Applied Mathematical Sciences

Gregory L. Naber

Topology, Geometry and Gauge fields

Interactions

Second Edition

Springer

Applied Mathematical Sciences

Volume 141

For other titles published in this series, go to
http://www.springer.com/series/34

Gregory L. Naber

Topology, Geometry and Gauge fields

Interactions

Second Edition

Springer

Gregory L. Naber
Drexel University
Department of Mathematics
Korman Center
3141 Chestnut Street
Philadelphia, Pennsylvania 19104-2875
USA
gln22@drexel.edu

ISSN 0066-5452
ISBN 978-1-4419-7894-3 e-ISBN 978-1-4419-7895-0
DOI 10.1007/978-1-4419-7895-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011923654

Mathematics Subject Classification (2010): 53-01, 53B21, 55-01, 70S15

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This book is for my son, Aaron, and my stepson, Mike Scarborough.

Preface

This volume is intended to carry on the program initiated in **Topology, Geometry and Gauge Fields**: **Foundations** (henceforth, [**N4**]). It is written in much the same spirit and with precisely the same philosophical motivation: Mathematics and physics have gone their separate ways for nearly a century now and it is time for this to end. Neither can any longer afford to ignore the problems and insights of the other. Why are Dirac magnetic monopoles in one-to-one correspondence with the principal $U(1)$-bundles over S^2? Why do Higgs fields fall into topological types? What led Donaldson, in 1980, to seek in the Yang-Mills equations of physics for the key that unlocks the mysteries of smooