \mathbf{R}^2$. We may parameterize $\mathbf{S}^1$ by $x = \cos\theta$ and $y = \sin\theta$ with $0 \le \theta \le 2\pi$.

DEFINITIONS 2.26. Let (X, x) be a pointed topological space. A morphism of pointed topological spaces from $(\mathbf{S}^1, s)$ to (X, x) is called a *loop (in X based at x)*. Given two loops α and β we define their product $\alpha \cdot \beta : (\mathbf{S}^1, s) \to (X, x)$ by

$$(\alpha \cdot \beta)(\theta) = \begin{cases} \alpha(2\theta) & 0 \le \theta \le \pi \\ \beta(2\theta - 2\pi) & \pi \le \theta \le 2\pi \end{cases}$$

Let $\langle\alpha\rangle$ denote the homotopy class of a loop α and let $\pi_1(X, x)$ denote the set of homotopy classes of loops in X based at x. The product on $\mathrm{Hom}_{\mathfrak{Top}_*}((\mathbf{S}^1, s), (X, x))$ induces a multiplication on $\pi_1(X, x)$ defined by $\langle\alpha\rangle \cdot \langle\beta\rangle = \langle\alpha \cdot \beta\rangle$. This multiplication is well-defined and associative. The homotopy class of the constant loop e at x_0 given by $e(\theta) = x_0$ for all $\theta \in [0, 2\pi]$ defines the identity element of this multiplication. Let γ be a loop in X based at x_0. The inverse of the homotopy class $\langle\gamma\rangle \in \pi_1(X, x_0)$ is defined by $\langle\gamma^{-1}\rangle$ where $\gamma^{-1}(\theta) = \gamma(2\pi - \theta)$ for all $\theta \in [0, 2\pi]$. The identity element and the inverse element are well-defined. We can check that $\langle e\rangle \cdot \langle\alpha\rangle = \langle\alpha\rangle = \langle\alpha\rangle \cdot \langle e\rangle$ and $\langle\gamma\rangle \cdot \langle\gamma^{-1}\rangle = \langle\gamma^{-1}\rangle \cdot \langle\gamma\rangle = \langle e\rangle$ for all $\langle\alpha\rangle$ and $\langle\gamma\rangle$ in $\pi_1(X, x_0)$. This gives $\pi_1(X, x)$ the structure of a group. Any morphism of pointed topological spaces $f : (X, x) \to (Y, y)$ induces a morphism of groups $f_* : \pi_1(X, x) \to \pi_1(Y, y)$ given by $f_*(\langle\alpha\rangle) = \langle f \circ \alpha\rangle$. Note that for any two loops α and β,

$$f_*(\langle\alpha\rangle \cdot \langle\beta\rangle) \quad = \quad f_*(\langle\alpha \cdot \beta\rangle) = \langle f \circ (\alpha \cdot \beta)\rangle = \langle f \circ \alpha\rangle \cdot \langle f \circ \beta\rangle = f_*(\langle\alpha\rangle) \cdot f_*(\langle\beta\rangle).$$

DEFINITION 2.27. Define the *homotopy category (with base points)*, denoted by $\mathfrak{HTop}_*$, as follows:

(1) $\mathrm{Ob}(\mathfrak{HTop}_*) = \mathrm{Ob}(\mathfrak{Top}_*)$, see Notation 1.3,
(2) For any two pointed topological spaces (X, x) and (Y, y),

$$\mathrm{Hom}_{\mathfrak{HTop}_*}((X,x),(Y,y)) = {}^{\mathrm{Hom}_{\mathfrak{Top}_*}((X,x),(Y,y))}\big/_{\sim_H}.$$

NOTATION 2.28. Let $h = \mathrm{Hom}_{\mathfrak{HTop}_*}((\mathbf{S}^1, s), -) : \mathfrak{HTop}_* \to \mathfrak{Set}$. Let $U : \mathfrak{Grp} \to \mathfrak{Set}$ be the forgetful functor defined in Examples 1.6.

DEFINITION 2.29. The fundamental group gives a functor $\pi_1 : \mathfrak{HTop}_* \to \mathfrak{Grp}$ such that the composition of functors $U \circ \pi_1 = h$. We call π_1 the *fundamental group functor*. Note that $((\mathbf{S}^1, s), \pi_1)$ is a cogroup object of $\mathfrak{HTop}_*$.

Remark 2.30. Let ${}^{op} : \mathfrak{Grp} \to \mathfrak{Grp}$ be the functor assigning to each group G its opposite group G^{op} and to each morphism $f : G_1 \to G_2$ the morphism of groups $f^{op} : G_1^{op} \to G_2^{op}$ defined by $f^{op}(g) = f(g^{-1})$. The morphism of groups $\alpha_G : G \to G^{op}$ given by $\alpha_G(g) = g^{-1}$ is an isomorphism. It gives an isomorphism of functors $\alpha : \mathrm{id}_{\mathfrak{Grp}} \to {}^{op}$.

Definition 2.31. The *opposite fundamental group functor* is the functor $\pi_1^{op} : \mathfrak{HTop}_* \to \mathfrak{Grp}$ defined by $\pi_1^{op} = {}^{op} \circ \pi_1$. Note that $((\mathbf{S}^1, x), \pi_1^{op})$ is a group object of $\mathfrak{HTop}_*^{op}$ and that $\pi_1^{op} = {}^{op} \circ \pi_1 \simeq \mathrm{id}_{\mathfrak{Grp}} \circ \pi_1 = \pi_1$ as functors.

Proposition 2.32. Let $((\mathbf{S}^1, s), \varpi)$ be a cogroup object of $\mathfrak{HTop}_*$. Then ϖ is isomorphic to π_1 or π_1^{op} as functors from $\mathfrak{HTop}_*$ to $\mathfrak{Grp}$.

Proof. By Proposition 1.14, to give a group object $((\mathbf{S}^1, s), \varpi)$ is equivalent to give a group structure on $(\mathbf{S}^1, s)$ in the opposite category of $\mathfrak{HTop}_*$, i.e. a *cogroup* structure on $(\mathbf{S}^1, s)$ in $\mathfrak{HTop}_*$. But by [**Argu**, Proposition 7.1, page 1679], there are only two cogroup structures on $\mathbf{S}^1$. This implies that $\mathfrak{HTop}_*$ has only two cogroup objects. We already gave these two: π_1 and π_1^{op}. $\square$

Remark 2.33. It is interesting to remark that the following diagram of functors (which speaks for itself)

$$\begin{tikzcd} \pi_1 \arrow[r, "\sim"] \arrow[d] & \pi_1^{op} \arrow[dl] \\ h & \end{tikzcd}$$

is not commutative.

Proposition 2.34. Let X be a path-connected topological space. Let $p \in X$ and $q \in X$ be points. Then $\pi_1(X, p)$ and $\pi_1(X, q)$ are isomorphic as groups. Therefore, for path-connected spaces X we may speak of *the fundamental group* $\pi_1(X)$. Also, for two path-connected topological spaces X and Y the fundamental group $\pi_1(X \times Y)$ is isomorphic to $\pi_1(X) \times \pi_1(Y)$.

Proof. [**Arms**, Theorem 5.6 and Theorem 5.14, pages 94 and 101]. $\square$

Definition 2.35. If X and Y are two manifolds, a *covering map* is a morphism $p : Y \to X$ of manifolds with the property that each point of X has an open neighborhood N such that $p^{-1}(N)$ is a disjoint union of open sets, each of which is mapped diffeomorphically by p onto N. Such a covering map p is called a *covering of* X and Y is called a *covering space of* X. For manifolds with a selected base point we require the covering map to preserve base points. A morphism of coverings $p : Y \to X$ and $q : Z \to X$ of a manifold X is a morphism $\varphi : Y \to Z$ of manifolds such that $q \circ \varphi = p$, i.e. such that the following diagram

$$\begin{tikzcd} Y \arrow[r, "\varphi"] \arrow[dr, "p"'] & Z \arrow[d, "q"] \\ & X \end{tikzcd}$$

is commutative.

Proposition 2.36. Any locally path-connected and connected topological space X is path-connected. As a consequence, any connected manifold as path-connected.

Proof. For any x in X, let U_x be the set of points in X such that there exists a path in X from x to that point. It suffices to show that $X = U_x$. Since X is locally path-connected, U_x is open. Note that the complement of U_x equals $\bigcup_{z \in X \setminus U_x} U_z$. Since this is an union of open subsets we conclude that U_x is closed. Since x is in U_x, U_x is nonempty. Since X is connected, $U_x = X$. Therefore X is path-connected. $\square$

THEOREM 2.37. **(Unique Lifting Property)** Let $p : (X, x_0) \to (Y, y_0)$ be a covering of two connected manifolds which preserves the base points. Let (Z, z_0) be a connected pointed manifold and $f : (Z, z_0) \to (Y, y_0)$ a morphism of connected manifolds which preserves the base points such that $f_*(\pi_1(Z)) \subset p_*(\pi_1(X))$. Then there exists a unique base point preserving morphism $\widetilde{f} : (Z, z_0) \to (X, x_0)$ of connected manifolds such that $p \circ \widetilde{f} = f$, i.e. such that the following diagram

$$\begin{tikzcd} & (X, x_0) \arrow[d, "p"] \\ (Z, z_0) \arrow[ur, dotted, "\widetilde{f}"] \arrow[r, "f"] & (Y, y_0) \end{tikzcd}$$

is commutative.

PROOF. [**Ful**, Proposition 13.5, pages 180-181]. □

DEFINITION 2.38. A path-connected topological space X is called *simply connected* if its fundamental group is $\{0\}$.

THEOREM 2.39. Let n be a positive integer. Let $\mathbf{S}^n := \{ (x_1, \dots, x_{n+1}) \in \mathbf{R}^{n+1} \mid x_1^2 + \dots + x_{n+1}^2 = 1 \}$ be the unit n-dimensional sphere in $\mathbf{R}^{n+1}$. Then $\mathbf{S}^n$ is path-connected for all n and simply connected for all $n \geq 2$. For $n = 1$, $\pi_1(\mathbf{S}^1) = \mathbf{Z}$.

PROOF. [**Arms**, Theorem 5.8 and Theorem 5.12, pages 96-99]. □

DEFINITION 2.40. **(Universal Covering)** Let X be a connected manifold. A covering $\widetilde{X} \to X$ is called a *universal covering* of X if $\widetilde{X}$ is a connected and simply connected manifold.

THEOREM 2.41. If X is a connected manifold then X has a universal covering space. The universal covering of a manifold is unique up to isomorphism. If we select a base point then the isomorphism is unique.

PROOF. For the existence of a universal covering see [**Ful**, Theorem 13.20, page 188-189]. The uniqueness is a direct consequence of the Unique Lifting Property, i.e. Theorem 2.37. □

EXAMPLE 2.42. Let $\mathbf{S}^1$ be the unit circle in $\mathbf{C}^*$, with the induced group structure. Then $\mathbf{R}$ is the universal covering of $\mathbf{S}^1$ and the covering map $\mathbf{R} \to \mathbf{S}^1$ is given by $x \mapsto e^{2\pi i x}$. For details we refer to [**Arms**, Section 5.3, pages 96-102].

4. Universal covering Lie groups

THEOREM 2.43. Let G be a connected Lie group. There exists a unique (up to an isomorphism of Lie groups) Lie group $\widetilde{G}$ such that $\widetilde{G}$ is a universal covering of G and the covering map $\widetilde{G} \to G$ is a morphism of Lie groups. From now on, for every Lie group G, $\widetilde{G}$ will denote this universal covering which we call the *universal covering Lie group of G.*

PROOF. Let G be a connected Lie group with identity element $e \in G$. By Theorem 2.41, G has a unique covering $p : \widetilde{G} \to G$ with $\widetilde{G}$ simply connected. Let $f : \widetilde{G} \times \widetilde{G} \to G$ be the morphism of manifolds defined by $f(\widetilde{\sigma}, \widetilde{\tau}) = p(\widetilde{\sigma})p(\widetilde{\tau})^{-1}$. Choose $\widetilde{e} \in p^{-1}(\{e\})$ as a base point for $\widetilde{G}$. By Proposition 2.34, $\widetilde{G} \times \widetilde{G}$ is simply connected. By Theorem 2.37, a unique mapping $\widetilde{f} : \widetilde{G} \times \widetilde{G} \to \widetilde{G}$ exists such that $p \circ \widetilde{f} = f$ and such that $\widetilde{f}(\widetilde{e}, \widetilde{e}) = e$. For $\widetilde{\sigma}$ and $\widetilde{\tau}$ in $\widetilde{G}$ we define two operations on $\widetilde{G}$:

$$\widetilde{\tau}^{-1} = \widetilde{f}(\widetilde{e}, \widetilde{\tau})$$

and

$$\widetilde{\sigma}\widetilde{\tau} = \widetilde{f}(\widetilde{\sigma}, \widetilde{\tau}^{-1}).$$

Consider the three maps $\widetilde{\sigma} \mapsto \widetilde{\sigma}\widetilde{e}$, $\widetilde{\sigma} \mapsto \widetilde{e}\widetilde{\sigma}$, and $\widetilde{\sigma} \mapsto \widetilde{\sigma}$ of $\widetilde{G}$ to itself. They all make the following diagram

$$\begin{tikzcd} & \widetilde{G} \arrow[d, "p"] \\ \widetilde{G} \arrow[ur, dotted] \arrow[r, "p"] & G \end{tikzcd}$$

commute, and they all send $\widetilde{e}$ to $\widetilde{e}$. So by the Unique Lifting Property (Theorem 2.37) we may conclude that the three maps $\widetilde{\sigma} \mapsto \widetilde{\sigma}\widetilde{e}$, $\widetilde{\sigma} \mapsto \widetilde{e}\widetilde{\sigma}$, and $\mathrm{id}_{\widetilde{G}} : \widetilde{\sigma} \mapsto \widetilde{\sigma}$ of $\widetilde{G} \to \widetilde{G}$ are identical. Hence $\widetilde{\sigma}\widetilde{e} = \widetilde{e}\widetilde{\sigma} = \widetilde{e}$ for all $\widetilde{\sigma}$ in $\widetilde{G}$. Similarly, it follows that $\widetilde{\sigma}\widetilde{\sigma}^{-1} = \widetilde{\sigma}^{-1}\widetilde{\sigma} = \widetilde{e}$ and that $(\widetilde{\sigma}\widetilde{\tau})\widetilde{\gamma} = \widetilde{\sigma}(\widetilde{\tau}\widetilde{\gamma})$ for all $\widetilde{\sigma}, \widetilde{\tau}, \widetilde{\gamma} \in \widetilde{G}$. Since $\widetilde{f}$ is a morphism of manifolds we conclude that $\widetilde{G}$ is a Lie group with the constructed group structure. Since $p(\widetilde{\tau}^{-1}) = p(\widetilde{\tau})^{-1}$ and $p(\widetilde{\sigma}\widetilde{\tau}) = p(\widetilde{\sigma})p(\widetilde{\tau})$ for any $\widetilde{\sigma}$ and $\widetilde{\tau}$ in $\widetilde{G}$ we conclude that $p : \widetilde{G} \to G$ is a morphism of Lie groups. By construction $\widetilde{G}$ is unique up to a unique isomorphism of Lie groups. □

Definitions 2.44. Let G be a group acting (from the left) on a manifold X. The action of G on X is said to be *differentiable* if the map from X to itself given by $x \mapsto gx$ is a diffeomorphism for any g in G. We say that G acts *evenly* on X if each point $x \in X$ has an open neighborhood U which satisfies $U \cap g(U) = \emptyset$ for all $g \in G - \{e\}$. If G acts evenly on X and U is an open neighborhood for x in X such that $U \cap g(U) = \emptyset$ for all $g \in G - \{e\}$ we say that U is an *even* neighborhood for x. The *orbit* of an element x in X, denoted by $G \cdot x$, is defined to be the set $\{\ y \in X \mid gx = y \text{ for some } g \text{ in } G\ \}$. The *orbit space*, denoted by X/G, is the set of orbits. There is a canonical projection $p : X \to X/G$ that maps a point to its orbit.

Remark 2.45. Let G be a group acting evenly on a manifold X and assume the action to be differentiable. The canonical projection $p : X \to X/G$ is continuous and open when the orbit space X/G is given the quotient topology. Furthermore, we can give X/G a differentiable structure such that the projection $p : X \to X/G$ is a local diffeomorphism.

Lemma 2.46. Let G act evenly on a manifold X and assume the action to be differentiable. Then the canonical projection $p : X \to X/G$ is a covering map.

Proof. Given $y \in X/G$, we choose a point $x \in p^{-1}(y)$ and an even neighborhood U of x in X. The quotient map p identifies all disjoint sets $\{\ g(U) \mid g \in G\ \}$ to a single open set $p(U)$. By the definition of the differentiable structure on X/G, p restricts to a diffeomorphism from $g(U)$ onto $p(U)$ for each g in G. □

Theorem 2.47. If a group G acts evenly on a connected and simply connected manifold X, then $\pi_1(X/G)$ is isomorphic to G.

Proof. Let $p : X \to X/G$ be the canonical projection. Fix a point $x_0 \in X$. Since X is path connected we can, given $g \in G$, join x_0 to gx_0 by a path γ. Note that $p \circ \gamma$ is a loop in X/G based at $p(x_0)$. Let $\phi : G \to \pi_1(X/G, p(x_0))$ be the map defined by $\phi(g) = \langle p \circ \gamma \rangle$. Since X is simply connected ϕ is well-defined.

For g_1 and g_2 in G join x_0 to g_1x_0 by a path γ_1 and x_0 to g_2x_0 by a path γ_2. Let $l_{g_1} : x \mapsto g_1x$ denote the left multiplication by g_1 on X. Note that $\gamma_1 \cdot (l_{g_1} \circ \gamma_2)$ is a path joining x_0 to $g_1g_2x_0$. This implies that ϕ is a morphism of groups.

By Lemma 2.46, the projection $p : X \to X/G$ is a covering map of connected manifolds. Let $\langle \alpha \rangle \in \pi_1(X/G, p(x_0))$ be the class of a loop α in X/G based at $p(x_0)$. By Theorem 2.37, there is a path γ in X which begins at x_0 and satisfies $p \circ \gamma = \alpha$. The endpoint $\gamma(1)$ lies in the orbit of x_0, so there is an element g in G such that $gx_0 = \gamma(1)$. By construction $\phi(g) = \langle \alpha \rangle$. This shows that ϕ is surjective. Let g be an element of G in the kernel of ϕ, i.e. assume that the loop $p \circ \gamma$, with γ a path joining x_0 to gx_0, is homotopic to the constant loop in X/G based at $p(x_0)$. By Theorem 2.37, there is a homotopy from γ to the constant loop in X based at x_0. By the uniqueness part of Theorem 2.37, γ must be the constant loop. This implies that $gx_0 = x_0$. Since G acts evenly on X we may conclude that g is the identity element of G. This shows that ϕ is injective. □

PROPOSITION 2.48. Let n be a a positive integer and let $\mathbf{Z}/2\mathbf{Z} = \{1, -1\}$ act on the unit n-dimensional sphere $\mathbf{S}^n$ by $-1 \cdot x = -x$ for any x in $\mathbf{S}^n$. By definition the orbit space $\mathbf{S}^n/(\mathbf{Z}/2\mathbf{Z})$ is the n-dimensional real projective space $\mathbf{P}^n_{\mathbf{R}}$. The quotient map $\mathbf{S}^n \to \mathbf{P}^n_{\mathbf{R}}$ is a double covering, i.e. every fiber of the covering map has 2 elements, of the n-dimensional projective space over $\mathbf{R}$. Moreover, for any $n \geq 2$ the fundamental group $\pi_1(\mathbf{P}^n_{\mathbf{R}})$ is isomorphic to $\mathbf{Z}/2\mathbf{Z}$.

PROOF. Let n be a positive integer. For any point x of $\mathbf{S}^n$ take U to be a hemisphere of $\mathbf{S}^n$ containing x. This is clearly an even neighborhood of x. By Lemma 2.46, the quotient map $\mathbf{S}^n \to \mathbf{P}^n_{\mathbf{R}}$ is a double covering. Since $\mathbf{S}^n$ is connected and simply connected for $n \geq 2$ the last part follows by applying Theorem 2.47. □

PROPOSITION 2.49. The Lie groups $\mathrm{SL}_n(\mathbf{C})$, $\mathrm{GL}_n(\mathbf{C})$, $\mathrm{SO}_n(\mathbf{R})$, $\mathrm{U}_n(\mathbf{R})$ and $\mathrm{SU}_n(\mathbf{R})$ are connected. The groups $\mathrm{O}_n(\mathbf{R})$ and $\mathrm{GL}_n(\mathbf{R})$ consist of two connected components distinguished by the sign of the determinant.

PROOF. [**Onis**, Proposition 4.4 of Chapter 1, page 21]. □

5. The Lie algebra of a Lie group

Let k be $\mathbf{R}$ or $\mathbf{C}$. Let X be a manifold over k, and let $p \in X$.

DEFINITIONS 2.50. We denote by $\mathcal{O}(X)$ the k-algebra of all differentiable maps (Definitions 2.14) from X to k. Two maps f and g defined on open sets of X containing p are said to have the same *germ* at p if they agree on some neighborhood of p. This introduces an equivalence relation on the set of differentiable maps defined on open neighborhoods of p, two differentiable functions being equivalent if and only if they have the same germ. The equivalence classes are called *germs*, and we denote the set of germs at p by $\mathcal{O}_{p,X}$. If f is a differentiable function on a neighborhood of p, then $\mathbf{f}$ will denote its germ. The operations of addition, scalar multiplication, and multiplication of functions induce on $\mathcal{O}_{p,X}$ the structure of a k-algebra. A germ $\mathbf{f}$ has a well-defined value $\mathbf{f}(p)$ at p, namely, the value at p of any representative of the germ. Note that $\mathcal{O}_{p,X}$ is a local ring. Its (unique) maximal ideal, denoted by m_p, is the set of germs vanishing at p. The residue field $\mathcal{O}_{p,X}/m_p$ is isomorphic to k. Since m_p is an $\mathcal{O}_{p,X}$-module, m_p/m_p^2 is an $\mathcal{O}_{p,X}/m_p$-module, i.e. a k-vector space.

DEFINITION 2.51. The *tangent space* to X at p, denoted by T_pX, is defined to be the k-vectorspace $\mathrm{Der}_k(\mathcal{O}_{p,X}, k)$ (Example 2.22). Elements of T_pX are called *tangent vectors (to X at p).*

PROPOSITION 2.52. The tangent space T_pX and $(m_p/m_p^2)^*$ are naturally isomorphic as k-vector spaces. (The symbol * denotes dual vector space.)

PROOF. If $v \in T_pX$, it is clear that v is a linear function on m_p vanishing on m_p^2. Conversely, for l in $(m_p/m_p^2)^*$, we define a tangent vector v_l at p by setting $v_l(\mathbf{f}) = l([\mathbf{f} - \mathbf{f(m)}])$. Here $\mathbf{f(m)}$ denotes the germ of the function with the constant value $\mathbf{f}(m)$, and [ ] is used to denote cosets in m_p/m_p^2. It is easy to see that v_l is a derivation and that the obtained mappings are inverse to each other. □

PROPOSITION 2.53. It holds that $\dim\, T_pX = \dim\, X$.

PROOF. Since the question is local, we may suppose that X is an open subset of $\mathbf{R}^n$ for some positive integer n and $p = 0 \subset \mathbf{R}^n$. For $i = 1, \ldots, n$ define ∂_i to be the element of $T_0X = \mathrm{Der}_k(\mathcal{O}_{0,X}, k)$ that sends $\mathbf{f}$ (or just f) to its ith partial derivative at 0 and let x_j be the jth coordinate function from X to $\mathbf{R}$. The elements ∂_i are linearly independent because of the relations $\partial_i(x_j) = \delta_{i,j}$. Here $\delta_{i,j}$ denotes the Kronecker symbol. We show that $\partial = \sum_{i=1}^{n} \partial(x_i)\partial_i$ for any ∂ in T_0X. To prove the identity, let f be in $\mathcal{O}_{0,X}$. Rewriting the identity

$$\int_0^1 \left(\frac{d}{dt} f(tx)\right) dt = f(x) - f(0),$$

with x in some neighborhood of 0, gives

$$f(x) = f(0) + \sum_{i=1}^{n} \partial_i(f)x_i + \sum_{i=1}^{n} x_i g_i(x),$$

where the g_i are differentiable, and $\partial(x_i g_i) = 0$ for each i. The identity $\partial = \sum_{i=1}^{n} \partial(x_i)\partial_i$ follows by applying δ to f. This completes the proof. $\square$

REMARK 2.54. We will treat tangent vectors as operating on functions rather than on their germs. If f is a differentiable function defined on a neighborhood of p, and $v \in T_pX$, we define $v(f) = v(\mathbf{f})$.

DEFINITION 2.55. Let $f : X \to Y$ be a morphism of manifolds, and let $p \in X$. Let f^* denote the induced morphism $f^* : \mathcal{O}(Y) \to \mathcal{O}(X)$ given by $f^*(g) = g \circ f$. The *tangent map (of f at p)*, denoted by T_pf, is the linear map $T_pf : T_pX \to T_{f(p)}Y$ defined by $T_pf(v) : g \mapsto v(f^*(g))$ for any v in T_pX. For $g : Y \to X$ a morphism of manifolds, $T_p(g \circ f) = T_pg \circ T_pf$.

REMARK 2.56. We have a functor from the category of manifolds (with base points) to the category of vector spaces. To each manifold we assign the tangent space to the manifold at the base point and to each morphism of manifolds we assign the tangent map at the base point.

REMARK 2.57. Let X be a submanifold of $\mathbf{R}^n$, $p \in X$ and f be as in Definition 2.2. Then the map

$$\ker Df(p) \to \mathrm{Der}(\mathcal{O}_{p,X}, \mathbf{R}), \quad x \mapsto \Big[g \mapsto g'(p) \cdot x\Big]$$

is a natural isomorphism. Here $g'(p)$ denotes the (well-defined) derivative $\frac{d}{dt}g(p)$. Consequently, the notions of tangent space for a submanifold and manifold coincide.

REMARK 2.58. Let G be a group (resp. Lie group). For g in G the left multiplication by g, denoted by $l_g : G \to G$, is a group (resp. Lie group) automorphism of G.

EXAMPLE 2.59. For g in $\mathrm{GL}_n(k)$ the tangent map of l_g at the identity element e of $\mathrm{GL}_n(k)$ is given by left multiplication on $\mathrm{M}_n(k)$, i.e. $T_el_g : a \mapsto ga$.

Let G be a Lie group.

LEMMA 2.60. The group of Lie group automorphisms $\mathrm{Aut}(G)$ of G acts (linearly) on the derivation algebra $\mathrm{Der}(\mathcal{O}(G))$ (Example 2.22): for f in $\mathrm{Aut}(G)$ and D in $\mathrm{Der}(\mathcal{O}(G))$, $fD = (f^{-1})^* \circ D \circ f^*$. Consequently, the set $\{\ D \in \mathrm{Der}(\mathcal{O}(G)) \mid l_f(D) = D \text{ for all } f \text{ in } \mathrm{Aut}(G)\ \}$ is a sub Lie algebra of $\mathrm{Der}(\mathcal{O}(G))$.

PROOF. This is an easy verification. $\square$

DEFINITION 2.61. A derivation D on $\mathcal{O}(G)$ is called *left invariant* if $g \cdot D := l_gD = D$ for all g in G.

PROPOSITION 2.62. Let $\mathfrak{g}$ denote the k-Lie algebra of left invariant derivations on $\mathcal{O}(G)$. Let T_eG be the tangent space to G at the identity. Then the map $\alpha : \mathfrak{g} \to T_eG$ defined by $\alpha(D) : f \mapsto (D(f))e$ for any $D \in \mathfrak{g}$ is an isomorphism of k-vector spaces. Consequently, dim $\mathfrak{g}$ = dim T_eG = dim G and $T_el_g \circ \alpha : \mathfrak{g} \to T_gG$ is an isomorphism of k-vector spaces for any g in G.

PROOF. It is clear that α is linear. We show that α is injective. Suppose that $D \in \mathfrak{g}$ is in the kernel of α, i.e. $D(f)(e) = 0$ for any morphism f in $\mathcal{O}(G)$. In view of the left invariance of G, for any morphism f in $\mathcal{O}(G)$ and any g in G,

$$(D(f))g = ((g \cdot D)(f))g = (D(f \circ l_g) \circ l_{g^{-1}})g = (D(f \circ l_g))e = 0.$$

This implies that $D = 0$ and that α is injective. We show that α is surjective. Let v be in T_eG. Let $D_v : \mathcal{O}(G) \to \mathcal{O}(G)$ be defined by $D_v(f) : g \mapsto \Big((T_e l_g)(v)\Big)f$ for any f in $\mathcal{O}(G)$. Then $\alpha(D_v) = v$. It remains to show that D_v is left invariant. For any g in G, f in $\mathcal{O}(G)$ and x in G,

$$((g \cdot D_v)(f))x = (D_v(f \circ l_g) \circ l_{g^{-1}})x = (D_v(f \circ l_g))(g^{-1}x) = \Big(T_e l_{g^{-1}x}(v)\Big)(f \circ l_g).$$

By the definition of the tangent map

$$((g \cdot D_v)(f))x = v(f \circ l_g \circ l_{g^{-1}x}) = v(f \circ l_x) = \Big(T_e l_x(v)\Big)f = (D_v(f))x. \qquad \square$$

DEFINITION 2.63. For v in T_eG let D_v denote the unique left invariant derivation given by v. Recall that D_v was defined by $D_v(f) : g \mapsto \Big((T_e l_g)(v)\Big)f$ for any f in $\mathcal{O}(G)$. The *Lie algebra of* G, denoted by $\mathrm{Lie}(G)$, is the tangent space T_eG equipped with the Lie bracket:

$$[\cdot,\cdot] : T_eG \times T_eG \to T_eG, \quad (a, b) \mapsto [v, w] := \Big[g \mapsto ([D_v, D_w](g))e\Big].$$

Clearly $\mathrm{Lie}(G)$ is a Lie algebra and $[D_v, D_w] = D_{[v,w]}$.

6. The Lie functor

Let k be $\mathbf{R}$ or $\mathbf{C}$.

PROPOSITION 2.64. We have a functor, denoted by Lie, from the category of Lie groups (over k) to the category of (k-)Lie algebras assigning to each Lie group G its Lie algebra $\mathrm{Lie}(G)$ and to each morphism $f : G_1 \to G_2$ of Lie groups the Lie algebra morphism $T_e f : \mathrm{Lie}(G_1) \to \mathrm{Lie}(G_2)$.

PROOF. Let $f : G_1 \to G_2$ be a morphism of Lie groups. Let v be in $\mathrm{Lie}(G_1)$ and let $v' = T_e f(v)$ be in $\mathrm{Lie}(G_2)$. Let D_v denote the unique left invariant derivation given by v and let $D_{v'}$ denote the unique left invariant derivation given by v'.

Since f is a morphism of groups $f \circ l_g = l_{f(g)} \circ f$ for any g in G_1. Hence $f^* \circ l^*_{f(g)} = l^*_g \circ f^*$ for any g in G_1. For F in $\mathcal{O}(G_2)$ and g in G_1,

$$\begin{aligned} ((f^* \circ D_{v'})F)g &= (D_{v'}(F))(f(g)) = \Big(T_e l_{f(g)}(v')\Big)F = v'(l^*_{f(g)}F) \\ &= \Big(T_e f(v)\Big)(l^*_{f(g)}F) = v(f^*(l^*_{f(g)}F)) = v(l^*_g(f^*F)) \\ &= \Big(T_e l_g(v)\Big))(f^*F) = (D_v(f^*F))g = ((D_v \circ f^*)F)g. \end{aligned}$$

We show that v' is the unique element of $\mathrm{Lie}(G_2)$ such that $D_v \circ f^* = f^* \circ D_{v'}$. Let v'' be an element of G_2 such that $D_v \circ f^* = f^* \circ D_{v''}$. Let $\mathrm{Eval}_e : \mathcal{O}(G_1) \to k$ be the *evaluation map* at $e \in G_1$ given by $h \mapsto h(e)$. Then $v' = \mathrm{Eval}_e \circ D_v \circ f^* = \mathrm{Eval}_e \circ f^* \circ D_{v''} = v''$.

Let w be another element in $\mathrm{Lie}(G_1)$ and let $w' = T_e f(w)$ be in $\mathrm{Lie}(G_2)$. It suffices to show that $[v, w]' := T_e f([v, w]) = [v', w']$. Note that $f^* \circ D_{[v,w]'} = [D_v, D_w] \circ f^*$ and that

$$\begin{aligned} f^* \circ D_{[v,w]'} &= D_v \circ D_w \circ f^* - D_w \circ D_v \circ f^* = D_v \circ f^* \circ D_{w'} - D_w \circ f^* \circ D_{v'} \\ &= f^* \circ D_{v'} \circ D_{w'} - f^* \circ D_{w'} \circ D_{v'} = f^* \circ [D_{v'}, D_{w'}] = f^* \circ D_{[v',w']}. \end{aligned}$$

By the uniqueness of $[v, w]'$ we may conclude that $[v, w]' = [v', w']$. $\square$

THEOREM 2.65. The functor Lie from Lie groups (over k) to (k-)Lie algebras induces an equivalence of categories from the category of connected and simply connected Lie groups to the category of finite dimensional Lie algebras.

PROOF. [**Edix**, Chapter 6, pages 38-43]. □

PROPOSITION 2.66. Let n be a nonnegative integer. The Lie algebra of $\mathrm{GL}_n(k)$ is isomorphic to the Lie algebra $M_n(k)$ of $(n \times n)$-matrices with coordinates in k. Furthermore, the Lie algebra of a sub Lie group of $\mathrm{GL}_n(k)$ is a sub Lie algebra of $\mathrm{M}_n(k)$.

PROOF. Let $G = \mathrm{GL}_n(k)$. Let L denote the tangent space $T_eG = \mathrm{M}_n(k)$ of G at the identity element e. Let a be in L. There is a unique left invariant derivation D_a on $\mathcal{O}(G)$ such that $a(f) = (D_a(f))e$ for any f in $\mathcal{O}(G)$. For any g in G and f in $\mathcal{O}(G)$,

$$(D_a(f))g = ((g \cdot D_a)(f))g = (D_a(f \circ l_g))e = a(f \circ l_g) = \Big(T_e l_g(a)\Big) f = (ga)(f).$$

Here ga denotes the product of g and a in L. Let $x_{i,j}$ be the coordinate functions on $\mathrm{M}_n(k)$. We compute $D_a(x_{i,j})$. For g in G and $\epsilon > 0$ a real number

$$x_{i,j}(g + \epsilon ga) = (g + \epsilon ga)_{i,j} = g_{i,j} + \epsilon (ga)_{i,j} = x_{i,j}(g) + \epsilon \sum_k g_{ik} a_{kj}.$$

We conclude $D_a(x_{i,j}) = \sum_k x_{i,k} a_{kj}$. Applying this identity twice gives

$$(D_a D_b) x_{i,j} = D_a \left(\sum_k x_{i,k} b_{kj} \right) = \sum_{k,k'} x_{i,k'} a_{k'k} b_{kj} = \sum_{k'} x_{i,k'} (ab)_{k'j} = D_{ab} x_{i,j}.$$

Hence $[D_a, D_b] x_{i,j} = D_{ab-ba} x_{i,j} = D_{[a,b]} x_{i,j}$. Two derivations D and E such that $Dx_{i,j} = Ex_{i,j}$ for all i and j are necessarily equal so we may conclude that $[D_a, D_b] = D_{[a,b]}$. We conclude that the Lie bracket on L is just the ordinary commutator of matrices.

Let H be a sub Lie group of G. The inclusion $H \to G$ clearly induces an inclusion $\mathrm{Lie}(H) \to \mathrm{M}_n(k)$. From the functoriality of the Lie algebra it follows that the Lie bracket for $\mathrm{Lie}(H)$ is the restriction of the one for $\mathrm{M}_n(k)$. □

From Remark 2.57 and our computations done in Proposition 2.13 we get the following Corollary.

COROLLARY 2.67. The Lie algebra of $\mathrm{SL}_n(k)$ is the k-Lie algebra of elements in $\mathrm{M}_n(k)$ with trace zero and the Lie algebra of $\mathrm{SU}_n(\mathbf{R})$ is the $\mathbf{R}$-Lie algebra of anti-hermitian elements in $\mathrm{M}_n(\mathbf{C})$ with trace zero. □

PROPOSITION 2.68. Let G be a sub Lie group of $\mathrm{GL}_n(k)$ and $\mathfrak{g} \subset \mathrm{M}_n(k)$ its Lie algebra. For g in G let $c_g : x \mapsto gxg^{-1}$ denote the conjugation on G by g. It is clear that G acts on itself by conjugation since c_g is an automorphism of G (as a Lie group) for any g in G. For any g in G the conjugation $c_g : G \to G$ induces (Section 6) a linear diffeomorphism $\mathrm{Lie}(c_g) : \mathfrak{g} \to \mathfrak{g}$ of the Lie algebra $\mathfrak{g}$ given by $\mathrm{Lie}(c_g)(X) = gXg^{-1}$ for any X in $\mathfrak{g}$. Furthermore, the morphism of Lie groups $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ given by $g \mapsto \mathrm{Lie}(c_g)$ induces a morphism of Lie algebras $\mathrm{ad} := \mathrm{Lie}(\rho) : \mathfrak{g} \to \mathrm{End}(\mathfrak{g})$ given by $\mathrm{ad}(x) = \Big\{ y \mapsto [x,y] = xy - yx \Big\}$.

PROOF. Let e be the identity element of G. For any X in $\mathfrak{g}$ and any real number $\epsilon > 0$ we may compute the product $(e + \epsilon X)(e - \epsilon X) = e + O(\epsilon^2)$ in $\mathrm{M}_n(k)$. So for $\epsilon > 0$ small enough $e + \epsilon X$ is in G. For $\epsilon > 0$ small enough

$$c_g(e + \epsilon X) = g(e + \epsilon X)g^{-1} = gg^{-1} + g\epsilon X g^{-1} = e + \epsilon g X g^{-1}$$

for any X in $\mathfrak{g}$. This shows that $\mathrm{Lie}(c_g)(X) = gXg^{-1}$.

By Proposition 2.66, the Lie algebra of $\mathrm{GL}(\mathfrak{g})$ is $\mathrm{End}(\mathfrak{g})$. Let x be an element of $\mathfrak{g}$ and let $\epsilon > 0$ be a real number. We already noted that $e + \epsilon x$ is in G if ϵ is small enough. For y in $\mathfrak{g}$ and ϵ small enough

$$\mathrm{Ad}(e + \epsilon x)(y) = (e + \epsilon x) y (e + \epsilon x)^{-1} = (e + \epsilon x) y (e - \epsilon x + O(\epsilon^2)) = e + \epsilon(xy - yx) + O(\epsilon^2).$$

This shows that $\mathrm{ad}(x) = \Big\{ y \mapsto [x,y] = xy - yx \Big\}$ for any x in $\mathfrak{g}$. □

CHAPTER 3

Representations

1. Definitions

Definitions 3.1. Let k be a field. A *group representation (over k)* of a group G is a pair (ρ, V) where V is a k-vector space and $\rho : G \to \mathrm{GL}(V)$ is a morphism of groups. A *Lie algebra representation (over k)* of a k-Lie algebra $\mathfrak{g}$ is a pair (V, ρ) where V is a k-vector space and $\rho : \mathfrak{g} \to \mathrm{End}_k(V)$ is a morphism of k-Lie algebras.

A *morphism* of two group representations (ρ, V) and (σ, W) is a k-linear map $f : V \to W$ such that $\sigma(g) \circ f = f \circ \rho(g)$, i.e. the following diagram

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\rho(g)"'] & W \arrow[d, "\sigma(g)"] \\ V \arrow[r, "f"'] & W \end{tikzcd}$$

is commutative for all g in G. Morphisms of Lie algebra representations for $\mathfrak{g}$ are defined similarly.

Definition 3.2. Let G be a Lie group and (ρ, V) be a finite dimensional group representation of the underlying abstract group of G. Then (ρ, V) is *differentiable* if the map $G \times V \to V$ given by $(g, v) \mapsto (\rho(g))(v) = gv$ is a morphism of manifolds. We shall always assume finite dimensional group representations of Lie groups to be differentiable.

Example 3.3. Let k be $\mathbf{R}$ or $\mathbf{C}$. For any finite dimensional vector space V over k the natural action of $\mathrm{GL}(V)$ on V gives us a (differentiable) group representation over k of $\mathrm{GL}(V)$ on V.

Example 3.4. With notation from Proposition 2.68, $(\mathrm{Ad}, \mathfrak{g})$ is called the *adjoint group representation* of G and $(\mathrm{ad}, \mathfrak{g})$ is called the *adjoint Lie algebra representation.*

Definition 3.5. Let G be a group and let V be a group representation of G. A subspace $W \subset V$ is *invariant* under the action of G if $\{g \cdot w \in V \mid g \in G,\ w \in W\} \subset W$. We call a group representation V of G *irreducible* if V has exactly two linear subspaces which are invariant under the action of G: $\{0\}$ and V.

We have a similar definition for Lie algebra representations.

2. Representations of $\mathbf{SL}_2(\mathbf{C})$ and its Lie algebra

Let k be a field and let d be a nonnegative integer. Let $k[x, y]_d$ denote the set of homogeneous polynomials in $k[x, y]$ of degree d. Note that $k[x, y]_d$ is a k-vector space with basis $(x^d, x^{d-1}y, \ldots, y^d)$.

Let k be an infinite field. Since k is infinite we may identify $k[x, y] = \bigoplus_d k[x, y]_d$ with the set of polynomial functions on k^2 to k. Now, consider the natural action of $\mathrm{SL}_2(k)$ on k^2 as in Example 3.3.

Proposition 3.6. Let d be a nonnegative integer. Let $\rho_d : \mathrm{SL}_2(k) \to \mathrm{GL}(k[x, y]_d)$ be the map given by $(\rho_d(g)(f))(v)) = f(g^t v)$ for all g in $\mathrm{SL}_2(k)$, f in $k[x, y]_d$ and v in k^2. Then $(\rho_d, k[x, y]_d)$ is a group representation of $\mathrm{SL}_2(k)$.

PROOF. For g and h in $\mathrm{SL}_2(k)$ we have, for f in $k[x,y]_d$ and v in k^2,

$$\Big(\rho_d(gh)(f)\Big)(v) = f((gh)^t v) = f(h^t g^t v) = \Big(\rho_d(h)(f)\Big)(g^t v) = \Big((\rho_d(g)\circ\rho_d(h))(f)\Big)(v). \qquad \square$$

By Corollary 2.67, the Lie algebra of $\mathrm{SL}_2(\mathbf{C})$ is the Lie algebra of trace zero elements in $\mathrm{M}_2(\mathbf{C})$. Let k be a (not necessarily infinite) field. Let L denote the k-Lie algebra of trace zero elements in $\mathrm{M}_2(k)$. Note that L has a k-basis (h, a_+, a_-) where

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ a_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } a_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The Lie bracket on L is given by:

$$[h, a_+] = 2a_+,\ [h, a_-] = -2a_-,\ [a_+, a_-] = h.$$

PROPOSITION 3.7. The action of $\mathrm{SL}_2(k)$ on $k[x,y]$ induces a Lie algebra representation over k of L on $k[x,y]$. The endomorphisms of the k-vector space $k[x,y]$ given by h, a_+ and a_- are given by

$$hx^iy^j = (i-j)x^iy^j,\ a_+x^iy^j = jx^{i+1}y^{j-1},\ a_-x^iy^j = ix^{i-1}y^{j+1}$$

for any two nonnegative integers i and j.

PROOF. Let a be in L and let $\epsilon > 0$ be a small real number such that $1+\epsilon a$ is in $\mathrm{SL}_2(k)$. We have $(1+\epsilon a)f = f + \epsilon(af) + O(\epsilon^2)$. For i and j two nonnegative integers we have:

$$(1+\epsilon h)x^iy^j = ((1+\epsilon)x)^i((1-\epsilon)y)^j = x^iy^j + \epsilon(i-j)x^iy^j + O(\epsilon^2).$$

This shows that $hx^iy^j = (i-j)x^iy^j$. The computations for a_+ and a_- are similar. $\square$

REMARK 3.8. The basis elements h, a_+, a_- of L act as derivations on $k[x,y]$ since for all f in $\mathbf{C}[x,y]$ we have:[1]

$$hf = \left(x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}\right)f,\ a_+ = \left(x\frac{\partial}{\partial y}\right)f,\ a_- = \left(y\frac{\partial}{\partial x}\right)f.$$

THEOREM 3.9. The Lie algebra representation $k[x,y]_d$ of L is irreducible for all nonnegative integers d if and only if k is of characteristic zero. When k is an infinite field we have a similar statement for the group representation $k[x,y]_d$ of $\mathrm{SL}_2(k)$.

PROOF. Let k be of characteristic zero. We show that the Lie algebra representation $k[x,y]_d$ of L is irreducible. Let d be a nonnegative integer. Let $V \neq 0$ be an invariant subspace of $k[x,y]_d$. We show that $V = k[x,y]_d$. Assume that x^d is in V. By Proposition 3.7 $a_- \cdot x^d = dx^{d-1}y$. From this, and the fact that k is of characteristic zero, it follows that $x^{d-1}y \in V$. Repeating this process gives us that $x^iy^j \in V$ for all nonnegative integers i and j. This shows that $V = k[x,y]_d$.

We show that $x^d \in V$. Since $V \neq 0$ there is an element v in V such that $v \neq 0$. Since a basis for $k[x,y]_d$ is given by $(x^d, x^{d-1}y, \ldots, y^d)$ there exist scalars $v_i \in k$ for $i = 0, \ldots, d$ such that $v = v_0y^d + v_1xy^{d-1} + \ldots + v_dx^d$. There is a minimal $i \in \{0, \ldots, d\}$ such that $v_i \neq 0$. Applying the matrix a_+ exactly $d-i$ times to v gives us that $a_+^{d-i}v$ is nonzero, since k is of characteristic zero. Moreover $a_+^{d-i}v \in kx^d$. This shows that $x^d \in V$.

Let k be of characteristic $p > 0$. It is not hard to see from the identities in Proposition 3.7 that the subspace generated by x^p and y^p of $k[x,y]_p$ is invariant under the action of L. Even better, the action of L on this subspace is trivial. This shows that $k[x,y]_d$ is not irreducible for all nonnegative integers d and finishes the first part of the theorem.[2]

The last part of the theorem follows from the observation that any nonzero subspace of $k[x,y]_d$ which is stable under the action of $\mathrm{SL}_2(k)$ is stable under the (induced) action of L. $\square$

[1] Differentiating polynomials makes sense over any field.

[2] Actually, it is not hard to see that $k[x,y]_d$ is irreducible for all nonnegative integers $d < p$ since the same argument used when k is of characteristic zero applies.

From now on, let k be a field of characteristic zero and let L denote the Lie algebra of trace zero elements in $M_2(k)$. Recall that L has a basis (h, a_+, a_-) as above.

DEFINITION 3.10. For V a finite dimensional Lie algebra representation over k of L and for λ in k we say that an element v of V is of *weight* λ if $hv = \lambda v$. Note that $0 \in V$ is of any weight. We denote the subspace of V of elements of weight λ by $V(\lambda)$.

PROPOSITION 3.11. Let V be a finite dimensional Lie algebra representation of L and let v be a nonzero element of V of some weight λ. For i a nonnegative integer let $v_i := a_-^i v$. Note that $v_0 = v$. Then

(1) $a_+v \in V(\lambda + 2)$,
(2) $a_-v \in V(\lambda - 2)$,
(3) $v_i \in V(\lambda - 2i)$ for all $i \geq 0$,
(4) if $a_+v = 0$ then $a_+v_i = i(\lambda - i + 1)v_{i-1}$ for all nonnegative integers i and there exists a nonnegative integer d such that the weight λ of v equals d, and the subrepresentation of V generated by v is isomorphic to $k[x, y]_d$.

PROOF. From the relation $[h, a_+] = 2a_+$ it follows that

$$h(a_+v) = a_+(hv) + 2a_+v = (\lambda + 2)a_+v.$$

For the second identity the computation is similar. The third identity follows by induction on i.

Assume that $a_+v = 0$. We show that $a_+v_i = i(\lambda - i + 1)v_{i-1}$ by induction on i. The case $i = 0$ being true by assumption we let i be a positive integer and assume that $a_+v_{i-1} = (i-1)(\lambda - i + 2)v_{i-2}$. Since $h = [a_+, a_-]$ we have

$$\begin{aligned} a_+v_i &= (h + a_-a_+)v_{i-1} = hv_{i-1} + a_-a_+v_{i-1} = (\lambda - 2(i-1))v_{i-1} + a_-(i-1)(\lambda - i + 2)v_{i-2} \\ &= (\lambda + 2 - 2i + (i-1)(\lambda - i + 2))v_{i-1} = (i\lambda - i^2 + i)v_{i-1} = i(\lambda - i + 1)v_{i-1}. \end{aligned}$$

We used the third identity and the induction hypothesis in obtaining the second line. Note that the nonzero v_i are all linearly independent by the first identity. Since V is finite dimensional there exists a smallest integer d such that $v_d \neq 0$ and $a_i v_d = v_{d+1} = 0$. Evidently $v_{d+j} = 0$ for all positive integers j. By the fourth identity,

$$0 = a_+v_{d+1} = (d+1)(\lambda - (d+1) + 1)v_d = (d+1)(\lambda - d)v_d.$$

This implies that $\lambda = d$. Now, $(v_0, v_1, v_2, \ldots, v_d)$ is a basis for the subrepresentation of V generated by $v = v_0$. Let $\langle v \rangle$ denote the Lie algebra representation of L generated by v. Consider the k-linear map $\langle v \rangle \to k[x, y]_d$ given by

$$\begin{cases} x^d & \mapsto & v = v_0 \\ x^{d-1}y & \mapsto & v_1 \\ \vdots & \vdots & \vdots \\ y^d & \mapsto & v_d. \end{cases}$$

By Proposition 3.7, this map is an isomorphism of Lie algebra representations of L. □

COROLLARY 3.12. Let V be an irreducible finite dimensional Lie algebra representation of L. The roots of the characteristic polynomial of h acting on V are all integers.

PROOF. For now (and only now) let L^k denote the Lie algebra of 2×2 matrices with trace zero and coordinates in k, i.e. $L^k \simeq kh \oplus ka_+ \oplus ka_-$. Let $k \subset k'$ be a field extension. By properties of the tensor product $k' \otimes_k L^k$ and $L^{k'}$ are isomorphic as Lie algebras.[3] Let V be an irreducible finite dimensional Lie algebra representation over k of L^k. Consider the roots of the characteristic polynomial of the action of $h \in L^k$ on V. Let λ be such a root and denote the endomorphism $v \mapsto \lambda v$ of V (also) by λ. For $k \to \bar{k}$ an algebraic closure of k we have that $\bar{k} \otimes_k V$ is a finite dimensional Lie algebra representation over $\bar{k}$ of $L^{\bar{k}}$. Consider this Lie algebra representation and let $v \in \ker(h - \lambda)$, i.e. the eigenspace belonging to λ. There is an i such that $a_+^{i+1}v = 0$ and

[3]This is called extension of scalars.

$a_+^i v \neq 0$. Apply Proposition 3.11 to the element $a_+^i v$ to see that λ (or actually $\lambda + 2i$) is an integer. $\square$

DEFINITION 3.13. Let V be an irreducible finite dimensional Lie algebra representation of L. Let d be the largest among the roots of the characteristic polynomial of h acting on V. Then d is called the *highest weight* of V. Any nonzero vector v in $V(d)$ is called a *maximal vector* of V.

PROPOSITION 3.14. Let V be an irreducible finite dimensional Lie algebra representation over k of L. Then V is isomorphic to some $k[x,y]_d$.

PROOF. Let d be the highest weight of V. By Corollary 3.12, d is an integer. Let v in $V(d)$ be a maximal vector. By Proposition 3.11, a_+v is in $V(d+2)$. This implies that $a_+v = 0$. Again by Proposition 3.11, there is an injective morphism of Lie algebra representations $f : k[x,y]_d \to V$ sending x^d to v since the subrepresentation generated by v is isomorphic to $k[x,y]_d$. The irreducibility of $k[x,y]_d$ and V implies that f is an isomorphism of Lie algebra representations. $\square$

PROPOSITION 3.15. Each representation V of L is isomorphic to a direct sum of irreducible representations.

PROOF. [**Edix**, Proposition 10.4.6, pages 58-60]. $\square$

The following theorem is an application of all the results in this section.

THEOREM 3.16. Every finite dimensional Lie algebra representation of L is isomorphic to a direct sum of copies of the $k[x,y]_d$. For k is $\mathbf{R}$ or $\mathbf{C}$ every finite dimensional group representation over k of $\mathrm{SL}_2(k)$ is isomorphic to a direct sum of copies of the $k[x,y]_d$. $\square$

COROLLARY 3.17. Let V be a (not necessarily finite-dimensional) representation of L, and suppose that each vector $v \in V$ lies in a finite-dimensional invariant subspace. Then V is the (possibly infinite) direct sum of finite-dimensional invariant subspaces on which L acts irreducibly.

PROOF. By hypothesis and Proposition 3.15 each member of V lies in a finite direct sum of irreducible invariant spaces. Thus $V = \sum_{s \in S} U_s$ where S is some index set and each U_s is an irreducible invariant subspace. We show that this sum is a direct sum. A subset R of S is said to be *independent* when the sum $\sum_{r \in R} U_r$ is a direct sum. It is clear that the union of any increasing chain of independent subsets of S is itself independent. By Zorn's Lemma there is a maximal independent subset T of S. By definition the sum $\sum_{t \in T} U_t$ is direct. We shall show that $V = \sum_{t \in T} U_t$. It suffices to show that $U_s \subset \sum_{t \in T} U_t$ for any $s \in S$. For $s \in T$ it is clear, so suppose that $s \notin T$. Then the maximality of T implies that $T \cup \{s\}$ is not independent. Consequently the sum $U_s + \sum_{t \in T} U_t$ is not direct, and we must have $U_s \cap \sum_{t \in T} U_t \neq (0)$. But this intersection is an invariant subspace of U_s. Since U_s is irreducible $U_s \cap \sum_{t \in T} U_t = U_s$ showing that $U_s \subset \sum_{t \in T} U_t$. $\square$

3. Quaternions

We define elements in $\mathrm{M}_2(\mathbf{C})$ by:

$$1 \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

Note that $i^2 = j^2 = k^2 = ijk = -1$. These equations imply that $ij = k$, $ji = -k$, $jk = i$, $kj = -i$, $ki = j$ and $ik = -j$. We conclude that the subspace with basis $(1, i, j, k)$ is a sub $\mathbf{R}$-algebra of $\mathrm{M}_2(\mathbf{C})$. This algebra is called the real quaternion algebra of Hamilton and is denoted by $\mathbf{H}$. Note that $\mathbf{R} \hookrightarrow \mathbf{H}$ canonically. Multiplying two elements

$$a + bi + cj + dk = \begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}$$

and

$$(e+fi+gj+hk) = \begin{pmatrix} e+fi & g+hi \\ -g+hi & e-fi \end{pmatrix}$$

of $\mathbf{H}$ gives $ae-bf-cg-dh+(af+be+ch-dg)i+(ag-bh+ce+df)j+(ah+bg-cf+de)k$ for any a, b, c, d, e, f, g and h in $\mathbf{R}$. Another way to describe $\mathbf{H}$ is to say that it is a $\mathbf{R}$-vector space with $(1,i,j,k)$ and an associative multiplication defined by $i^2=j^2=k^2=ijk=-1$.

For z in $\mathbf{C}$ let $\bar{z}$ denote the complex conjugate of z. The elements of $\mathbf{H}$ are precisely the matrices of the form $\begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}$ with $a,b\in\mathbf{C}$. Since the determinant of an element $\begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}$ in $\mathbf{H}$ equals $\mid a\mid^2+\mid b\mid^2$ any nonzero element in $\mathbf{H}$ is invertible with inverse $\begin{pmatrix} \bar{a} & \bar{b} \\ -b & a \end{pmatrix}$ in $\mathbf{H}$. We conclude that the quaternion algebra $\mathbf{H}$ is a division algebra.

For x in $\mathbf{H}$ let $x^*=\bar{x}^t=\overline{x^t}$ denote the hermitian conjugate of x. Let $*:\mathbf{H}\to\mathbf{H}$ be the $\mathbf{R}$-linear map given by $x\mapsto x^*$.

Proposition 3.18. The linear map $*$ is an involution and $(xy)^*=y^*x^*$ for any x and y in $\mathbf{H}$.

Proof. It is clear that $*$ is an involution since $(x^t)^t=x$ and $\bar{\bar{x}}=x$ for any x in $\mathbf{H}$. The second part follows from the observation that for any x and y in $\mathbf{H}$ it holds that $(xy)^*=\overline{(xy)^t}=\overline{y^tx^t}=\overline{y^t}\cdot\overline{x^t}=y^*x^*$. □

Let $N:\mathbf{H}\to\mathbf{R}$ be the map defined by $N(x)=\det(x)$.

Proposition 3.19. The map N defines a norm on $\mathbf{H}$.

Proof. The underlying vector space of $\mathbf{H}$ may be identified with $\mathbf{R}^4$ by identifying $a+bi+cj+dk$ in $\mathbf{H}$ with (a,b,c,d) in $\mathbf{R}^4$. It is clear that the norm N under this identification is the usual Euclidean norm (squared) on $\mathbf{R}^4$, i.e. $N(a+bi+cj+dk)=|(a,b,c,d)|^2=a^2+b^2+c^2+d^2$. □

Remark 3.20. The above norm is a quadratic form and gives us an inner product

$$\langle x,y\rangle := \frac{N(x+y)-N(x)-N(y)}{2}$$

on $\mathbf{H}$. One can show by direct computation that $\langle x,y\rangle=\dfrac{x^*y+y^*x}{2}$.

Proposition 3.21. The groups $\{\ x\in\mathbf{H}\mid N(x)=1\ \}\subset\mathbf{H}^*$ and $\mathrm{SU}_2(\mathbf{R})\subset\mathbf{H}^*$ are isomorphic. Note that $\{\ x\in\mathbf{H}\mid N(x)=1\ \}=\{\ (a,b)\in\mathbf{C}^2\mid |a|^2+|b|^2=1\ \}$ is the unit sphere $\mathbf{S}^3\subset\mathbf{C}^2$.

Proof. Since $xx^*=\det(x)$ for any x in $\mathbf{H}$ it is clear that $\{\ x\in\mathbf{H}\mid N(x)=1\ \}\subset\mathrm{SU}_2(\mathbf{R})$. Let $x=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an element of $\mathrm{SU}_2(\mathbf{R})$. It suffices to show that x is in $\mathbf{H}$ since $x^*=x^{-1}$ and $\det(x)=ad-bc=1$. Note that $x^*=\begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix}=\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}=x^{-1}$ implies that $x=\begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}$. We conclude that x is an element of $\mathbf{H}$. □

We end this section with the following remark.

Remark 3.22. By Example 2.20, $\mathbf{H}$ has the structure of a $\mathbf{R}$-Lie algebra. The Lie bracket on $\mathbf{H}$ is given by $[x,y]=xy-yx$. Since $\mathbf{H}^*$ is open in $\mathbf{H}$ the group $\mathbf{H}^*$ is a real Lie group with Lie algebra $\mathbf{H}$.

4. Complex representations of $\mathbf{SU}_2(\mathbf{R})$

Let L be the $\mathbf{C}$-Lie algebra of trace zero elements in $\mathrm{M}_2(\mathbf{C})$, i.e. the Lie algebra of $\mathrm{SL}_2(\mathbf{C})$. Let L' be the Lie algebra of $\mathrm{SU}_2(\mathbf{R})$, i.e. the $\mathbf{R}$-Lie algebra of anti-hermitian elements in $\mathrm{M}_2(\mathbf{C})$ with trace zero. Note that L' has a basis (I, J, K) where

$$I = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \ J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ K = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

The Lie bracket on L' is given by:

$$[I, J] = 2K, \ [J, K] = 2I, \ [K, I] = 2J.$$

Lemma 3.23. The inclusion of L' in L induces an isomorphism of Lie algebras from the complexification $L'_{\mathbf{C}} = L' \otimes_{\mathbf{R}} \mathbf{C}$ of L' to L.

Proof. Note that $h = -iI$, $a_+ = \frac{J-iK}{2}$ and $a_- = -\frac{J+iK}{2}$. These identities imply that I, J and K form a $\mathbf{C}$-basis for L. Thus applying the functor $- \otimes_{\mathbf{R}} \mathbf{C}$ to the injective morphism $i : L' \to L$ gives us an isomorphism $i_{\mathbf{C}} : L'_{\mathbf{C}} \to L$ of complex Lie algebras. $\square$

Proposition 3.24. Restriction from L to L' induces an isomorphism from the category of finite dimensional representations over $\mathbf{C}$ of L to that of L'. By Theorem 3.16, every finite dimensional complex vector space with an action of $\mathrm{SU}_2(\mathbf{R})$ is isomorphic to a direct sum of copies of the $\mathbf{C}[x, y]_d$.

Proof. Giving a representation over $\mathbf{C}$ of L' is the same as giving a representation over $\mathbf{C}$ of the complexification $L'_{\mathbf{C}}$. By Lemma 3.23, the representations over $\mathbf{C}$ of L' are the same as those of L. $\square$

Lemma 3.25. Let $F : \mathcal{C} \to \mathcal{D}$ be an equivalence of categories. Then for all $X, Y \in \mathcal{C}$ the morphism of sets $\mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$ is bijective.

Proof. Let $F : \mathcal{C} \to \mathcal{D}$ be an equivalence of categories. Then there exists a functor $G : \mathcal{D} \to \mathcal{C}$ and an isomorphism of functors $\alpha : GF \to \mathrm{id}_{\mathcal{C}}$[4]. Let X and Y be two objects of $\mathcal{C}$. For any $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, we have $f = \alpha(Y) \circ GF(f) \circ \alpha(X)^{-1}$ since the following diagram

$$\begin{tikzcd} GF(X) \arrow[r, "GF(f)"] \arrow[d, "\alpha(X)"'] & G(Y) \arrow[d, "\alpha(Y)"] \\ X \arrow[r, "f"'] & Y \end{tikzcd}$$

is commutative. Hence the morphism of sets $\mathrm{Hom}_C(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y)$ is injective. Likewise we can show that the morphism of sets $\mathrm{Hom}_D(X, Y) \to \mathrm{Hom}_{\mathcal{C}}(G(X), G(Y)$ is injective. Moreover, let $g \in \mathrm{Hom}_{\mathcal{D}}(X, Y)$. Define a morphism $f : X \to X$ by $f = \alpha(Y) \circ G(g) \circ \alpha(X)^{-1}$. Since the morphism of sets $\mathrm{Hom}_D(X, Y) \to \mathrm{Hom}_{\mathcal{C}}(G(X), G(Y)$ is injective, we have $F(f) = g$ if and only if $GF(f) = G(g)$. Since $\alpha : GF \to \mathrm{id}_{\mathcal{C}}$ is a natural transformation $GF(f) = \alpha(Y)^{-1} \circ f \circ \alpha(X)$. By the definition of f, $G(g) = GF(f)$. This implies that $F(f) = g$ and that the morphism of sets $\mathrm{Hom}_C(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y)$ is surjective. $\square$

Theorem 3.26. Let V be a finite dimensional complex vector space with an action of $\mathrm{SU}_2(\mathbf{R})$. Then V is isomorphic to a direct sum of copies of the $\mathbf{C}[x, y]_d$.

Proof. By Proposition 3.21, $\mathrm{SU}_2(\mathbf{R})$ is connected and simply connected so the theorem follows by Theorem 2.65 and Lemma 3.25. $\square$

[4]We shall not need the isomorphism of functors $FG \to \mathrm{id}_{\mathcal{D}}$.

5. Complex representations of $\mathrm{SO}_3(\mathbf{R})$

Let $\mathbf{H}$ be the real quaternion algebra of Hamilton. We showed that $\mathbf{H}$ is a normed space. Let V be the orthogonal complement of $\mathbf{R}$. Consider the orthogonal decomposition $\mathbf{H} = \mathbf{R} \bigoplus V$. Note that $V = \mathbf{R}i \oplus \mathbf{R}j \oplus \mathbf{R}k$ is a three-dimensional normed space which we call the space of *pure quaternions*.

For $x \in \mathbf{H}^*$, let the conjugation by x from $\mathbf{H}$ to itself be given by $c_x : y \mapsto xyx^{-1}$. Note that the center of $\mathbf{H}$ is equal to $\mathbf{R}$. For, assume that x is in $\mathbf{H}$ and commutes with all elements of $\mathbf{H}$. Since elements that commute with $i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$ have to be diagonal and diagonal matrices that commute with $j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ have to be scalar matrices we may identify the center of $\mathbf{H}$ with $\mathbf{R}$.

Consider the adjoint group representation $\mathrm{Ad} : \mathbf{H}^* \to \mathrm{GL}(\mathbf{H})$ of $\mathbf{H}^*$. Since the center of $\mathbf{H}$ is $\mathbf{R}$ restricting c_x to $(0) \oplus V = V$ gives us a morphism $\mathrm{Ad} : \mathbf{H}^* \to \mathrm{GL}(V)$ given by $x \mapsto c_x$. Since $N(c_x(y)) = N(xyx^{-1}) = N(x)N(y)N(x^{-1}) = N(x)N(y)N(x)^{-1} = N(y)$ for all y in $\mathbf{H}$, the map c_x is orthogonal. Let $\mathrm{O}(V)$ denote the group of orthogonal endomorphisms of V. Restriction of Ad to the subgroup $\mathrm{SU}_2(\mathbf{R})$ of $\mathbf{H}^*$ gives us a representation (again) denoted by $\mathrm{Ad} : \mathrm{SU}_2(\mathbf{R}) \to \mathrm{O}(V)$ of $\mathrm{SU}_2(\mathbf{R})$ on V.

Lemma 3.27. The image of Ad is contained in $\mathrm{SO}(V)$.

Proof. Choosing the orthogonal basis (i, j, k) for V gives us a (non-unique) isomorphism of $\mathrm{SO}(V)$ with $\mathrm{SO}_3(\mathbf{R})$. Therefore $\mathrm{SO}(V)$ is connected. Note that $\mathrm{SO}(V)$ and the set of orthogonal maps with determinant -1 are the two connected components of $\mathrm{O}(V)$. Since $\mathrm{SU}_2(\mathbf{R}) \simeq \mathbf{S}^3$ is connected, Ad maps $\mathrm{SU}_2(\mathbf{R})$ into the connected component of $\mathrm{O}(V)$ containing the identity. This is precisely $\mathrm{SO}(V)$. $\square$

Lemma 3.28. The sub Lie algebra V of $\mathbf{H}$ is the Lie algebra of $\mathrm{SU}_2(\mathbf{R})$. In particular, $\mathrm{Ad} : \mathrm{SU}_2(\mathbf{R}) \to \mathrm{O}(V)$ is the adjoint group representation of $\mathrm{SU}_2(\mathbf{R})$.

Proof. This follows from the fact that $ij - ji = 2k$, $jk - kj = 2i$ and $ki - ik = 2j$. $\square$

Proposition 3.29. The (induced) morphism ad from V to the Lie algebra of $\mathrm{SO}(V)$ is an isomorphism of Lie algebras.

Proof. Since $\mathrm{SU}_2(\mathbf{R})$ and $\mathrm{SO}(V)$ are both Lie groups of dimension three it suffices to show that the kernel of the induced morphism ad is trivial. By Proposition **??**, $\mathrm{ad}(y) = \big[v \mapsto yv - vy\big]$ for all y in V. Let y be in V and assume that $\mathrm{ad}(y) = 0$. Then $yv - vy = 0$ for all v in V. Since the center of $\mathbf{H}$ is $\mathbf{R}$ we must have that y is in $\mathbf{R}$. But this implies that $y = 0$, since V is the orthogonal complement of $\mathbf{R}$. $\square$

Lemma 3.30. Let G be a connected Lie group. Let U be an open neighborhood of e and let U^n consist of all n-fold products of elements of U. Then $G = \bigcup_{n=1}^{\infty} U^n$.

Proof. Let $V = U \cap U^{-1}$ where $U^{-1} = \{\ g^{-1} \mid g \in U\ \}$. Let $H = \bigcup_{n=1}^{\infty} V^n$. Then $H \subset \bigcup_{n=1}^{\infty} U^n$, H is an abstract subgroup of G and an open subset of G since $hV \subset H$ for all h in H. By the latter each coset mod H is open in G. Now H is the complement of the union of all these cosets and therefore H is a closed subset of H. Since H is nonempty we conclude that $H = G$. This shows that $G = \bigcup_{n=1}^{\infty} U^n$. $\square$

Lemma 3.31. Let G and H be connected Lie groups and let $\mathfrak{g}$ (respectively $\mathfrak{h}$) be the Lie algebra of G (respectively H). A morphism of Lie groups $\phi : G \to H$ is surjective if the induced morphism $\mathrm{Lie}(\phi) : \mathfrak{g} \to \mathfrak{h}$ is an isomorphism of Lie algebras.[5]

[5] We can actually prove that ϕ is a covering map, but since we do not need this we shall not give the proof.

PROOF. Suppose that $\mathrm{Lie}(\phi) : \mathfrak{g} \to \mathfrak{h}$ is an isomorphism of Lie algebras. Then ϕ is everywhere a local diffeomorphism. Thus, by Lemma 3.30, ϕ maps G onto H since ϕ is a morphism of Lie groups and since $\phi(G)$ contains a neighborhood of the identity in the connected Lie group H. □

THEOREM 3.32. The map $\mathrm{Ad} : \mathrm{SU}_2(\mathbf{R}) \to \mathrm{SO}(V)$ is surjective with kernel $\{1, -1\} \subset \mathrm{SU}_2(\mathbf{R})$. We conclude that $\mathrm{SU}_2(\mathbf{R})/\{1, -1\}$ and $\mathrm{SO}(V)$ are isomorphic as Lie groups.

PROOF. By Lemma 3.31, Ad is surjective. The kernel is given by the intersection $\mathbf{R} \cap \mathrm{SU}_2(\mathbf{R}) = \{1, -1\}$. □

COROLLARY 3.33. The map $\mathrm{Ad} : \mathrm{SU}_2(\mathbf{R}) \to \mathrm{SO}(V)$ is a double covering of $\mathrm{SO}(V)$. We conclude that $\mathrm{SU}_2(\mathbf{R})$ is the universal covering Lie group of $\mathrm{SO}_3(\mathbf{R})$ and that $\pi_1(\mathrm{SO}_3(\mathbf{R})) = \mathbf{Z}/2\mathbf{Z}$.

PROOF. This is an application of Proposition 2.48. □

Since $\mathrm{SO}_3(\mathbf{R})$ is isomorphic to the quotient of $\mathrm{SU}_2(\mathbf{R})$ by its subgroup $\{1, -1\}$ we have the following theorem.

THEOREM 3.34. The irreducible complex representations of $\mathrm{SO}_3(\mathbf{R})$ are the $\mathbf{C}[x, y]_d$ with $d \geq 0$ even. All finite dimensional complex representations of $\mathrm{SO}_3(\mathbf{R})$ are direct sums of irreducible representations.

PROOF. Consider the action of $\mathrm{SU}_2(\mathbf{R})$ on $\mathbf{C}[x, y]_d$. For any nonnegative integer i,

$$-1 \cdot x^i y^{d-i} = (-x)^i (-y)^{d-i} = (-1)^{i+d-i} x^i y^{d-i} = (-1)^d x^i y^{d-i}.$$

Since any morphism of groups $f : G \to G'$ factors uniquely through the quotient G/H if H is a normal subgroup contained in the kernel of f, giving a representation of $\mathrm{SO}_3(\mathbf{R})$ is to give a representation of $\mathrm{SU}_2(\mathbf{R})$ that is trivial on $\{-1, -1\}$. □

6. Résumé

In this section we briefly summarize what we did in this chapter. Please note that we do not make very precise statements and we shall be very informal in this section.

Basically, the first thing we did was define a representation of $\mathrm{SL}_2(\mathbf{C})$. This gave us a representation of its Lie algebra L. Working with this Lie algebra turned out to be much easier since the representations of L are direct sums of exactly those representations induced by $\mathrm{SL}_2(\mathbf{C})$. This means that translating this back to $\mathrm{SL}_2(\mathbf{C})$ we get a complete description of the representations of $\mathrm{SL}_2(\mathbf{C})$. Now, what about the representations of $\mathrm{SU}_2(\mathbf{R})$? Well, before obtaining those we studied quaternions. Basically, the quaternions are $\mathbf{C} \times \mathbf{C}$ with a multiplication defined on its elements. The quaternions show a lot of similarity with $\mathbf{C}$, which is just $\mathbf{R} \times \mathbf{R}$ with a multiplication defined on its elements. For example, we have the notion of pure quaternions just like we have imaginary numbers. Furthermore, the norm of a quaternion can be compared with the modulus of a complex number. Anyway, one of the main purposes quaternions served is showing that $\mathrm{SU}_2(\mathbf{R})$ is homeomorphic to $\mathbf{S}^3$. Now since $\mathbf{S}^3$ is connected and simply connected, giving representations of $\mathrm{SU}_2(\mathbf{R})$ isn't hard with the aid of the equivalence of categories of the category of connected and simply connected Lie groups and the category of Lie algebras. As a last subject we turned to the representations of the rotation group $\mathrm{SO}_3(\mathbf{R})$. We found these by looking at the adjoint representation of the Lie group $\mathbf{H}^*$. The main idea is showing that this representation induces an isomorphism of Lie groups $\mathrm{SU}_2(\mathbf{R})/\{1, -1\} \simeq \mathrm{SO}_3(\mathbf{R})$. Hence to give representations of $\mathrm{SO}_3(\mathbf{R})$ is to give a representation of $\mathrm{SU}_2(\mathbf{R})$ where -1 acts trivially. For the physicists among us, the theory of the (relativistic) electron should be recognizable in the fact that $\mathrm{SO}_3(\mathbf{R})$ has no representations of odd dimension.

CHAPTER 4

Homotopy groups of $\mathrm{SU}_2(\mathbf{R})$

In this chapter we briefly study the homotopy groups of $\mathrm{SU}_2(\mathbf{R})$. For a more detailed treatment of the material in this section we refer to [**Hat**, Chapter 4].

1. Definitions

DEFINITION 4.1. **(Wedge Sum)** The *wedge sum* of two pointed topological spaces (X, x) and (Y, y), denoted by $X \vee Y$, is defined to be the quotient of the disjoint union $X \coprod Y$ obtained by identifying x with y. Its base point is the class of x (or y).

REMARK 4.2. The wedge sum is the coproduct in the category of pointed topological spaces $\mathfrak{Top}_*$. This implies that $X \vee -$ is actually a functor from $\mathfrak{Top}_*$ to itself for any $X \in \mathfrak{Top}_*$.

DEFINITION 4.3. **(Higher Homotopy Groups)** Let n be a positive integer and let (X, x) be a pointed topological space. Choose s_0 to be the base point of $\mathbf{S}^n$. The set of homotopy classes of maps $f : (\mathbf{S}^n, s_0) \to (X, x)$ is denoted by $\pi_n(X, x_0)$. It has a group structure defined as follows: for two maps $f, g : \mathbf{S}^n \to (X, x_0)$ the sum $f + g$ is the composition $\mathbf{S}^n \xrightarrow{c} \mathbf{S}^n \vee \mathbf{S}^n \xrightarrow{f \vee g} X$ where c collapses the equator $\mathbf{S}^{n-1}$ in $\mathbf{S}^n$ to a point and we choose the base point to lie in this equator. We call $\pi_n(X, x)$ the *n-th homotopy group* of X based at x. Note that $\pi_n(X, x)$ is the fundamental group of X based at x for $n = 1$.

REMARK 4.4. The n-th homotopy functor can be defined similarly to the fundamental group functor as

$$\pi_n = \mathrm{Hom}_{\mathfrak{HTop}_*}((\mathbf{S}^n, s), -) : \mathfrak{HTop}_* \to \mathfrak{Grp}.$$

REMARK 4.5. The group $\pi_n(X, x)$ is abelian for $n \geq 2$ and independent of x when X is path-connected.

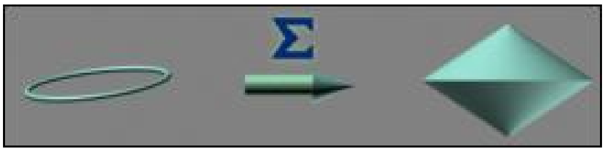
PROPOSITION 4.6. Let $p : \widetilde{X} \to X$ be the universal covering of a connected manifold X. Then $\pi_n(\widetilde{X})$ and $\pi_n(X)$ are isomorphic groups for any positive integer $n \geq 2$. In particular, the n-th homotopy group $\pi_n(\mathbf{S}^1) = \{0\}$ for any integer $n \geq 2$.

PROOF. Let $f : \mathbf{S}^n \to X$ be a morphism. By Theorem 2.37 and since $\mathbf{S}^n$ is simply connected for all $n \geq 2$, there exists a unique morphism $\widetilde{f} : \mathbf{S}^n \to \widetilde{X}$ such that $p \circ \widetilde{f} = f$, i.e. the following diagram

$$\begin{tikzcd} & \widetilde{X} \arrow[d, "p"] \\ \mathbf{S}^n \arrow[ur, dotted, "\widetilde{f}"] \arrow[r, "f"] & X \end{tikzcd}$$

is commutative. This shows that the induced morphism of groups $\pi_n(p) : \pi_n(\widetilde{X}) \to \pi_n(X)$ is bijective for all $n \geq 2$. The second part follows by the fact that the universal cover of $\mathbf{S}^1$ is $\mathbf{R}$. Since $\mathbf{R}$ is contractible, $\pi_n(\mathbf{R}) = 0$ for any positive integer n. More explicitly, let $f : \mathbf{S}^n \to \mathbf{R}$ be a morphism. Then the morphism $F : \mathbf{S}^n \times [0, 1] \to \mathbf{R}$ given by $F(x, t) = f(x)t$ is a homotopy from f to the constant map $\mathbf{S}^n \to \mathbf{R}$ given by $x \mapsto 0 \in \mathbf{R}$. $\square$

Definition 4.7. **(Suspension)** Let X be a topological space. We define an equivalence relation, denoted by $\sim$, on the product space $X \times [-1,1]$ by $(x,-1) \sim (y,-1)$ and $(x,1) \sim (y,1)$ for any x and y in X. The *suspension* of X, denoted by ΣX, is defined to be the quotient space $\Sigma X = X \times [-1,1]/\sim$. We denote the equivalence class of an element (x,t) in $X \times [-1,1]$ by $[x,t]$. The *suspension* of a morphism $f : X \to Y$, denoted by Σf, is the map $\Sigma f : \Sigma X \to \Sigma Y$ defined by $\Sigma f[x,t] = [f(x),t]$. Note that Σ is a functor from the category of topological spaces to itself.



Proposition 4.8. Let $\mathbf{S}^n$ be the unit n-dimensional sphere in $\mathbf{R}^{n+1}$. The suspension $\Sigma\mathbf{S}^n$ is homeomorphic to $\mathbf{S}^{n+1}$.

Proof. Let $\alpha : \mathbf{S}^n \times [-1,1] \to \mathbf{S}^{n+1}$ be the map defined by

$$\alpha(x_1,\dots,x_{n+1},t) = (\sqrt{1-t^2}x_1,\dots,\sqrt{1-t^2}x_{n+1},t).$$

It is clear that α is surjective and continuous. Since α is constant on the boundary of $[-1,1]$ it factors through the quotient. Since $\mathbf{S}^n$ and $[-1,1]$ are compact separated spaces α induces a homeomorphism $\Sigma\mathbf{S}^n \xrightarrow{\sim} \mathbf{S}^{n+1}$. $\square$

2. The Hopf map

Let $V = \mathbf{R}i \oplus \mathbf{R}j \oplus \mathbf{R}k$ be the space of pure quaternions. Let

$$\mathrm{Ad} : \mathrm{SU}_2(\mathbf{R}) \to \mathrm{SO}(V), \quad x \mapsto c_x$$

be the surjective map studied in Section 5. Let $\mathrm{Eval}_i : \mathrm{SO}(V) \to V$ be the evaluation map given by $f \mapsto f(i)$. The composition $\mathrm{Eval}_i \circ \mathrm{Ad}$ is denoted by $\hat{H}$. It is given by $\hat{H}(x) = xix^{-1}$ for any x in $\mathrm{SU}_2(\mathbf{R})$. Since i has unit norm the image of H is contained in the space of pure quaternions. It is clear that we have isomorphisms $\mathbf{S}^3 \simeq \mathrm{SU}_2(\mathbf{R})$ and $\mathbf{S}^2 \simeq \{\ x \in V \mid N(x) = 1\ \}$. The composition

$$\mathbf{R}^4 \supset \mathbf{S}^3 \xrightarrow{\sim} \mathrm{SU}_2(\mathbf{R}) \xrightarrow{\hat{H}} \{\ x \in V \mid N(x) = 1\ \} \xrightarrow{\sim} \mathbf{S}^2 \subset \mathbf{R}^3$$

is denoted by H and is called the *Hopf map.*

It is easy to see that the Hopf map $H : \mathbf{S}^3 \to \mathbf{S}^2$ can be given explicitly by $H(a,b,c,d) = (a^2+b^2-c^2-d^2, 2(ad-bc), 2(bd+ac))$.

View $\mathbf{S}^1$ as the unit circle in $\mathbf{C}^*$ and view $\mathbf{S}^3$ as the unit sphere in $\mathbf{C}^2$. Consider $\mathbf{P}^1_{\mathbf{C}}$ as the quotient space of $\mathbf{S}^3$ under the equivalence relation $(z,w) \sim \lambda(z,w)$ for any $\lambda \in \mathbf{S}^1$. Identifying $\mathbf{S}^2$ with $\mathbf{P}^1_{\mathbf{C}}$ the Hopf map can be taken to be the projection map $p : \mathbf{S}^3 \to \mathbf{P}^1_{\mathbf{C}}$ which sends (z,w) to its equivalence class $[z,w]$.

Identify $\mathbf{S}^2$ with the Riemann sphere $\mathbf{C} \cup \{\infty\}$. The Hopf map $H : \mathbf{S}^3 \to \mathbf{S}^2$ can be taken to be $H(z,w) = \frac{z}{w}$.

Definition 4.9. Let F be a manifold. A *fiber bundle* structure on a manifold E, with *fiber* F, consists of a surjective morphism of manifolds $p : E \to B$ such that each point of B has a neighborhood U for which there is a diffeomorphism $h : p^{-1}(U) \to U \times F$ making the diagram

$$\begin{tikzcd}
p^{-1}(U) \arrow[r, "h"] \arrow[dr, "p"'] & U \times F \arrow[d, "\mathrm{pr}_1"] \\
 & U
\end{tikzcd}$$

or equivalently the diagram

$$\begin{tikzcd} U \times F \arrow[r, "\sim"] \arrow[dr, "\mathrm{pr}_1"'] & U \times_B E \arrow[r] \arrow[d] & E \arrow[d] \\ & U \arrow[r] & B \end{tikzcd}$$

commutative. Here pr_1 is the projection on the first factor and $U \times_B E$ is the fiber product of U and E over B. The manifold B is called the *base space*, and E is called the *total space*. We usually denote fiber bundles as "exact sequences of manifolds": $F \to E \to B$.

REMARK 4.10. With notation as in Definition 4.9. Commutativity of the following diagram

$$\begin{tikzcd} p^-(U) \arrow[r, "h"] \arrow[dr, "p"'] & U \times F \arrow[d, "\mathrm{pr}_1"] \\ & U \end{tikzcd}$$

means that h carries each fiber $F_b = p^{-1}(b)$ diffeomorphically onto the copy $\{b\} \times F$ of F. Thus the fibers F_b are arranged locally as in the product of the bundle. Since the first coordinate of h is just p, h is determined by its second coordinate, a map $p^{-1}(U) \to F$ which is a diffeomorphism on each fiber F_b.

EXAMPLE 4.11. A fiber bundle with fiber a discrete space is a covering map. Conversely, a covering map whose fibers all have the same cardinality is a fiber bundle with discrete fiber.

LEMMA 4.12. The Hopf map $H : \mathbf{S}^3 \to \mathbf{S}^2$ is a fiber bundle with fiber $\mathbf{S}^1$. We denote this bundle by $\mathbf{S}^1 \to \mathbf{S}^3 \to \mathbf{S}^2$ and call it the *Hopf bundle*.

PROOF. Viewing the Hopf map $H : \mathbf{S}^3 \to \mathbf{S}^2$ as the projection $p : \mathbf{S}^3 \to \mathbf{P}^1_{\mathbf{C}}$ it is easy to see that the fibers are copies of $\mathbf{S}^1$. Define $U_i \subset \mathbf{P}^1_{\mathbf{C}}$ for $i = 1, 2$ to be the open set of equivalence classes $[z_1, z_2]$ with $z_i \neq 0$. Define $h_i : p^{-1}(U_i) \to U_i \times \mathbf{S}^1$ by $h_i(z_1, z_2) = \left([z_1, z_2], \frac{z_i}{|z_i|}\right)$. This map takes fibers to fibers, and is a diffeomorphism since its inverse is the map $([z_1, z_2], \lambda) \mapsto \lambda |z_i| z_i^{-1}(z_1, z_2)$. $\square$

REMARK 4.13. There are only three other fiber bundles with base, fiber and total space spheres. These are also called Hopf bundles. They are $\mathbf{S}^0 \to \mathbf{S}^1 \to \mathbf{S}^1$, $\mathbf{S}^3 \to \mathbf{S}^7 \to \mathbf{S}^4$ and $\mathbf{S}^7 \to \mathbf{S}^{15} \to \mathbf{S}^8$. For a proof we refer the reader to [**Adams**].

3. $\pi_4(\mathbf{S}^3) = \mathbf{Z}/2\mathbf{Z}$.

THEOREM 4.14. **(Freudenthal suspension theorem)** The suspension map $\pi_i(\mathbf{S}^n) \to \pi_{i+1}(\mathbf{S}^{n+1})$ is an isomorphism for $i < 2n - 1$ and a surjection for $i = 2n - 1$.

PROOF. [**Hat**, Corollary 4.24, pages 360-361]. $\square$

THEOREM 4.15. Every fiber bundle $F \to E \to B$ of connected manifolds gives rise to a long exact sequence of groups

$$\ldots \to \pi_n(F) \to \pi_n(E) \to \pi_n(B) \to \pi_{n-1}(F) \to \ldots \to \pi_1(B) \to \{0\}.$$

PROOF. [**Hat**, Theorem 4.41 and Proposition 4.48, pages 376-380]. $\square$

THEOREM 4.16. The homotopy group $\pi_3(\mathbf{S}^2)$ is isomorphic to $\mathbf{Z}$ and generated by the homotopy class of H.

PROOF. By Theorem 4.15, the Hopf bundle gives a long exact sequence:

$$\ldots \to \pi_n(\mathbf{S}^1) \to \pi_n(\mathbf{S}^3) \to \pi_n(\mathbf{S}^2) \to \pi_{n-1}(\mathbf{S}^1) \to \ldots \to \pi_1(\mathbf{S}^1) \to \{0\}.$$

By this long exact sequence and Proposition 4.6, we have the exact sequence

$$0 \to \pi_2(\mathbf{S}^3) \to \pi_2(\mathbf{S}^2) \to \pi_1(\mathbf{S}^1) \to \pi_1(\mathbf{S}^3) \to \ldots.$$

Since $\mathbf{S}^3$ is simply connected, $\pi_1(\mathbf{S}^3) = \{0\}$. By the Freudenthal suspension theorem the suspension map induces an isomorphism $\pi_1(\mathbf{S}^2) \simeq \pi_2(\mathbf{S}^3)$. Since $\mathbf{S}^2$ is simply connected we may conclude that $\pi_2(\mathbf{S}^3) = \{0\}$. We get a short exact sequence

$$0 \to \pi_2(\mathbf{S}^2) \to \pi_1(\mathbf{S}^1) \to 0.$$

This implies that $\pi_2(\mathbf{S}^2) \simeq \pi_1(\mathbf{S}^1) \simeq \mathbf{Z}$. Now, consider the suspension map $\pi_2(\mathbf{S}^2) \to \pi_3(\mathbf{S}^3)$. By the Freudenthal suspension theorem this map is an isomorphism. We conclude that $\pi_3(\mathbf{S}^3) = \mathbf{Z}$. Again by the above long exact sequence and Proposition 4.6, we have the short exact sequence of abelian groups

$$0 \to \pi_3(\mathbf{S}^3) \to \pi_3(\mathbf{S}^2) \to 0.$$

Hence $\pi_3(\mathbf{S}^2) \simeq \pi_3(\mathbf{S}^3) = \mathbf{Z}$ and $\pi_3(\mathbf{S}^2)$ is generated by the Hopf map H.[1] This completes the proof of the theorem. □

THEOREM 4.17. The image of H under the suspension map $\pi_3(\mathbf{S}^2) \to \pi_4(\mathbf{S}^3)$ generates $\pi_4(\mathbf{S}^3)$ and is of order two.

PROOF. Apply the Freudenthal suspension theorem to see that the suspension map $\pi_3(\mathbf{S}^2) \to \pi_4(\mathbf{S}^3)$ is surjective (it is not an isomorphism). Therefore $\pi_4(\mathbf{S}^3)$ is a cyclic group generated by the image of the Hopf map under the suspension map $\pi_3(\mathbf{S}^2) \to \pi_4(\mathbf{S}^3)$. It remains to show that the image of H under this suspension map is of order two in $\pi_4(\mathbf{S}^3)$. This is left to the reader as a (very difficult) exercise. □

[1] Actually it is not clear why the Hopf map generates $\pi_3(\mathbf{S}^2)$. Let us just remark that it follows from the definition of the connection maps in the long exact sequence.

CHAPTER 5

Introduction to Yang-Mills theories

1. Intermezzo

We have reached the "physical" part of this thesis. In this short section we briefly note on notation that is commonly used in physics. We will also give a theorem that links matrix groups to their Lie algebra through the matrix exponential.

Let us first note that from now on all Lie groups are defined over real manifolds. Thus when we speak of a Lie group we shall mean a real Lie group.

The notation we used for Lie groups such as the unimodular group $\mathrm{U}_1(\mathbf{R})$ will no longer be used. Instead we shall denote the $U_1(\mathbf{R})$ by $U(1)$ and denote the group $\mathrm{SU}_2(\mathbf{R})$ by $SU(2)$.

Since we shall only work with matrix groups we briefly introduce the (matrix) exponential map in order to easily communicate between a Lie group and its Lie algebra. For any $(n \times n)$-matrix A we define the exponential of A to be

$$e^A = \exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{6} + \ldots = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

We shall not concern ourselves to convergence problems. We do note that $\exp(A)$ is well defined for any matrix A.

Let O be the zero matrix. It is easy to see that $\exp(O) = 1$. If the matrices A and B commute, then

$$\exp(A + B) = \exp(A)\exp(B).$$

We will just indicate a proof by looking at the first few terms.

$$\begin{aligned}
\exp(A+B) &= I + A + B + \frac{A^2}{2} + AB + \frac{B^2}{2} + \frac{A^3}{6} + \frac{A^2B}{2} + \frac{B^3}{6} + \ldots \\
\exp(A)\exp(B) &= (I + A + \frac{A^2}{2} + \frac{A^3}{6} + \ldots)(I + B + \frac{B^2}{2} + \frac{B^3}{6} + \ldots) \\
&= I + A + B + \frac{A^2}{2} + AB + \frac{B^2}{2} + \frac{A^3}{6} + \frac{A^2B}{2} + \frac{B^3}{6} + \ldots
\end{aligned}$$

Note that this implies that $\exp(A)$ is nonsingular with inverse $\exp(-A)$ for any $(n \times n)$-matrix. Therefore exp maps the Lie algebra of all $(n \times n)$-matrices on the Lie group of invertible matrices.

When two matrices A and B do not commute one must invoke the so called Campbell-Hausdorff[1] formula. The Campbell-Hausdorff formula gives a formula for C where $\exp(A)\exp(B) = \exp(C)$, in terms of repeated commutators of A and B such as $[A, B]$, $[A, [A, B]]$, $[B, [A, B]]$, $[A, [B, [A, B]]]$, etc. We give the formula up to order 3. This will give us the insight that the Lie bracket is, up to a factor $\frac{1}{2}$, the quadratic term.

Theorem 5.1. For any two $(n \times n)$-matrices A and B it holds that $\exp(A)\exp(B) = \exp(C)$, where

$$C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] + [B, [B, A]]) + \ldots.$$

Proof. [**Serre**]. The proof given there applies to any Lie group. □

[1] In some texts this formula is also called the Baker-Campbell-Hausdorff formula.

The following theorem links a Lie group to its Lie algebra by using the exponential map.

THEOREM 5.2. Let $G \subset \mathrm{GL}_n(\mathbf{R})$ be a (matrix) Lie group. The Lie algebra of G is given by $\{X \in M_n(\mathbf{R}) |\, \exp(tX) \in G \text{ for all } t \in \mathbf{R}\}$. A similar statement holds for a Lie group $G \subset \mathrm{GL}_n(\mathbf{C})$.

PROOF. [**Edix**, Chapter 3, pages 27-30]. The proof given there applies to any Lie group. □

An important example of a Lie group is $SU(2)$. In the preceding chapters we saw that the (mathematical) Lie algebra of $SU(2)$ is the real vector space with basis $(i\sigma^3, i\sigma^2, i\sigma^1)$. Here

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli-matrices. This is easily confirmed by Theorem 5.2 when one uses the equations

$$AA^\dagger = A^\dagger A, \quad \exp(A)^\dagger = \exp(A^\dagger), \quad \det(\exp(tA)) = \exp(t\mathrm{tr}(A)).$$

The generators of $SU(2)$ are given by $(-\frac{1}{2}\sigma^3, -\frac{1}{2}\sigma^2, -\frac{1}{2}\sigma^1)$ and can be replaced by $(\frac{1}{2}\sigma^1, \frac{1}{2}\sigma^2, \frac{1}{2}\sigma^3)$. It is easy to see that the structure constants are given by $f^{ijk} = \epsilon^{ijk}$.

Now, since we shall not be doing any explicit calculations it is easier and more convenient to let $\hbar = c = 1$.

We conclude with some final remarks on prerequisites. We shall assume the reader to have some familiarity with classical electrodynamics, relativity and classical quantum mechanics. We refer the reader to [**Grif1**] and [**Grif2**] for a basic overview. Furthermore, the Einstein summation convention is used. We also assume the reader to be comfortable with representation theory since we shall abuse the standard language sometimes. For example, when we say that x is an element of the adjoint representation of $SU(2)$ we shall mean that x is an element of $\mathbf{C}^3$ transforming under the adjoint action of $SU(2)$.

2. Electrodynamics and relativity

In this section we give a formulation of Maxwell's equations in terms of the electromagnetic field tensor $F_{\mu\nu}$. In this formulation it is clear that the Maxwell equations are Lorentz-invariant and above all are gauge invariant.

Since we shall only consider the electromagnetic field in vacuum we use the so-called rationalized cgs units where $\epsilon_0 = 1$ and $\mu_0 = 1$.

Let

$$\mathbf{E} = \mathbf{E}(\mathbf{x}, t) = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}, \text{ respectively } \quad \mathbf{B} = \mathbf{B}(\mathbf{x}, t) = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix},$$

be the electric, respectively magnetic, field. The homogeneous Maxwell equations are given by

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0.$$

The first equation excludes the existence of magnetic monopoles while the second expresses Faraday's law of induction.

Let $\rho = \rho(\mathbf{x}, t)$ be the electric charge density and let $\mathbf{J} = \mathbf{J}(\mathbf{x}, t)$ be the current density. The inhomogeneous Maxwell equations are given by

$$\nabla \cdot \mathbf{E} = \rho, \quad \nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{J}.$$

It is well known that the continuity equation $\dfrac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$ is a consequence of the inhomogeneous Maxwell equations. Moreover, Gauss's law follows by applying the integral theorem of Gauss to the equation $\nabla \cdot \mathbf{E} = \rho$. The homogeneous and inhomogeneous Maxwell equations give a complete description of classical electrodynamics when we add Lorentz's force law $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$.

To solve the homogeneous Maxwell equations we can express the electric field and the magnetic field in terms of a scalar potential $\phi = \phi(\mathbf{x}, t)$ and a vector potential $\mathbf{A} = (A_x, A_y, A_z)$. Indeed, the homogeneous equations are satisfied for

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}.$$

We have a four-vector potential $A_\mu = (\mathbf{A}, i\phi) = (A_x, A_y, A_z, i\phi)$ which possesses a certain *gauge* freedom. Under a *gauge transformation*

$$A_\mu \to A_\mu + \partial_\mu \xi(x)$$

the electric and magnetic fields are unchanged. The parameter $\xi(x)$ in this *gauge transformation* is an arbitrary function of $\mathbf{x}$ and t. Due to this gauge freedom $A_\mu(x)$ is generally called a *gauge field* and Maxwell's theory of electromagnetism is just the simplest example of a gauge theory. Under the same gauge transformation the *field tensor* $F_{\mu\nu}$ defined by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (\mu, \nu = 1, 2, 3, 4)$$

is also unchanged. Therefore it might be interesting to try and understand $F_{\mu\nu}$ more.

It is clear that $F_{\mu\nu}$ is antisymmetric, i.e. $F_{\nu\mu} = -F_{\mu\nu}$. In particular, $F_{\mu\mu} = 0$ for all $\mu = 0, 1, 2, 3$. Furthermore, it is easy to see that

$$F_{12} = \partial_1 A_2 - \partial_2 A_1 = (\nabla \times \mathbf{A})_z = B_z, \quad F_{41} = \partial_4 A_1 - \partial_1 A_4 = i(-\nabla\phi - \frac{\partial \mathbf{A}}{\partial t})_x = iE_1.$$

Determining all other components of $F_{\mu\nu}$ is completely similar. Since $F_{\mu\nu}$ is a second-rank tensor we may express it as a matrix

$$F^{\mu\nu} = \begin{pmatrix} 0 & B_3 & -B_2 & -iE_1 \\ -B_3 & 0 & B_1 & -iE_2 \\ B_2 & -B_1 & 0 & -iE_3 \\ iE_1 & iE_2 & iE_3 & 0 \end{pmatrix}.$$

One can check that the homogeneous Maxwell equations can be obtained from the equations

$$\partial_\mu F_{\mu\lambda} + \partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} = 0.$$

Introducing the four-vector current $J^\mu = (\mathbf{J}, i\rho) = (J_x, J_y, J_z, i\rho)$ we can easily check that the equation $\partial_\nu F_{\mu\nu} = J^\mu$ implies the inhomogeneous Maxwell equations. Note that under the assumption that A^μ and J^μ transform as four-vectors it is immediate that the Maxwell equations are Lorentz invariant.

Our next objective is to study quantum electrodynamics, i.e. the unification of Maxwell's theory and the theory of particles with spin $\frac{1}{2}$. We will see that quantum electrodynamics is a "simple" case of a *Yang-Mills theory.*

3. Particles with spin $\frac{1}{2}$

Particles with spin $\frac{1}{2}$ obey the Dirac equation. The solutions to the Dirac equation are called Dirac spinors. The Lagrangian that produces the Dirac equation will turn out to play an important role later on.

In particle mechanics one introduces Lagrangians L as functions of the (generalized) coordinates q_i and their time derivatives $\dot{q}_i$. In field theory we work with a Lagrangian (technically, a Lagrangian *density*) $\mathcal{L}$, which is a function of the fields ϕ_i and their x, y, z and t derivatives: $\partial_\mu \phi_i$. The Euler-Lagrange equations generalize to

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} = \frac{\partial \mathcal{L}}{\partial \phi_i} \quad (i = 1, 2, \ldots).$$

We are going to give the Lagrangian that produces the so called Dirac equation. The Dirac equation describes (in quantum field theory) a particle of spin $\frac{1}{2}$ and mass m. Before giving the Dirac equation we need to introduce the gamma matrices.

The *gamma matrices* are denoted by γ^0, γ^1, γ^2 and γ^3. They are (4×4)-matrices that satisfy

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}.$$

Here

$$g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

is the Minkowski metric and curly brackets denote the anticommutator: $\{A, B\} = AB + BA$.

We are now ready to give the **Dirac equation**:

$$i\gamma^\mu \partial_\mu \psi - m\psi = 0.$$

Here m denotes the mass of the particle and the wave function ψ is a four-element column matrix

$$\psi = \psi_\alpha = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}.$$

We call it a *Dirac spinor*.

If ψ is a Dirac spinor one might suspect ψ to be a four-vector, but this is not the case. Since we shall not need it, the explicit (relativistic) transformation formula for ψ when you go from one inertial system to another will not be given here. We just note that under a Lorentz transformation it transforms as the the four-dimensional $\frac{1}{2} \oplus \frac{1}{2}$ representation of Lie$(SL_2(\mathbf{C})$. More on this in section 7. Note that when we let $\psi^\dagger$ be the Hermitian conjugate of ψ and define the *adjoint* spinor of ψ by[2]

$$\overline{\psi} = (\psi^\dagger)^\alpha (\gamma^0)_\alpha$$

we can form a relativistic invariant (Lorentz scalar): $\overline{\psi}_\alpha \psi_\alpha = \overline{\psi}\psi$. Of course we can not justify this without the explicit transformation formula. In the *standard* representation of the gamma matrices[3], i.e.

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^k = \begin{pmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{pmatrix},$$

this reads $\overline{\psi}\psi = \mid \psi_1 \mid^2 + \mid \psi_2 \mid^2 - \mid \psi_3 \mid^2 - \mid \psi_4 \mid^2$ since $\overline{\psi}_\alpha = \overline{\psi} = (\psi_0^*, \psi_1^*, -\psi_2^*, -\psi_3^*)$ in the standard representation. Another useful hermitian (4×4)-matrix is $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ which satisfies $\{\gamma^\mu, \gamma^5\} = 0$ and may be expressed as

$$\gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

in the standard representation. Introducing the matrices

$$\sigma^{\mu\nu} = -\frac{1}{2} i [\gamma_\mu, \gamma_\nu]$$

we see that $(1, \gamma^5, \gamma^\mu, \gamma^\mu\gamma^5, \sigma^{\mu\nu})$ is a basis for the vector space of (4×4) complex matrices.

Note we have another relativistic invariant given by

$$\overline{\psi}\gamma^5\psi = \overline{\psi}_\alpha (\gamma^5)_{\alpha\beta} \psi_\beta = -\psi_0^*\psi_3 - \psi_1^*\psi_4 + \psi_2^*\psi_1 + \psi_3^*\psi_2.$$

Now, consider the (relativistic) Lagrangian

$$\mathcal{L}_{Dirac} = \overline{\psi} i \gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi.$$

[2]Note that the notation $\overline{\psi}$ was used to denote the complex conjugate. Here it is something completely different.

[3]Here 1 denotes the (2×2) identity matrix and σ^k are the Pauli spin matrices.

We treat ψ and $\overline{\psi}$ as independent field variables. This is exactly what characterizes Dirac fermions. Applying the (generalized) Euler-Lagrange equations to $\overline{\psi}$, we find

$$\frac{\partial \mathcal{L}_{Dirac}}{\partial(\partial_\mu \overline{\psi})} = 0, \quad \frac{\partial \mathcal{L}_{Dirac}}{\partial \overline{\psi}} = \gamma^\mu i \partial_\mu \psi - m\psi$$

so that

$$\gamma^\mu i \partial_\mu \psi - m\psi = 0.$$

This is the Dirac equation. Meanwhile, if we apply the Euler-Lagrange equations to ψ we obtain the adjoint of the Dirac equation:[4]

$$\overline{\psi}(-i\gamma^\mu \overleftarrow{\partial_\mu} - m) = 0.$$

We conclude this section by making the observation that the Lagrangian $\mathcal{L}_{Dirac}$ is invariant under the following phase transformations

$$\begin{aligned} \psi &\rightarrow e^{i\alpha}\psi, \\ \overline{\psi} &\rightarrow \overline{\psi}e^{-i\alpha}. \end{aligned}$$

Here α is an arbitrary real number. Hence the Lagrangian $\mathcal{L}_{Dirac}$ is invariant under the action of the group $U(1)$ on the fields ψ and $\overline{\psi}$. Recall that $U(1)$ is homeomorphic to the unit circle $\mathbf{S}^1$ in $\mathbf{C}^*$. Since this symmetry is independent of the position we say that this is a *global* symmetry of $\mathcal{L}$.

4. The electromagnetic field

In this section we give the field theoretic Lagrangian that produces the Maxwell equations in vacuum. For completeness we also give the Maxwell Lagrangian for a massless vector field with source.

The electromagnetic field A_μ is a massless vector field. Instead of just giving this Lagrangian we start with Lagrangian that describes particles with spin 1 and mass m. That is, we look at vector fields V_μ with mass m and free Lagrangian given by the so called *Proca Lagrangian*

$$\mathcal{L}_{Proca} = -\frac{1}{4}(\partial_\mu V_\nu - \partial_\nu V_\mu)^2 - \frac{1}{2}m^2 V_\mu^2.$$

It is well known (See [**WitSm**]) that this Lagrangian describes a particle of spin 1 with mass m. The Euler-Lagrange equations yield the so called *Proca equation*

$$\partial_\mu(\partial_\mu V_\nu - \partial_\nu V_\mu) + m^2 V^\nu = 0.$$

Now, the idea is to take the massless limit and see if this leads us to Maxwells equations. Putting $m = 0$ and identifying V_μ with the electromagnetic field A_μ one gets the Maxwell equations for empty space, i.e.

$$\mathcal{L} = -\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 = -\frac{1}{4}F_{\mu\nu}^2$$

is the Maxwell Lagrangian. When one introduces a source J^μ the complete Lagrangian would be

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 - J^\mu A_\mu.$$

The Euler-Lagrange equations correctly yield the Maxwell equations given by

$$\partial_\nu F_{\mu\nu} = J^\mu.$$

In the following section the goal will be to merge the Lagrangian for a spin $\frac{1}{2}$ particle and the Lagrangian for the electromagnetic field. This way we will be able to give a description of the interaction of Dirac fields with Maxwell fields. First we will need the notion of renormalization.

---

[4] The arrow on the differential operator means one has to apply the operator to the function on the left, i.e. $f\overleftarrow{\partial_\mu} = \partial_\mu f$.

5. Gauging the Dirac equation

We already saw that the Lagrangian

$$\mathcal{L} = \overline{\psi} i\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi.$$

produces the Dirac equation. We also observed that $\mathcal{L}$ is invariant under the (simultaneous) global $U(1)$ transformations

$$\begin{aligned} \psi &\rightarrow e^{i\alpha}\psi, \\ \overline{\psi} &\rightarrow \overline{\psi} e^{-i\alpha}. \end{aligned}$$

By "gauging" the Dirac equation we mean that we look at how we have to change our Lagrangian in order to be invariant under the local phase transformation (or "gauge transformation") given by

$$\begin{aligned} \psi &\rightarrow e^{i\alpha(x)}\psi, \\ \overline{\psi} &\rightarrow \overline{\psi} e^{-i\alpha(x)}. \end{aligned}$$

We shall see that the result will turn out to be very interesting.

Our goal[5] is to construct the most general renormalizable[6] Lagrangian invariant under the (simultaneous) *local* transformations

$$\begin{aligned} \psi &\rightarrow e^{i\alpha(x)}\psi, \\ \overline{\psi} &\rightarrow \overline{\psi} e^{-i\alpha(x)}, \end{aligned}$$

starting from the Dirac Lagrangian. The difference here with the global symmetries being that the phase factor α may depend on x. It is easily seen that $\mathcal{L}$ is not invariant under the local transformations

$$\begin{aligned} \psi &\rightarrow e^{i\alpha(x)}\psi, \\ \overline{\psi} &\rightarrow \overline{\psi} e^{-i\alpha(x)} \end{aligned}$$

since we pick up a term $\sim \partial_\mu \alpha(x)$. More precisely,

$$\mathcal{L} \rightarrow \mathcal{L} - \partial_\mu(\alpha(x))\overline{\psi}\gamma^\mu\psi.$$

Now the remedy for our Lagrangian $\mathcal{L}$ to be invariant under the local phase transformation (also called gauge transformation) is to add a new field A_μ, a so called *gauge* field, in such a way that the new Lagrangian

$$\mathcal{L} = [\overline{\psi} i\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi] + (q\overline{\psi}\gamma^\mu\psi)A_\mu$$

is invariant under the gauge transformation. Note that we added a *coupling constant* q. Also note the similar notation we used for the vector potential in the section on electrodynamics. By letting $D_\mu = \partial_\mu - iqA_\mu$ we can write our Lagrangian as

$$\mathcal{L} = \overline{\psi} i\gamma^\mu D_\mu \psi - m\overline{\psi}\psi$$

showing the resemblance with the Lagrangian we started with. The combination $D_\mu = \partial_\mu - iqA_\mu$ is called the *covariant derivative.* The name is well chosen since the transformation law for D_μ (actually $D_\mu\psi$) should be the same as the one for ψ if we want our Lagrangian to be gauge invariant, i.e. D_μ should transform as

$$D_\mu\psi \rightarrow e^{i\alpha(x)} D_\mu \psi$$

or in operator language

$$D_\mu \rightarrow e^{i\alpha(x)} D_\mu e^{-i\alpha(x)}.$$

The covariance requirement on D_μ

$$D_\mu \rightarrow D'_\mu = \partial_\mu + iA'_\mu = e^{i\alpha(x)}(\partial_\mu + iA_\mu)e^{-i\alpha(x)}$$

[5]Apparently "localizing" global symmetries might be interesting.

[6]Informally speaking, a theory is renormalizable if one can get rid of all infinities that appear at each order of perturbation theory.

becomes a transformation property of A_μ. We conclude that the gauge field A_μ transforms as

$$A_\mu \to A_\mu + \frac{1}{q}\partial_\mu \alpha(x)$$

which is exactly what we need for our "new and improved" Lagrangian to be invariant under the gauge transformation.

In order to clarify the notion of the covariant derivative more we briefly take on a more geometric approach. We introduce a scalar function $U(x,y)$ which transforms as

$$U(x,y) \to \exp(i\alpha(x))U(x,y)\exp(-i\alpha(y))$$

and satisfies $U(x,x) = 1$. We call U a *comparator.* Given a vector n^μ, we look for the generalization of a "directional derivative" for a Dirac field ψ in the direction of n^μ. The ordinary definition is

$$n^\mu \partial_\mu \psi = \lim_{\epsilon\to 0} \frac{1}{\epsilon}(\psi(x+\epsilon n) - \psi(x)).$$

As a pure formal statement, an interesting object would be

$$n^\mu D_\mu \psi = \lim_{\epsilon\to 0} \frac{1}{\epsilon}(\psi(x+\epsilon n) - U(x+\epsilon n, x)\psi(x)).$$

Note that we added the comparator to this definition in order to make ψ and the object $n^\mu D_\mu \psi$ have the same transformation law. Expanding in the small parameter ϵ gives us

$$U(x+\epsilon n, x) = U(x,x) + \epsilon n^\mu \partial_\mu U(x+\epsilon n, x)\mid_{\epsilon=0} + \mathcal{O}(\epsilon^2) = 1 + \epsilon n^\mu \partial_\mu U(x+\epsilon n, x)\mid_{\epsilon=0} + \mathcal{O}(\epsilon^2).$$

Introducing a new vector field A_μ satisfying $\epsilon n^\mu \partial_\mu U(x+\epsilon n, x)\mid_{\epsilon=0} = iq\epsilon n^\mu A_\mu$, it holds that

$$\begin{aligned} n^\mu D_\mu \psi &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}(\psi(x+\epsilon n) - U(x+\epsilon n, x)\psi(x)) \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}(\psi(x+\epsilon n) - (1 + \epsilon n^\mu \partial_\mu U(x+\epsilon n, x)\mid_{\epsilon=0} + \mathcal{O}(\epsilon^2))\psi(x)) \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}(\psi(x+\epsilon n) - \psi(x)) - iqn^\mu A_\mu \psi = n^\mu(\partial_\mu - iqA_\mu)\psi. \end{aligned}$$

By construction $D_\mu = \partial_\mu - iqA_\mu$ has the same transformation law as ψ. We have therefore constructed the covariant derivative in a much more geometric setting. In the language of differential geometry, the vector field A_μ is known as a *connection.* The transformation law for A_μ follows from the transformation law for the comparator. We have that

$$U(x+\epsilon n, x) \to e^{i\alpha(x+\epsilon n)}U(x+\epsilon n, x)e^{-i\alpha(x)} = 1 - iq\epsilon n^\mu\left(A_\mu + \frac{1}{q}\partial_\mu\alpha(x)\right) + \mathcal{O}(\epsilon^2).$$

From this it follows that

$$A_\mu \to A_\mu + \frac{1}{q}\partial_\mu \alpha(x).$$

So we found a "better" Lagrangian which is invariant under any gauge transformation. The covariant derivative along with the gauge field were introduced ad hoc and from a geometric point of view.

Now let's "improve" our Lagrangian even more. Let's turn our attention momentarily to vector fields. In particular, consider a vector field V_μ whose free Lagrangian is given by the Proca Lagrangian

$$\mathcal{L}_{Proca} = -\frac{1}{4}(\partial_\mu V_\nu - \partial_\nu V_\mu)^2 - \frac{1}{2}m^2 V_\mu^2.$$

Now, it is easily seen that the massless limit $-\frac{1}{4}(\partial_\mu V_\nu - \partial_\nu V_\mu)^2$ of $\mathcal{L}_{Proca}$ is invariant under the gauge transformation and that the mass term violates gauge invariance. Going back to our gauge field A_μ, which is a vector field, and introducing the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

it is easy to see that we can add a "free" part to our Lagrangian to get a "newer and improved" Lagrangian

$$\mathcal{L} = \overline{\psi} i\gamma^\mu D_\mu \psi - m\overline{\psi}\psi \ - \frac{1}{4} F_{\mu\nu}F^{\mu\nu}.$$

The reason we mention the Proca Lagrangian is to emphasize the fact that there could be a possibility of our gauge field to be taken massive, but A_μ is forced to be massless in order for gauge invariance of the Lagrangian to hold.

Can we improve our Lagrangian any further? The answer to this question is no. The proof is based on the fact that operators of dimension larger than four enforce coupling constants with negative mass dimension. And since any theory with coupling constants of negative mass dimension is non-renormalizable we can not add any more (nontrivial) terms to our Lagrangian.

The last term we might add to our Lagrangian would be $\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu}$, which is invariant under the gauge transformation and is also renormalizable. But this term is not invariant under parity and therefore we choose not to add it to our Lagrangian.

We conclude that the most general Lagrangian invariant under parity and the gauge transformation

$$\begin{array}{ccc} \psi & \rightarrow & e^{i\alpha(x)}\psi, \\ \overline{\psi} & \rightarrow & \overline{\psi}e^{-i\alpha(x)}, \\ A_\mu & \rightarrow & A_\mu + \dfrac{1}{q}\partial_\mu \alpha(x), \end{array}$$

is given by

$$\mathcal{L} = \overline{\psi} i\gamma^\mu D_\mu \psi - m\overline{\psi}\psi \ - \frac{1}{4} F_{\mu\nu}F^{\mu\nu}.$$

Now, what does this Lagrangian describe? The answer is beautiful and very surprising. The Lagrangian produces the Maxwell equations, with the four-current being $J^\mu = q(\overline{\psi}\gamma^\mu\psi)$, and the Dirac equation. Thus we have found the Lagrangian for quantum electrodynamics: electrons and positrons interacting with photons! Or better said, Dirac fields interacting with Maxwell fields:

$$\mathcal{L} = \underbrace{\overline{\psi} i\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi}_{Dirac} + \underbrace{\frac{-1}{4} F_{\mu\nu}F^{\mu\nu}}_{Maxwell} + \underbrace{(q\overline{\psi}\gamma^\mu\psi)A_\mu}_{Interaction}.$$

Note that our coupling constant q comes into play here. For electrons $q = -e$ and for protons $q = +e$. The gauge field A_μ is the wave function of the photon.

Recall that we started from the observation that the action of $U(1)$ on the fields ψ and $\overline{\psi}$ left the Dirac Lagrangian (globally) invariant. Then we made the extra assumption that this invariance should hold locally, i.e. we localized the global symmetry, and we altered our Lagrangian accordingly. Then we saw that the "new and improved" Lagrangian produces quantum electrodynamics. This was our first example of a *Yang-Mills theory*: the $U(1)$ Yang-Mills theory.

The physics behind what we just did can be seen as follows. We started with the Dirac Lagrangian which basically describes all matter. In particular, electrons and protons. One then sees that we can add a free part to this Lagrangian which produces the Maxwell equations, i.e. the electromagnetic field. Therefore $U(1)$ can be identified with the electromagnetic field. When we impose gauge invariance we see that we must add a gauge field (or force carrier) which is just the electromagnetic force. Therefore we could state that gauge invariance leads us to forces. The idea of the next section will be to impose gauge invariance on the action of $SU(2)$ on two-particle fields. This will lead us to a description of the electroweak force. In short, $U(1)$ is the electromagnetic field and gauge invariance under $U(1)$ gives us the electromagnetic force.

6. $SU(2)$ Yang-Mills theory

In this section we introduce Yang-Mills theories. We do this from a general point of view, i.e. we give a mathematical approach. Every Yang-Mills theory is characterized by a gauge group which *is always a compact Lie group. There are two gauge groups of major interest to us and those are $U(1)$ and $SU(2)$.*

Our goal will be to give a general description of a Yang-Mills theory. To make this general idea of Yang-Mills theories clear we study the important example of a $SU(2)$ Yang-Mills theory. Firstly, let us motivate why an $SU(2)$ Yang-Mills theory would be interesting.

Suppose that we have two spin-$\frac{1}{2}$ fields ψ_1 and ψ_2 with mass m_1 and mass m_2, respectively. The Lagrangian, in the absence of any interactions, is given by

$$\mathcal{L} = [\overline{\psi_1} i\gamma^\mu \partial_\mu \psi_1 - m_1 \overline{\psi_1}\psi_1] + [\overline{\psi_2} i\gamma^\mu \partial_\mu \psi_2 - m_2 \overline{\psi_2}\psi_2].$$

Applying the Euler-Lagrange equations to $\mathcal{L}$ shows that ψ_1 and ψ_2 obey the Dirac equation with the appropriate mass. By combining ψ_1 and ψ_2 into a two-component column vector[7]

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

we can write our Lagrangian much more compactly

$$\mathcal{L} = \overline{\psi} i\gamma^\mu \partial_\mu \psi - \overline{\psi} M \psi.$$

Here $\overline{\psi}$ is the adjoint spinor given by

$$\overline{\psi} = (\overline{\psi_1}\ \overline{\psi_2})$$

and M denotes the so called mass matrix given by

$$M = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}.$$

Now, it starts to get interesting when we assume m_1 and m_2 to be equal. We speak of a so called *doublet* ψ and our Lagrangian equals

$$\mathcal{L} = \overline{\psi} i\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi.$$

But this *looks* exactly like the one-particle Dirac Lagrangian we had been studying earlier.

We already saw that the one-particle Dirac Lagrangian is globally invariant under the action of $U(1)$ on the (one-particle) fields ψ and $\overline{\psi}$, i.e. under the global phase transformation given by

$$\psi \to U\psi$$

with U in $U(1)$. Our new (two-particle Dirac) Lagrangian is invariant under the action of $U(2)$ on the (two-particle) fields ψ and $\overline{\psi}$, i.e. under the global phase transformation given by

$$\psi \to U\psi$$

with U in $U(2)$. This is easily seen from the fact that any unitary matrix satisfies $U^\dagger U = 1$ and therefore the transformation

$$\overline{\psi} \to \overline{\psi} U^\dagger$$

leaves the combination $\psi\overline{\psi}$ invariant. Analogously to the $U(1)$ Yang-Mills theory, the idea will be to gauge this global invariance and construct a (most general) Lagrangian invariant under gauge transformations.

Recall that for any unitary matrix U there exists an hermitian matrix H such that[8] $U = e^{iH}$. Also, since any hermitian matrix H can be written as $H = \theta 1 + a^k \sigma^k$ we may write U as

$$U = e^{i\theta} e^{ia^k \sigma^k}.$$

Since we have already explored the implications of phase transformations given by $e^{i\theta}$ all there remains is to look at the global $SU(2)$ transformations

$$\psi \to e^{ia^k \sigma^k} \psi.$$

The latter is exactly what Yang and Mills did in their original paper [**YaMi**].

[7] Note that ψ actually consists of 8 entries. Therefore we could use a double index notation to label the particle and the spinor component. We choose not to do this since it is not necessary for our purposes.

[8] The Lie algebra of $U(2)$ is given by the space of (2×2) **anti**-hermitian matrices, i.e. $\{X \in M_2(\mathbf{C}) \mid X^\dagger = -X\}$.

Let us begin constructing the most general renormalizable Lagrangian invariant under $SU(2)$ gauge transformations and parity. Let us recall some basic facts about $SU(2)$.

We have already seen in Section 4 of Chapter 3 that the generators of $SU(2)$ are given by

$$T^1 = \frac{1}{2}\sigma^1, \quad T^2 = \frac{1}{2}\sigma^2, \quad T^3 = \frac{1}{2}\sigma^3.$$

The Lie bracket is given by

$$[T^1, T^2] = iT^3, \quad [T^2, T^3] = iT^1, \quad [T^3, T^1] = iT^2$$

as can be easily checked. By definition, for any X in $SU(2)$ there exist real numbers θ_1, θ_2 and θ_3 such that $X = \exp(i\theta_1 T_1 + i\theta_2 T_2 + i\theta_3 T_3)$.

Let

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

be a doublet, which transforms as

$$\psi \to \exp\left(i\theta^k \frac{\sigma^k}{2}\right)\psi.$$

Note the resemblance with the previous section where the transformation looked like $\psi \to e^{i\alpha}\psi$. The exponential map is always used to write the action of a Lie group in terms of the generators of the Lie algebra. In this case there are three generators and in the case of $U(1)$ there was only one generator. The idea is again to extend this global symmetry to a local symmetry by letting $\theta^k \to \theta^k(x)$. This means that the fields transform according to

$$\psi \to \exp\left(i\theta^k(x)\frac{\sigma^k}{2}\right)\psi.$$

Noting that the Lagrangian is not invariant under this gauge transformation we start with constructing the covariant derivative. Using the geometric approach this goes completely analogous to just before with the only difference being that the comparator $U(x, y)$ is now a (2×2)-matrix, since ψ is a two-component object. Letting $V(x) = \exp\left(i\theta^k(x)\frac{\sigma^k}{2}\right)$ the comparator has the transformation property $U(x, y) \to V(x)U(x, y)V(y)^\dagger$ and satisfies $U(x, x) = 1$. Here 1 denotes the (2×2) identity matrix. Expanding the comparator in ϵ we find

$$U(x, y) = 1 + ig\epsilon n^\mu A^k_\mu \frac{\sigma^k}{2} + \mathcal{O}(\epsilon^2),$$

where g is a (coupling) constant. Thus, the covariant derivative becomes $D_\mu = \partial_\mu - igA^k_\mu \frac{\sigma^k}{2}$. We see that we have three gauge fields, i.e. every generator of the group $SU(2)$ gives a gauge field. Again we can read of the transformation law for the gauge fields from the transformation law of the comparator. Thus, we find

$$A^k_\mu \frac{\sigma^k}{2} \to A^k_\mu \frac{\sigma^k}{2} + \frac{1}{g}(\partial_\mu \theta^k)\frac{\sigma^k}{2} + i[\theta^k \frac{\sigma^k}{2}, A^l_\mu \frac{\sigma^l}{2}] + \ldots.$$

This is seen to correspond to the fact that

$$A_\mu \to A_\mu + D_\mu \theta.$$

The field strength tensor $F^k_{\mu\nu}$ also has to be modified in order for gauge invariance to hold. One finds that

$$F^k_{\mu\nu}\frac{\sigma^k}{2} = \partial_\mu A^k_\nu \frac{\sigma^k}{2} - \partial_\nu A^k_\mu \frac{\sigma^k}{2} - ig[A^k_\mu \frac{\sigma^k}{2}, A^l_\nu \frac{\sigma^l}{2}].$$

This can be written as

$$F^k_{\mu\nu} = \partial_\mu A^k_\nu - \partial_\nu A^l_\mu + g\epsilon^{klm} A^l_\mu A^m_\nu$$

when we use the commutation relations $[\sigma^k,\sigma^l]=2i\epsilon^{klm}\frac{\sigma^m}{2}$. Note that $F^k_{\mu\nu}$ is gauge covariant and not gauge invariant. But since

$$\mathrm{tr}[(F^k_{\mu\nu}\frac{\sigma^k}{2})^2]=\frac{1}{4}(F^k_{\mu\nu})^2$$

is gauge invariant we conclude that the most general renormalizable Lagrangian invariant under the $SU(2)$ symmetry is

$$\mathcal{L}=\overline{\psi}i\gamma^\mu D_\mu\psi-\frac{1}{4}(F^k_{\mu\nu})^2-m\overline{\psi}\psi.$$

Notice the similar "form" of this Lagrangian and the QED Lagrangian. Before exploring general Yang-Mills theories we comment on our Lagrangian.

Note that we have a Lagrangian $\mathcal{L}_{\text{"}Maxwell\text{"}}=\frac{1}{4}(F^k_{\mu\nu})^2$ which resembles the Maxwell Lagrangian and when we "add" fermions (in our case a single doublet of fermions) to our theory we should add the interaction term and work with the Lagrangian given by

$$\mathcal{L}=\overline{\psi}i\gamma^\mu D_\mu\psi-\frac{1}{4}(F^k_{\mu\nu})^2-m\overline{\psi}\psi=\overline{\psi}i\not{D}\psi-\frac{1}{4}(F^k_{\mu\nu})^2-m\overline{\psi}\psi.$$

Note that we introduced the Feynman slash notation, i.e. $\gamma^\mu D_\mu=\not{D}$. We have found a generalization of electrodynamics as a $U(1)$ Yang-Mills theory. The "generalized" Maxwell equations now involve **three** charge densities, three current densities, three scalar potentials, three vector potentials, three "electric" fields and three "magnetic" fields in this theory.

Summarizing $SU(2)$ Yang-Mills theory (also called $SU(2)$ *gauge theory*) we have the following. We have a doublet of Dirac fermions ψ, a covariant derivative $D_\mu=\partial_\mu-igA^k_\mu\frac{\sigma^k}{2}$, three gauge fields A^k_μ, three field strength tensors $F^k_{\mu\nu}=\partial_\mu A^k_\nu-\partial_\nu A^l_\mu+g\epsilon^{klm}A^l_\mu A^m_\nu$ and a Lagrangian consisting of a generalized Maxwell Lagrangian, a generalized Dirac Lagrangian and a generalized interaction Lagrangian, given by

$$\mathcal{L}=\underbrace{\overline{\psi}i\gamma^\mu\partial_\mu\psi-m\overline{\psi}\psi}_{\text{"}Dirac\text{"}}+\underbrace{\frac{-1}{4}(F^k_{\mu\nu})^2}_{\text{"}Maxwell\text{"}}+\underbrace{(g\overline{\psi}\gamma^\mu\psi)A_\mu}_{\text{"}Interaction\text{"}}.$$

Therefore $SU(2)$ gauge theory describes two equal-mass Dirac fields in interaction with three massless vector gauge fields.

The physics in this theory can be seen as follows. Just as was the case with $U(1)$ Yang-Mills theory we can identify the terms in our Lagrangian by fields and forces. In this case the group $SU(2)$ can be identified with the *weak field* and the gauge fields are the well known W^+, W^- and Z bosons. They are the force carriers. Again we make the identification where $SU(2)$ can be identified with the field(s) and gauge invariance gives the forces.

We now give the general idea of a Yang-Mills theory.

Consider a compact Lie group G, which we call the *gauge* group, with generators T^a and structure constants f^{abc} given by

$$[T^a,T^b]=if^{abc}T^c.$$

Suppose we have fields transforming according to an irreducible representation ρ of G. The first step is to replace the "derivative" in our Lagrangian with a "covariant derivative" which one finds to have the general form (summation over a implied)

$$D_\mu=\partial_\mu-igA^a_\mu T^a.$$

Then we define a field strength tensor

$$F^a_{\mu\nu}=\partial_\mu A^a_\nu-\partial_\nu A^a_\mu+gf^{abc}A^b_\mu A^c_\nu$$

and give the most general renormalizable Lagrangian invariant under parity and gauge transformations

$$\mathcal{L}=\overline{\psi}i\gamma^\mu D_\mu\psi-m\overline{\psi}\psi-\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}.$$

We speak of G gauge theory. As we have already shown taking $G = U(1)$ reproduces quantum electrodynamics. This is easily seen since we have only one generator. Therefore we have only one gauge field and all structure constants are zero. From a physical point of view, this can be seen as the fact that we have only one force carrier: the photon. In case $G = SU(2)$ we have three generators, three gauge fields and structure constants given by $f^{abc} = \epsilon^{abc}$. We already noted that the theory describes the weak force in this case.

Note that until now we are using the fundamental (spinor) representation of the Lie group to define the action on the fields. Later we shall consider $SU(2)$ gauge theories with higher $SU(2)$ representations.

Yang and Mills started from the assumption that there exist two elementary spin-$\frac{1}{2}$ particles of equal mass and had a very nice application in mind. It was the nucleon (proton and neutron) system which they were trying to understand. The small difference in the mass of the proton and the mass of the neutron would be attributed to electromagnetic symmetry-breaking. Unfortunately, their model for the nucleon system did not hold. So in its original form Yang-Mills theory turned out to be of little use. So the identifications we made here weren't done by Yang and Mills themselves. But is was definitely their idea of "gauging" global symmetries which was late used in the context of color $SU(3)$ symmetry and *weak* isospin-hypercharge $(SU(2) \otimes U(1))$ symmetry in the weak interactions.

To conclude this section we note on the color $SU(3)$ symmetry mentioned above. Differently put we will discuss quantum chromodynamics, i.e. $SU(3)$ gauge theory. The fields are the color triplets *blue, green, red.* Since $SU(3)$ is an eight-dimensional Lie group we have eight gauge fields which are just the gluons. The gluons are thus the force carriers. For the last time we note that we see that we can make the following identification: $SU(3)$ represents the colors (fields) and the gauge invariance represents the forces.

7. Helicity and chirality

We give a short introduction to Weyl and Dirac fermions. We also discuss chirality.

Let us go back to the complex representation theory of $SU(2)$ studied in Section 4 of Chapter 3. As we know all representations of $SU(2)$ are characterized by their highest weight. The highest weight d of a representation for $SU(2)$ is either an integer or half-integer, i.e. $d = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$. The highest weight for the fundamental/spinor representation equals $\frac{1}{2}$. The dimension of a representation of highest weight d for $SU(2)$ equals $2d + 1$.

We can label representations of the well known Lorentz group as (j, j') since the Lie algebra of the Lorentz group is isomorphic to the Lie algebra of $SU(2) \times SU(2)$. Here j and j' are elements of $\{0, \frac{1}{2}, 1, \frac{3}{2}, \ldots\}$.

We already noted in Section 3 that Dirac fermions transform as $(\frac{1}{2}, \frac{1}{2})$ under a Lorentz transformation. Now, we define a left handed fermion to be a solution of the Dirac equation that transforms as $(\frac{1}{2}, 0)$ and a right handed fermion a solution that transforms as $(0, \frac{1}{2})$ under a Lorentz transformation. Note that this is just a label, and not real handedness. Note that it is clear that left handed fermions transform into right handed fermions under a parity transform and vice versa. Equivalently, we can use the γ^5 matrix to define the handedness. The γ^5 matrix has eigenvalues ± 1. Therefore, when we apply the projection operator $\frac{1}{2}(1 - \gamma^5)$ (respectively $\frac{1}{2}(1 + \gamma^5)$) to our Dirac fermion we project out the left handed (respectively right handed) component. The handedness of a Dirac fermion is called chirality in this case. Therefore left handed (and right handed) fermions are called chiral or Weyl fermions.

We end this short section with some final remarks. First of all, for some unknown reason, it turns out that only Weyl fermions are coupled to the weak force. Also note that Weyl fermions are given by irreducible representations whereas Dirac fermions are given by reducible representations. An important fact that we shall use is that a doublet of left handed Weyl fermions is the same as a

Dirac fermion in an $SU(2)$ gauge theory. Finally, for any general field ψ we shall use the notation

$$\psi_L = \frac{1}{2}(1-\gamma^5)\psi$$

and

$$\psi_R = \frac{1}{2}(1+\gamma^5)\psi.$$

8. Berezin integration

We generalize the notion of a Feynman path integral to field theory for fermions. To this extent we shall define the Grasmman algebra and an operation on it called Berezin integration.

It is important to note that all spinor fields take their values in the so called Grassmann algebra. The Grassmann algebra consists of so called Grassmann numbers which are to be treated as regular (real or complex) numbers with the difference being that all Grassmann numbers anti-commute. This fact is exactly what characterizes the difference between path integrals for bosons and fermions. In the literature Grassmann numbers are also called anti-commuting variables or g-numbers.

The Grassmann algebra consists of Grassmann numbers θ that mutually anticommute. Thus, for any Grassmann number θ it holds that $\theta^2 = 0$. When our field is not of characteristic 2 it holds that $\theta^2 = 0$ for all Grassmann numbers θ if and only if $\theta\theta' = -\theta'\theta$ for all Grassmann numbers θ and θ'. In particular, over $\mathbf{R}$ and $\mathbf{C}$ the two conditions are equivalent. We shall consider the Grassmann algebra as an (infinite dimensional) algebra over $\mathbf{C}$. Therefore Grassmann numbers commute with complex numbers. We wish to define integration over Grassmann numbers, such that the normalization of the wave function ψ is as usual

$$\langle\psi \mid \psi\rangle = \mid a_0^2 \mid + \mid a_1^2 \mid = \int d\theta\psi^*(\theta)\psi(\theta).$$

Since we wish to study spinor fields we consider (smooth) functions of a single Grassmann number. Let $f(\theta)$ be a function of a Grassmann variable. Since $\theta^2 = 0$, this function has a finite Taylor series given by $f(\theta) = a_0 + a_1\theta$. Since a_0 and a_1 are complex numbers this shows that $f(\theta)$ actually spans a two dimensional complex vector space.

Now, we will define a certain operation on this space which we shall call Berezin integration. The name indicates that this is like usual Lesbesque integration on a space, but this is not the case. But as the reader will see, Berezin integration does resemble usual integration in some ways.

The first thing we really want is that when we integrate a function $f(\theta) = a_0 + a_1\theta$ with respect to the Grassmann variable θ we want to get a_1. Therefore it seems plausible to define the integration on Grassmann numbers by

$$\int d\theta\theta = 1, \quad \int d\theta 1 = 0.$$

Demanding linearity of Berezin integration we see that the function $f(\theta)$ does integrate to a_1. We note that $d\theta$ is considered as an independent Grassmann number. This implies that when one considers multiple Grassmann integrations (as we shall do in a moment) one may must strictly take in account the order of integration. We work out some examples for the readers convenience.

Example 5.3. It is clear from the definition that for a a regular number and b a Grassmann number it holds that

$$\int d\theta(a+\theta b) = \int d\theta a + \int d\theta(\theta b) = 0 + (\int d\theta\theta)b = 1\cdot b = b.$$

Note that the ordering of b with respect to θ is important. With the opposite ordering one would obtain

$$\int d\theta(b\theta) = \int d\theta(-\theta b) = -b.$$

EXAMPLE 5.4. Since $\theta^2 = 0$ for any Grassmann number, for any Grassmann number x, it holds that

$$\int d\theta \exp(\theta x) = \int d\theta(1 + \theta x + \frac{\theta x \theta x}{2!} + \ldots) = \int d\theta(1 + \theta x - \frac{\theta^2 x^2}{2!} + \ldots) = \int d\theta(1 + \theta x) = x.$$

Note that the ordering of x with respect to θ is important. With the opposite ordering one would obtain

$$\int d\theta \exp(x\theta) = \int d\theta(1 + x\theta) = \int d\theta(1 - \theta x) = -x.$$

EXAMPLE 5.5. Let y be a regular number and x be a Grassmann number. It holds that

$$\int d\theta \exp(\theta x + y) = x \exp(y).$$

The reader can easily verify this by using the identity

$$\exp(\theta x + y) = \sum_{n=0}^{\infty} \frac{1}{n!}(\theta x + y)^n = \sum_{n=0}^{\infty} \frac{1}{n!} y^n + \theta x \sum_{n=1}^{\infty} \frac{1}{(n-1)!} y^{n-1} = \exp(y) + \theta x \exp(y).$$

Let us consider multiple Berezin integration, i.e. take $\theta_1, \theta_2, \ldots, \theta_n$ to be Grassmann numbers for which

$$\int d\theta_i 1 = 0, \qquad \int d\theta_i \theta_i = 1 \ (i \text{ not summed}), \quad i = 1, 2, \ldots, n.$$

When the measure of integration and integrand involve more than one variable, as is the case here, one performs the integration according to a nested procedure. Thus, for instance

$$\begin{aligned}\int d\theta_1 \int d\theta_2 \theta_1 \theta_2 &= \int d\theta_1 \left(\int d\theta_2 \theta_1 \theta_2 \right) = \int d\theta_1 \left(\int d\theta_2 - \theta_2 \theta_1 \right) \\ &= \int d\theta_1 - \theta_1 = -1.\end{aligned}$$

In general we shall avoid writing multiple integral signs. Hence the above integral will be denoted by

$$\int d\theta_1 \int d\theta_2 \theta_1 \theta_2 = \int d\theta_1 d\theta_2 \theta_1 \theta_2.$$

The following examples will prove to be the most important examples of Berezin integration.

EXAMPLE 5.6. We recall that we use the Einstein summation convention. Consider $2n$ independent Grassmann numbers $\eta_1, \overline{\eta}_1, \ldots, \eta_n, \overline{\eta}_n$. For any $(n \times n)$-matrix $A = (A_{ij})_{i,j=1}^n$ with complex entries it holds that

$$\int d\eta_1 d\overline{\eta_1} \ldots d\eta_n d\overline{\eta_n} \exp(\overline{\eta_i} A_{ij} \eta_j) = \det(A).$$

To prove this identity we consider the Taylor expansion

$$\exp(\overline{\eta_i} A_{ij} \eta_j) = 1 + \overline{\eta_i} A_{ij} \eta_j + \frac{(\overline{\eta_i} A_{ij} \eta_j)^2}{2} + \ldots + \frac{(\overline{\eta_i} A_{ij} \eta_j)^n}{n!} + \ldots.$$

Note that terms of order higher than n are zero. They all contain more than n Grassmann numbers implying that they contain at least one square of a Grassmann number. Furthermore, all terms of order smaller than n integrate to zero since the integrand in that case will be independent of at least one Grassmann variable. Therefore, the only terms that contribute to the integral are the terms of order n ,i.e.

$$\begin{aligned}\int d\eta_1 d\overline{\eta_1} \ldots d\eta_n d\overline{\eta_n} \exp(\overline{\eta_i} A_{ij} \eta_j) &= \int d\eta_1 d\overline{\eta_1} \ldots d\eta_n d\overline{\eta_n} \frac{(\overline{\eta_i} A_{ij} \eta_j)^n}{n!} \\ &= \frac{1}{n!} \int d\eta_1 d\overline{\eta_1} \ldots d\eta_n d\overline{\eta_n} (\overline{\eta_i} A_{ij} \eta_j)^n.\end{aligned}$$

Since the square of a Grassmann number equals zero the only terms that do not integrate to zero are of the form

$$\int d\eta_1 d\overline{\eta_{\sigma(1)}} \ldots d\eta_n d\overline{\eta_{\sigma(n)}} A_{1\sigma(1)} \ldots A_{n\sigma(n)} \eta_1 \overline{\eta_{\sigma(1)}} \ldots \eta_n \overline{\eta_{\sigma(n)}}$$

where $\sigma \in S_n$ is a permutation of the set $\{1, 2, \ldots, n\}$. Since there are $n!$ ways of permuting the set $\{1, 2, \ldots\}$ there are exactly $n!$ terms of this form. Noting that

$$\int d\eta_1 d\overline{\eta_{\sigma(1)}} \ldots \eta_n d\overline{\eta_{\sigma(n)}} \eta_1 \overline{\eta_{\sigma(1)}} \ldots \eta_n \overline{\eta_{\sigma(n)}} = \text{sign}(\sigma)$$

we have that

$$\begin{aligned}
\int d\eta_1 d\overline{\eta_1} \ldots d\eta_n d\overline{\eta_n} \exp(\overline{\eta_i} A_{ij} \eta_j) &= \frac{1}{n!} \int d\eta_1 d\overline{\eta_{\sigma(1)}} \ldots d\eta_n d\overline{\eta_{\sigma(n)}} \sum_{\sigma \in S_n} n! A_{1\sigma(1)} \ldots A_{n\sigma(n)} \eta_1 \overline{\eta_{\sigma(1)}} \ldots \overline{\eta_n \eta_{\sigma(n)}} \\
&= \sum_{\sigma \in S_n} A_{1\sigma(1)} \ldots A_{n\sigma(n)} \int d\eta_1 d\overline{\eta_{\sigma(1)}} \ldots \eta_n d\overline{\eta_{\sigma(n)}} \eta_1 \overline{\eta_{\sigma(1)}} \ldots \eta_n \overline{\eta_{\sigma(n)}} \\
&= \sum_{\sigma \in S_n} A_{1\sigma(1)} \ldots A_{n\sigma(n)} \text{sign}(\sigma) = \det(A).
\end{aligned}$$

EXAMPLE 5.7. Given any n Grassmann numbers $\theta_1, \ldots, \theta_n$ and antisymmetric $(n \times n)$-matrix A with complex coordinates it holds that

$$\det(A) = \left(\int d\theta_1 \ldots d\theta_n \exp(\frac{1}{2} \theta_i A_{ij} \theta_j) \right)^2,$$

implying that

$$\pm\sqrt{\det(A)} = \int d\theta_1 \ldots d\theta_n \exp(\frac{1}{2} \theta_i A_{ij} \theta_j).$$

To this extent we substitute $\eta_i = \frac{1}{\sqrt{2}}(\theta_i + i\overline{\theta_i})$ and $\overline{\eta_i} = \frac{1}{\sqrt{2}}(\theta_i - i\overline{\theta_i})$ for all $i = 1, \ldots, n$ in the identity of Example 5.6. Note that i denotes the imaginary number when it is not used as an index. The reader may verify that

$$d\eta_1 d\overline{\eta_1} \ldots d\eta_n d\overline{\eta_n} = \frac{1}{\det(B_n)} d\theta_1 d\overline{\theta_1} \ldots d\theta_n d\overline{\theta_n} = \frac{1}{(-i)^n} d\theta_1 d\overline{\theta_1} \ldots d\theta_n d\overline{\theta_n}$$

with B_n the $(2n) \times (2n)$ transformation matrix given by

$$B_n = \frac{1}{\sqrt{(2)}} \begin{pmatrix} 1 & i & 0 & 0 & \ldots & \ldots \\ 1 & -i & 0 & 0 & \ldots & \ldots \\ 0 & 0 & 1 & i & \ldots & \ldots \\ 0 & 0 & 1 & -i & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \ldots & \ldots & \ldots & \ldots & 1 & i \\ \ldots & \ldots & \ldots & \ldots & 1 & -i \end{pmatrix}$$

It holds that

$$\begin{aligned}
\overline{\eta_i} A_{ij} \eta_j &= \frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j} - i \frac{1}{2} \theta_i A_{ij} \overline{\theta_j} + i \frac{1}{2} \overline{\theta_i} A_{ij} \theta_j \\
&= \frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j} + i \frac{1}{2} \overline{\theta_j} A_{ij} \theta_i + i \frac{1}{2} \overline{\theta_i} A_{ij} \theta_j \\
&= \frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j} - i \frac{1}{2} \overline{\theta_i} A_{ij} \theta_j + i \frac{1}{2} \overline{\theta_i} A_{ij} \theta_j \\
&= \frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j}
\end{aligned}$$

since A is antisymmetric and Grassmann numbers anticommute. This shows that

$$\det(A) = \int d\eta_1 d\overline{\eta_1} \ldots d\eta_n \overline{d\eta_n} \exp(\overline{\eta_i} A_{ij} \eta_j) = \frac{1}{(-i)^n} \int d\theta_1 d\overline{\theta_1} \ldots d\theta_n d\overline{\theta_n} \exp(\frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j}).$$

Now we invoke the Campbell-Hausdorff formula to see that

$$\begin{aligned}
\exp(\frac{1}{2} \theta_i A_{ij} \theta_j) \exp(\frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j}) &= \exp(\frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j} + \frac{1}{2} [\frac{1}{2} \theta_i A_{ij} \theta_j, \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j}]) \\
&= \exp(\frac{1}{2} \theta_i A_{ij} \theta_j + \frac{1}{2} \overline{\theta_i} A_{ij} \overline{\theta_j}).
\end{aligned}$$

To justify the last step it suffices to note that for any four Grassmann numbers g_1, g_2, g_3 it holds that

$$(g_1 g_2)(g_3) = (g_3)(g_1 g_2).$$

That is, the product of two Grassmann numbers is a regular number. Since

$$d\theta_1 d\overline{\theta_1} \dots d\theta_n d\overline{\theta_n} = (-1)^{(n-1)+\dots+1} d\theta_1 \dots d\theta_n d\overline{\theta_1} \dots d\overline{\theta_n} = (-1)^{\frac{n(n-1)}{2}} d\theta_1 \dots d\theta_n d\overline{\theta_1} \dots d\overline{\theta_n},$$

we may conclude that

$$\begin{aligned}
\det(A) &= \frac{1}{(-i)^n} \int d\theta_1 d\overline{\theta_1} \dots d\theta_n d\overline{\theta_n} \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \exp(\frac{1}{2}\overline{\theta_i} A_{ij}\overline{\theta_j}) \\
&= \frac{1}{(-i)^n}(-1)^{\frac{n(n-1)}{2}} \int d\theta_1 \dots d\theta_n d\overline{\theta_1} \dots d\overline{\theta_n} \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \exp(\frac{1}{2}\overline{\theta_i} A_{ij}\overline{\theta_j}) \\
&= \frac{1}{(-i)^n}(-1)^{\frac{n(n-1)}{2}} \int d\theta_1 \dots d\theta_n \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \int d\overline{\theta_1} \dots d\overline{\theta_n} \exp(\frac{1}{2}\overline{\theta_i} A_{ij}\overline{\theta_j}) \\
&= \frac{1}{(-i)^n}(-1)^{\frac{n(n-1)}{2}} \left(\int d\theta_1 \dots d\theta_n \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \right)^2 .
\end{aligned}$$

When n is odd it holds that $\det(A) = 0$, since $\det(A) = \det(A^t) = \det(-A) = (-1)^n \det(A) = -\det(A)$. Therefore the identity

$$\det(A) = \left(\int d\theta_1 \dots d\theta_n \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \right)^2$$

holds. When n is even we may conclude that

$$(-i)^n (-1)^{-\frac{n(n-1)}{2}} \det(A) = (-1)^{n-\frac{n^2}{2}} \det(A) = \det(A) = \left(\int d\theta_1 \dots d\theta_n \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \right)^2 .$$

Examples 5.6 and 5.7 can be generalized to operators. For all operators $\hat{A} = A$ that will be of interest to us we have the (formal) equalities

$$\det(A) = \int d\eta_1 d\overline{\eta_1} \dots d\eta_n d\overline{\eta_n} \exp(\overline{\eta_i} A_{ij}\eta_j) = \left(\int d\theta_1 \dots d\theta_n \exp(\frac{1}{2}\theta_i A_{ij}\theta_j) \right)^2 .$$

The determinant is taken to be the formal product of all the eigenvalues of $\hat{A}$. In most cases this product shall have to be regularized in order for the integral to be able to be defined properly.

The next chapter builds on the fact that formally one has

$$\pm\sqrt{\det(A)} = \int d\theta_1 \dots d\theta_n \exp(\frac{1}{2}\theta_i A_{ij}\theta_j).$$

We will actually see that the square root of the determinant can not be defined satisfactorily in some cases. This is what leads to *Witten's anomaly*.

CHAPTER 6

Witten's global $SU(2)$ anomaly

1. Towards Witten's anomaly

In this chapter we consider the Yang-Mills theory with gauge group $SU(2)$ in Minkowski space. In chapter 4 we showed that $\pi_4(SU(2))$ is nontrivial and equals $\mathbf{Z}/2\mathbf{Z}$. The nontrivial element of $\pi_4(SU(2))$ gives us a topologically nontrivial gauge transformation. We will show that due to the existence of this gauge transformation we cannot have an odd number of doublets (in the fundamental representation) of Weyl fermions in our theory. If we do have an odd number of Weyl doublets the theory will be mathematically inconsistent. We shall use a result on Berezin integration, the Schwinger-Dyson equations and the Atiyah-Singer index theorem. We shall also discuss generalizations of this so called anomaly to other Yang-Mills theories and study $SU(2)$ gauge theories with higher $SU(2)$ representations.

Spontaneous chiral symmetry breaking in an $SU(2)$ gauge theory permits massless Weyl fermions to pair up and become massive Dirac fermions. This fact suggests there might be something strange to an $SU(2)$ gauge theory with just one doublet of Weyl fermions. As we shall see this strange fact shall translate itself into mathematical inconsistency of the theorem. We say that the theory contains an anomaly, i.e. a breakdown of a (classical) symmetry. This anomaly is called Witten's anomaly. See [**Witt**].

The reader shall see that Witten's anomaly arises from a problem somewhat analogous to the Adler-Bell-Jackiw anomaly. The Adler-Bell-Jackiw causes the fact that the Weyl Lagrangian

$$\mathcal{L}_{Weyl} = -\frac{1}{2g^2}\mathrm{tr}\ F_{\mu\nu}^2 + \psi_L^\dagger \sigma^\mu(\partial_\mu + iA_\mu(x))\psi_L$$

is not renormalizable. In other words, coupling the gauge field A_μ gauge invariantly to a lefthanded Weyl fermion doublet leads to a non renormalizable theory. As a consequence the theory is inconsistent. But as we have seen the Dirac Lagrangian

$$\mathcal{L}_{Dirac} = -\frac{1}{2g^2}\mathrm{tr}\ F_{\mu\nu}^2 + \psi^\dagger \sigma^\mu(\partial_\mu + iA_\mu(x))\psi$$

on the contrary is renormalizable and gives us quantum electrodynamics. In other words, coupling A_μ gauge invariantly to a Dirac four component field leads to QED when A_μ is identified with the photon, ψ with the electron and g with the electric charge.

Witten's anomaly causes the fact that an $SU(2)$ theory with a single doublet of Weyl fermions is ill-defined. Just as is the case with the Adler-Bell-Jackiw anomaly, the theory is well-defined when the gauge fields are coupled to Dirac fermions.

Although the Adler-Bell-Jackiw anomaly has been used to make an analogy with Witten's anomaly we must state that the latter is of a different nature. It is caused by the existence of a topologically non-trivial gauge transformation, i.e. it is topological of nature. As will be clarified later on, we can categorize gauge transformations in an $SU(2)$ gauge theory by being small or large. This is reflected in the fact that $\pi_4(SU(2)) = \mathbf{Z}/2\mathbf{Z}$. Since all higher homotopy groups of $U(1)$ are trivial it is easy to see that the Adler-Bell-Jackiw anomaly, which is seen to arise in $U(1)$ gauge theories, is not topological (or homotopic) of nature.

2. A topologically non-trivial gauge transformation

We may characterize all gauge transformations in $SU(2)$ gauge theory by the fourth homotopy group of $SU(2)$. Since the fourth homotopy group of $SU(2)$ was seen to be $\mathbf{Z}/2\mathbf{Z}$ we may characterize gauge transformations topologically by being "small" or "large".

It can be shown that all gauge transformations in an $SU(2)$ gauge theory can be classified by the fourth homotopy group of $SU(2) = \mathbf{S}^3$. The reader may look at [**Bar**]. Since $\pi_4(SU(2)) = \mathbf{Z}/2\mathbf{Z}$ a gauge transformation either belongs to the class $\overline{0}$ or $\overline{1}$. A gauge transformation is said to be *small* when it can be connected to the identity, i.e. when it belongs to the class $\overline{0}$. A gauge transformation is said to be *large* when it can not be connected to the identity, i.e. when it belongs to the class $\overline{1}$. That is, in addition to null-homotopic gauge transformations we have gauge transformations that wrap around $SU(2)$ in such a way that when we wrap them around $SU(2)$ twice they are null-homotopic. Note that we let gauge transformations take their values in the gauge group instead of the Lie algebra of the gauge group. This goes through the adjoint representation of course. Large gauge transformations are also said to be topologically nontrivial.

Let A_μ be a gauge field. Let $U(x)$ be a topologically nontrivial element of $\pi_4(SU(2))$. A gauge transformation looks like $A_\mu \to U^{-1}A_\mu U - iU^{-1}\partial_\mu U$. The latter is called the conjugate gauge field (of A_μ) and is denoted by $A_\mu^U = U^{-1}A_\mu U - iU^{-1}\partial_\mu U$. Note that this notation hides the fact that the action of U is the adjoint action of the Lie group on its Lie algebra. For example, $U(x)A_\mu$ should be seen as the adjoint action of $U(x) \in SU(2)$ on A_μ. We already noted A_μ takes its values in the Lie algebra of $SU(2)$.

3. Intuitive approach

We elucidate the properties of an $SU(2)$ gauge theory with a single doublet of Weyl fermions in the fundamental representation. We shall see that this theory is mathematically inconsistent. The mathematical details are discussed in the next section.

We start with the partition function for an $SU(2)$ gauge theory with a single doublet of Weyl fermions

$$Z = \int d\psi d\overline{\psi} \int dA_\mu \times \exp\left(- \int d^4x[\frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}^2 + \overline{\psi}i\not{D}\psi]\right).$$

Let us recall that $\not{D} = \gamma^\mu D_\mu$ is the Dirac operator and that the action integral is carried out over Euclidean four-space. We would like to integrate out the (Weyl) fermions and discuss the effective theory with the fermions eliminated. The basic integral, for a theory with a doublet of Weyl fermions, is

$$I_{Weyl} = \int (d\psi d\overline{\psi})_{Weyl}\ \exp(\overline{\psi}i\not{D}\psi).$$

Since the gauge group is $SU(2)$ a single doublet of Dirac fermions is the same as two Weyl fermions. Therefore we expect the basic integral, for a theory with a doublet of Dirac fermions, to be

$$I_{Dirac} = \int (d\psi d\overline{\psi})_{Dirac}\ \exp(\overline{\psi}i\not{D}\psi) = \left(\int (d\psi d\overline{\psi})_{Weyl}\ \exp(\overline{\psi}i\not{D}\psi)\right)^2 = I_{Weyl}^2.$$

It is well known that for a theory with a doublet of Dirac fermions the basic integral equals[1]

$$I_{Dirac} = \det(i\not{D}).$$

Here the determinant is the formal product of all of its eigenvalues, i.e.

$$\det(i\not{D}) = \prod_{\lambda \text{ eigenvalue of } i\not{D}} \lambda.$$

Since a doublet of Dirac fermions could have a gauge invariant bare mass we may regularize the above expression with the aid of Pauli-Villars regularization. This process is performed by adding

[1]See the section on Berezin integration, i.e. Section 8 of Chapter 5. In that section we proved this equality for matrices. This equality is a generalization to operators.

auxiliary massive particles to the Lagrangian. This is possible due to spontaneous chiral symmetry breaking, i.e. since a doublet of Dirac fermions could have a gauge invariant bare mass. Having defined the determinant in this way it becomes completely gauge invariant, i.e. invariant under small and large gauge transformations. We conclude that an $SU(2)$ theory with a single doublet of Dirac fermions can be defined satisfactorily.

Going back to $SU(2)$ theories with a single doublet of Weyl fermions, the fermion integration with a single doublet of Weyl fermions would give

$$\int (d\psi d\overline{\psi})_{Weyl}\ \exp(\overline{\psi} i \not{D} \psi) = \pm\sqrt{I_{Dirac}} = \pm \det(i\not{D})^{\frac{1}{2}}.$$

Note the ambiguity in the sign of the square root. Given a gauge field A_μ we are free to define the sign of the square root in an arbitrary way for A_μ. The Schwinger-Dyson equations imply that

$$\begin{aligned} 0 &= \int dA_\mu \frac{\partial}{\partial A_\mu}\left(\det(i\not{D}(A_\mu))^{\frac{1}{2}} e^{i\overline{\psi} i\not{D}(A_\mu)\psi}\right) \\ &= \int dA_\mu \left(\frac{\partial}{\partial A_\mu}\det(i\not{D}(A_\mu))^{\frac{1}{2}}\right) e^{i\overline{\psi} i\not{D}(A_\mu)\psi} + i\int dA_\mu \det(i\not{D}(A_\mu))^{\frac{1}{2}} e^{i\overline{\psi} i\not{D}(A_\mu)\psi}\frac{\partial}{\partial A_\mu}(\overline{\psi} i\not{D}(A_\mu)\psi) \\ &= \left\langle \frac{\partial}{\partial A_\mu}\det(i\not{D}(A_\mu))^{\frac{1}{2}}\right\rangle + i\left\langle \det(i\not{D}(A_\mu))^{\frac{1}{2}}\frac{\partial}{\partial A_\mu}(\overline{\psi} i\not{D}(A_\mu)\psi)\right\rangle. \end{aligned}$$

When we demand smooth gauge field dependence, by the Schwinger-Dyson equations, which can be seen as a sort of Gauss's Law for functional integrals, there is no further freedom after we choose the sign. We must define the fermion integral $\det(i\not{D})^{\frac{1}{2}}$ to vary smoothly as A_μ is varied. More precisely, suppose we have fixed the sign. When we perform a gauge transformation the second term

$$\left\langle \det(i\not{D}(A_\mu))^{\frac{1}{2}}\frac{\partial}{\partial A_\mu}(\overline{\psi} i\not{D}(A_\mu)\psi)\right\rangle$$

either changes sign or does not change at all. As a consequence, the first term in the above equation must change accordingly. Since that term is nothing but the derivative of our determinant this implies that as our determinant changes its derivative changes accordingly.

Let us discuss the effect of small and large gauge transformations on the fermion integral. First of all, it is easy to see that the fermion integral is invariant under small gauge transformations. Under a small gauge transformation

$$A_\mu \to A'_\mu = A_\mu - D_\mu\theta(x),$$

the covariant derivative transforms as

$$D_\mu\psi \to \exp\left(i\theta(x)\right) D_\mu\psi.$$

This implies that $\det(i\not{D})^{\frac{1}{2}}$ is certainly gauge invariant under small gauge transformations - since the sign does not change abruptly. On the other hand, as will be explained the effect of large gauge transformations leads to a change of sign in the fermion integral.

THEOREM 6.1. Let U be a nontrivial element of $\pi_4(SU(2))$. For any gauge field A_μ, it holds that

$$[\det i\not{D}(A_\mu)]^{\frac{1}{2}} = -[\det i\not{D}(A_\mu^U)]^{\frac{1}{2}}.$$

The proof of Theorem 6.2 is postponed until the next section. For now, we discuss its implications.

The partition function for $SU(2)$ gauge theory with a single doublet of Weyl fermions is

$$\begin{aligned} Z &= \int d\psi d\overline{\psi}\int dA_\mu \times \exp\left(-\int d^4x[\frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}^2 + \overline{\psi} i\not{D}\psi]\right) \\ &= \int dA_\mu (\det\ i\not{D}(A_\mu))^{\frac{1}{2}}\exp\left(-\int d^4x[\frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}^2 + \overline{\psi} i\not{D}\psi]\right) = 0 \end{aligned}$$

since the contribution of any gauge field A_μ is exactly cancelled by the equal and opposite contribution of A_μ^U. Also, given a gauge invariant operator X the path integral

$$Z_X = \int dA_\mu (\det\ i\not{D})^{\frac{1}{2}} X \exp\left(-\int d^4x [\frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}^2 + \overline{\psi} i\not{D}\psi]\right)$$

identically equals zero. On these grounds Witten concluded that expectation values $\langle X\rangle = \frac{Z_X}{Z} = \frac{0}{0}$ are indeterminate and that the theory is mathematically inconsistent or ill-defined.

What is the remedy for this anomaly? First of all, since the problem lies within the the sign of $(\det\ i\not{D})^{\frac{1}{2}}$ we could, in order to avoid the anomaly, define the fermion integral as the absolute value of $(\det\ i\not{D})^{\frac{1}{2}}$. Unfortunately the resulting theory would not obey the Schwinger-Dyson equations because there will be a Delta delta function in the equation. So this will not not work. Second of all, we could try to avoid integrating over that part of field space which contains A_μ^U. But since A_μ and A_μ^U are continuously connected, i.e. A_μ and A_μ^U lie in the same sector of field space, this is not possible. Finally, we already noted that the theory with a single doublet of Dirac fermions can be defined satisfactorily with the aid of Pauli-Villars regularization. Since two doublets of Weyl fermions are the same as a single doublet of Dirac fermions one can expect that our theory should contain an even number of doublets of Weyl fermions. We will comment on this more detailed in Section 5.

In the next section we provide the proof for Theorem 6.2. This is the only gap left in establishing the global $SU(2)$ anomaly in an $SU(2)$ gauge theory with a single doublet of Weyl fermions. After that we will consider generalizations to other Yang-Mills theories and discuss theories with other $SU(2)$ representations.

4. Mathematical details

We provide a mathematical proof for Theorem 6.2. In the course of doing this we shall use a certain form of the Atiyah-Singer index theorem for a certain five-dimensional Dirac operator.

For the readers convenience we repeat Theorem 6.2.

THEOREM 6.2. Let U be a nontrivial element of $\pi_4(SU(2))$. For any gauge field A_μ, it holds that

$$[\det i\not{D}(A_\mu)]^{\frac{1}{2}} = -[\det i\not{D}(A_\mu^U)]^{\frac{1}{2}}.$$

To prove this theorem we shall study the complete eigenvalue spectrum of the (4-dimensional) Dirac operator $i\not{D}(A_\mu) = i\not{D}$ in a given gauge field A_μ. Let us recall that we are studying $SU(2)$ gauge theory in Minkowski space. Taking space-time to be a large enough sphere, with a coupled point at infinity, will force the eigenvalue spectrum of $i\not{D}$ to be discrete. Since $i\not{D}$ is nothing but the Hamiltonian operator one can compare this reasoning to the case of a particle in a box. In the case of a particle in a box of length L the momentum operator $p = \frac{h}{\lambda} = \frac{nh}{L}$ is seen to have a discrete eigenvalue spectrum due to similar reasons. Since $i\not{D}$ is an Hermitian operator all of its eigenvalues are real. Let λ be an eigenvalue of $i\not{D}$ with eigenfunction ψ,

$$i\not{D}\psi = \lambda\psi.$$

Since $\not{D}\gamma^5 = -\gamma^5\not{D}$ it holds that

$$i\not{D}(\gamma^5\psi) = -\gamma^5(i\not{D}\psi) = -\lambda(\gamma^5\psi).$$

Therefore we may conclude that for any eigenvalue λ of $i\not{D}$ with eigenfunction ψ we have that $-\lambda$ is also an eigenvalue of $i\not{D}$ only with eigenfunction $\gamma^5\psi$, i.e. all nonzero eigenvalues come in pairs $(\lambda, -\lambda)$. Note that when $i\not{D}$ has a zero eigenvalue Theorem 6.2 is automatically true. Therefore we may exclude any zero eigenvalues and satisfactorily define the square root of the determinant $\det i\not{D}(A_\mu) = \det i\not{D}$ as the product of all its positive eigenvalues,

$$(\det(i\not{D}))^{\frac{1}{2}} = \prod_{\lambda\ \textbf{positive}\ \text{eigenvalue of}\ i\not{D}} \lambda.$$

Note that this is equivalent to picking out only "half" of the eigenvalues. That is, for any pair of eigenvalues $(\lambda, -\lambda)$ one chooses the positive one when taking the square root.

Now we consider a smooth path from A_μ to its conjugate gauge field A_μ^U and study the spectral flow. Let $0 \le t \le 1$ and define $A_\mu^t = (1-t)A_\mu + tA_\mu^U$. We recall that A_μ^U defines a topologically nontrivial gauge transformation. By gauge invariance it must hold that the eigenvalue spectrum of $i\not{D}$ in the gauge field $A_\mu^0 = A_\mu$ does not differ from the eigenvalue spectrum of $i\not{D}$ in the gauge field $A_\mu^1 = A_\mu^U$. Although the eigenvalue spectrum is not different at $t = 0$ and $t = 1$ it might rearrange itself. Thanks to a very deep theorem in mathematics we know exactly what kind of rearrangements are not allowed.

It is the Atiyah-Singer index theorem (see [**AtiSin**]) which tells us what kind of rearrangements may occur.

Let λ_1, λ_2, ... be the positive eigenvalues of $i\not{D}$. A possible rearrangement of the eigenvalue spectrum might be

$$\lambda_1(t=1) = -\lambda_1(t=0) = -\lambda_1$$
$$\lambda_i(t=1) = \lambda_i(t=0) \text{ for all } i \ge 2.$$

One sees that solely one positive eigenvalue *flows* to its negative partner. When one calculates the determinant one sees that there will be a sign change. An example of a rearrangement which is not allowed by the Atiyah-Singer index theorem is given by

$$\lambda_i(t=1) = -\lambda_i(t=0) \text{ for } i = 1, 2$$
$$\lambda_i(t=1) = \lambda_i(t=0) \text{ for all } i \ge 3.$$

In this case there is no sign change since two eigenvalues change sign. Now, although the Atiyah-Singer index theorem allows very complicated rearrangements we may assume that all rearrangements look like[2]

$$\lambda_i(t=1) = \pm\lambda_i(t=0).$$

Therefore it is clear that when one studies the spectral flow of $i\not{D}$ that the ambiguity of the sign arises from the fact that (apparently) only an odd number of eigenvalues may change sign.

Let ν be the number of eigenvalues (of $i\not{D}$) which have changed sign from positive to negative (or from negative to positive). This number being finite follows from the fact that $i\not{D}$ is a so called elliptic operator. Note that it is clear that

$$[\det i\not{D}(A_\mu)]^{\frac{1}{2}} = (-1)^\nu [\det i\not{D}(A_\mu^U)]^{\frac{1}{2}}.$$

THEOREM 6.3. Consider an $SU(2)$ gauge theory with a single doublet of Weyl fermions. It holds that $\nu \equiv 1 \pmod 2$, i.e. the number of eigenvalues that change sign from positive to negative is odd. As a consequence Theorem 6.2 follows.

Before we give the proof we introduce the five-dimensional Dirac operator $\not{D}^5$ defined by

$$\not{D}^5 = \sum_{i=1}^{5} \gamma_i \left(\partial_i + \sum_{a=1}^{3} A_i^a T^a \right).$$

Recall that T^1, T^2, T^3 are the generators of $SU(2)$. The five-dimensional Dirac equation is by definition given by

$$\not{D}^5 \psi = \sum_{i=1}^{5} \gamma_i \left(\partial_i + \sum_{a=1}^{3} A_i^a T^a \right) \psi = 0.$$

Note that the spinor ψ has eight components, the gamma matrices can be taken to be real, symmetric 8×8 matrices and the anti-hermitian generators T^a of $SU(2)$ can be taken to be real, anti-symmetric matrices. This can be seen from the fact that the spinor representation of $O(4)$ is a four-dimensional pseudoreal representation and the spinor representation of $SU(2)$ is a two-dimensional pseudoreal representation. Hence the spinor representation of $O(4) \otimes_{\mathbf{R}} SU(2)$ is real.

[2]This is actually not true. But the reader should just keep in mind that ν (which will be defined shortly) is finite and therefore what really counts is just how many eigenvalues change their sign.

One could also take ψ to be an element of the $\mathbf{H}\times\mathbf{H}$ where $\mathbf{H}$ is the quaternion algebra and act on the left by a $U_2(\mathbf{H}) = O(4)$ and on the right by $U_1(\mathbf{H}) = SU(2)$. Both arguments let us conclude that the five-dimensional Dirac operator $\not{D}^5$ for an $SU(2)$ doublet is a real, anti-symmetric operator acting on an infinite-dimensional space.

Let T be a real, antisymmetric operator acting on an infinite-dimensional space. Suppose that $T\psi = \lambda\psi$. Since T is real and antisymmetric, $\lambda^* = -\lambda$. Therefore all eigenvalues are imaginary. We conclude that the eigenvalues of T either vanish or are imaginary and occur in complex conjugate pairs.

Definition 6.4. The *mod two index* of T is the number of zero eigenvalues of T modulo 2.

Consider the five-dimensional cylinder $\mathbf{S}^4 \times \mathbf{R}$. We denote an element of $\mathbf{S}^4 \times \mathbf{R}$ by (x^σ, τ) where $\sigma = 1, 2, 3, 4$. The idea is to define an instanton-like gauge field $(A_\mu(x^\sigma, \tau), A_\tau)$ on $\mathbf{S}^4 \times \mathbf{R}$ such that at "one side" of the cylinder we have A_μ and at the "other side" we have A_μ^U. Thus, for all $(x^\sigma, \tau) \in \mathbf{S}^4 \times \mathbf{R}$ we take $A_\tau = 0$. Furthermore, we define $A_\mu(x^\sigma, \tau)$ for $\mu = 1, 2, 3, 4$ to be, for any fixed τ, a four-dimensional gauge field described as follows. For $\tau \to -\infty$ we make $A_\mu(x^\sigma, \tau)$ approach the four-dimensional gauge field $A_\mu = A_\mu(t = 0)$. For $\tau \to \infty$ we make $A_\mu(x^\sigma, \tau)$ approach the four-dimensional gauge field $A_\mu^U = A_\mu(t = 1)$. Finally, we assume that as τ varies from $-\infty$ to ∞ the five-dimensional gauge field $A_\mu(x^\sigma, \tau)$ varies *adiabatically* from A_μ to its conjugate A_μ^U. One might wonder why we did not consider the cylinder $\mathbf{S}^4 \times [0, 1]$ instead of $\mathbf{S}^4 \times \mathbf{R}$ and let the gauge field vary from $\tau = 0$ to $\tau = 1$. This is because we need a smooth interval without boundaries since we wish to differentiate with respect to τ.

Proposition 6.5. The mod two index of $\not{D}^5$ is a topological invariant and, in the five-dimensional gauge field $(A_\mu(x^\sigma, \tau), A_\tau)$ defined above, equals 1.

Proof. To show that the mod two index of $\not{D}^5$ is a topological invariant it suffices to note that the number of zero eigenvalues of $\not{D}^5$ changes if and only if a complex conjugate pair of eigenvalues move away or towards the origin. For the second part of this proposition we refer the reader to the Atiyah-Singer index theorem in [**AtiSin**]. The proof is based on the fact that $\not{D}^5$ belongs to a special class of operators called elliptic operators. □

Proof. (**Theorem 6.3**) By Proposition 6.5, it suffices to show that the mod two index of $\not{D}^5$ equals ν modulo 2. Note that we do not show that ν equals the mod two index of $\not{D}^5$ which is generally not true. They solely belong to the same class of $\mathbf{Z}/2\mathbf{Z}$.

Let us begin by noting that the Dirac equation $\not{D}^5\psi = 0$ can be written as

$$\sum_{i=1}^{4} \gamma_i \left(\partial_i + \sum_{a=1}^{3} A_i^a T^a \right) \psi = -\gamma^\tau (\partial_\tau + \sum_{a=1}^{3} A_\tau^a T^a)\psi = -\gamma^\tau \frac{d}{d\tau}\psi$$

since the fifth coordinate is the "time" τ and A_τ was chosen to be zero. Introducing the four-dimensional Dirac operator

$$\not{D}^4 = \sum_{\mu=1}^{4} \gamma^\mu D_\mu = \sum_{i=1}^{4} \gamma_i \left(\partial_i + \sum_{a=1}^{3} A_i^a T^a \right)$$

we can write the Dirac equation as

$$\frac{d}{d\tau}\psi = -\gamma^\tau \not{D}^4 \psi$$

since $\gamma^\tau\gamma^\tau = 1$. Since $A_\mu(x^\sigma, \tau)$ evolves adiabatically we may solve the Dirac equation in the adiabatic approximation. Let $\phi^\tau(x^\mu)$ be a smoothly evolving solution of the eigenvalue equation

$$\gamma^\tau \not{D}^4 \phi^\tau(x^\mu) = \lambda(\tau)\phi^\tau(x^\mu).$$

Substituting $\psi(x^\mu, \tau) = F(\tau)\phi^\tau(x_\mu)$ in the Dirac equation shows us that

$$\frac{d}{d\tau}F(\tau)\phi^\tau(x^\mu) = \phi^\tau(x^\mu)\frac{d}{d\tau}F(\tau) = -\gamma^\tau \not{D}^4(F(\tau)\phi^\tau(x^\mu)) = -F(\tau)\gamma^\tau \not{D}^4 \phi^\tau(x^\mu) = -\lambda(\tau)F(\tau)\phi^\tau(x^\mu).$$

Therefore in the adiabatic limit the Dirac equation reduces to

$$\frac{d}{d\tau}F(\tau) = -\lambda(\tau)F(\tau).$$

The latter can be written as

$$\frac{dF(\tau)}{F(\tau)} = -\lambda(\tau)d\tau$$

showing that

$$\log(F(\tau)) + C = -\int_0^\tau \lambda(\tau')d\tau'.$$

The solution therefore yields

$$F(\tau) = F(0)\exp\left(-\int_0^\tau \lambda(\tau')d\tau'\right).$$

When trying to normalize this one sees that the eigenvalue $\lambda(\tau)$ should be positive when $\tau \to \infty$. Conversely one must have that $\lambda(\tau)$ is negative when $\tau \to -\infty$. Now, recall that for $\tau \to \infty$ our gauge field is equal to A_μ^U whereas when $\tau \to -\infty$ our gauge field equals A_μ. Since $\lambda(\tau)$ is continuous, the equation $\lambda(\tau) = 0$ has a solution implying that as we pass from A_μ to A_μ^U our eigenvalue changes sign.

Thus to give a solution of the Dirac equation is the same as to say that our eigenvalue $\lambda(\tau)$ changes sign. In other words we have that in the adiabatic approximation the number of zero eigenvalues of $\not{D}^5$, i.e. the number of solutions to the Dirac equation, is exactly equal to ν. When we make corrections to our solution we see that the number of zero eigenvalues can only be counted modulo two. Therefore we lose the exact equality and only keep the equality modulo 2, i.e. the number of zero eigenvalues of $\not{D}^5$ is equal to ν modulo two. $\square$

5. Generalizations

We start with generalizing the result of Section 3 to an $SU(2)$ gauge theory with an arbitrary number of doublets of Weyl fermions. We shall see that only the theories for which the number of doublets of Weyl fermions is odd are anomalous. Also, we shall consider other gauge theories and determine whether the global anomaly is absent or not. We conclude with a result on theories that contain higher dimensional $SU(2)$ representations.

Consider an $SU(2)$ theory with m doublets of Dirac fermions. The fermion integral is easily seen to be equal to

$$I_{Dirac} = \int (d\psi_1 d\overline{\psi}_1 \ldots d\psi_m d\overline{\psi_m})_{Dirac}\ \exp(\overline{\psi} i\not{D}\psi) = \det(i\not{D})^m.$$

As before, in this case the theory may be defined satisfactorily with the aid of Pauli-Villars regularization. Now, consider an $SU(2)$ theory with n doublets of Weyl fermions. The fermion integral in this case is analogously to before seen to equal (det $i\not{D})^{\frac{n}{2}}$. When n is even the fermion integral equals some integer power of $\det(i\not{D})$ and we may again use Pauli-Villars regularization to define this expression satisfactorily. When n is odd the theory suffers from the same inconsistency as before. More precisely, given a gauge field A_μ, it holds that

$$[\det i\not{D}(A_\mu)]^{\frac{n}{2}} = (-1)^{\frac{n}{2}}[\det i\not{D}(A_\mu^U)]^{\frac{n}{2}}.$$

We conclude that the theories for which n is even are anomaly free and the theories for which n is odd are anomalous. In sum, the global $SU(2)$ anomaly is absent for any number of Dirac fermions and only arises for an odd number of Weyl fermions.

We have shown that a pure $SU(2)$ gauge theory is inconsistent if the number of Weyl fermion doublets is odd. Since the fermion integration in this case is necessarily either even or odd under the large gauge transformation U, if it is odd under U in a pure $SU(2)$ gauge theory, it remains odd under U if additional gauge[3] or Yukawa[4] couplings are smoothly switched on. Let us consider the standard $SU(3) \times SU(2) \times U(1)$ model of strong, weak, and electromagnetic interaction. By the above the standard model would be ill-defined, i.e. contain a global anomaly, if the number of Weyl fermion doublets were odd. Evidently the standard model is consistent and all there remains is to check the number of Weyl fermions. Firstly we note that the $SU(3)$ part is harmless because the coupling of the quark currents to the gluonic gauge fields is vector-like so that, effectively, only Dirac fermions are involved. For the electroweak subgroup of the standard group, $SU(2) \times U(1)$, the total number of weak isospin doublets - for the quarks and leptons of a single generation - equals 4. For example, for the quarks and leptons of the (d, u) generation we have 3 lefthanded chiral quark doublets (red, blue and green) and 1 is a lefthanded lepton doublet (one "color" per lepton). So for each generation of quarks and leptons, the numbers of chiral (weak isospin) doublets are separately odd but the total number is even so that Witten's anomaly is absent.

Recall that the key point in obtaining the global anomaly was Theorem 6.2. Therefore we expect non-trivial conditions in gauge theories for which the fourth homotopy group of the gauge group equals $\mathbf{Z}/2\mathbf{Z}$ to be anomalous. Since $\pi_4(Sp(N)) = \mathbf{Z}/2\mathbf{Z}$ we conclude that non-trivial conditions arise for $Sp(N)$ gauge theories. Note that $Sp(1) = SU(2)$. Since $\pi_4(SU(N)) = 0$ for any $N > 2$ and $\pi_4(O(N)) = 0$ for any $N > 5$ we do not expect global anomalies in these gauge theories. As a consequence theories such as quantum electrodynamics and quantum chromodynamics do not contain Witten's global anomaly.

It has been shown that an $SU(2)$ gauge theory with a single doublet of Weyl fermions in the fundamental[5] representation is anomalous. Let us turn to $SU(2)$ theories with a single doublet of Weyl fermions in a higher dimensional representation. Among the three generators of $SU(2)$ we normalize the diagonal generator T_3 such that the trace of its square tr $T_3^2 = \frac{1}{2}$ in the fundamental representation. More explicitly, we normalize T_3 such that

$$T_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}$$

in the fundamental representation. Let us now consider an $SU(2)$ theory with a single doublet of Weyl fermions in the representation with highest weight d. The Atiyah-Singer index theorem (See [**AtiSin**]) gives us that only those representations for which

$$2\mathrm{tr}\ T_3^2 \equiv 1 \pmod 2$$

are anomalous. This agrees perfectly with the fact that the fundamental representation is anomalous since $2\mathrm{tr}\ T_3^2 = 1 \equiv 1 \pmod 2$ in the fundamental representation.

THEOREM 6.6. The representations for which the highest weight $d = 2l + \frac{1}{2}$ $(l = 0, 1, 2, \ldots)$ are anomalous. For all other representations the anomaly is absent.

PROOF. With the above normalization the diagonal generator T_3 acts as

$$T_{3,d} = \begin{pmatrix} d & 0 & \ldots & 0 \\ 0 & d-1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -d \end{pmatrix}$$

in the representation of highest weight d. Note that we added a d in the index to emphasize that we are working in the representation of highest weight d. It holds that

$$2\mathrm{tr}\ T_{3,d}^2 = 2\sum_{i=0}^{2d} (d-i)^2 = \frac{4}{3}d^3 + 2d^2 + \frac{2}{3}d.$$

[3]In this case we mean additional $U(1)$ gauge and $SU(3)$ gauge couplings.

[4]Given two spin $\frac{1}{2}$ fields we can form either a spin 0 or a spin 1 combination. The couplings to a spin 0 field are the Yukawa couplings.

[5]The two-dimensional representation. This representation coincides with the spinor representation. The action of $SU(2)$ on this two-dimensional space is just the natural action.

By the result above, only those representations for which $2\mathrm{tr}\, T_{3,d} \equiv 1 \pmod 2$ are anomalous. Let us begin with the representations for which d is an integer, i.e. $d = 0, 1, 2, \ldots$. Then it is easy to see that $\frac{4}{3}d^3 + 2d^2 + \frac{2}{3}d \equiv 0 \pmod 2$. From this we may conclude that the integer representations are anomaly free. Let us turn to the half-integer representations. For convenience we write $d = f - \frac{1}{2}$ with $f = 1, 2, \ldots$. Substituting this in our expression gives

$$\frac{4}{3}d^3 + 2d^2 + \frac{2}{3}d = \frac{4}{3}f^3 - \frac{1}{3}f \equiv -\frac{1}{3}f \pmod 2 \equiv f \pmod 2.$$

We see that when f is odd the theory is anomalous. We conclude that only those half-integer representations for which $d = \underbrace{2l+1}_{f \text{ is odd}} - \frac{1}{2} = 2l + \frac{1}{2}$ $(l = 0, 1, 2 \ldots)$ are anomalous. □

Not only the fundamental representation with $d = \frac{1}{2}$, also the representations with $d = \frac{5}{2}, \frac{9}{2}, \frac{13}{2}$ etc. suffer from Witten's global anomaly. One might wonder what causes only "half" of the half-integer representations to be anomalous. In [**Bar**] it is attempted to explain this unusual pattern. One actually tries to find a general formula for the number of eigenvalues that change sign as a function of the highest weight of the $SU(2)$ representation. Unfortunately this is not a very easy task.

Bibliography

[Edix] S.J. Edixhoven, *Lie groups and Lie algebras.* Syllabus DEA 2000-2001. At author's homepage:

$http://www.math.leidenuniv.nl/edix/public_html_rennes/cours/dea0001.html.$

[Lke] M. Lübke, *Introduction to Manifolds.* Syllabus. At author's homepage:

$http://www.math.leidenuniv.nl/\sim lubke/IntroMani/Introtomanifolds.pdf.$

[Dor] F. Doray, *Homological Algebra.* Syllabus. At author's homepage:

$http://www.math.leidenuniv.nl/\sim doray/HomAlg/index.html.$

[Adams] J. F. Adams, *Vector fields on spheres,* Ann. of Math. 75 (1962), 603632.

[Argu] M. Arkowitz and M. Gutierrez, *Comultiplications on Free Groups and Wedges of Circles.* Transactions of the American Mathematical Society, Volume 350, Number 4, pages 1663-1680.

[Arms] M.A. Armstrong, *Basic Topology.* Springer-Verlag 1983, New York.

[AtiSin] M.F. Atiyah, I.M. Singer, Ann. Math. 93 (1971), page 139.

[Bar] O. Bär, *On Witten's global anomaly for higher $SU(2)$ representations.* Nuclear Physics B 650 [PM] 2003, 522-542.

[Ful] W. Fulton, *Algebraic Topology: A First Course.* Springer-Verlag 1995, New York.

[Grif1] D.J. Griffiths, *Introduction to Electrodynamics, Third Edition.* Prentice-Hall International, Inc. 1999, London.

[Grif2] D.J. Griffiths, *Introduction to Elementary Partiles.* Wiley-VCH Verlag, Weinheim.

[Hart] R. Hartshorne, *Algebraic Geometry.* Springer Science + Business Media 2006, New York.

[Hat] A. Hatcher, *Algebraic Topology.* Cambridge University Press 2002.

[Mac] Saunders Mac lane, *Categories for the Working Mathematician.* Springer-Verlag 1971, New York.

[Mars] R.E. Marshak, *Conceptual Foundations of Modern Particle Physics.* World Scientific Publishing Co. Pte. Ltd. 1993, Singapore.

[Onis] A.L. Onishchik, V.V. Gorbatsevich and E.B. Vinberg. *Lie groups and Lie algebras I.* Springer Verlag 1991, New York.

[Serre] J-P. Serre *Lie Algebras and Lie Groups.* Springer Verlag Lecture Notes in Mathematics 1500, 2nd edition, 1992.

[War] F.W. Warner, *Foundations of Differentiable Manifolds and Lie Groups.* Scott, Foresman and Company 1971, Illinois.

[WitSm] B. de Wit and J. Smith, *Field Theory in Particle Physics, Volume 1.* Elsevier Science Publishing Company 1986, Amsterdam.

[Witt] E. Witten, *An $SU(2)$ Anomaly.* Phys. Lett. B 117 (1982), pages 324-328.

[YaMi] C.N. Yang and R. Mills, Phys. Rev. 96, 191 (1954).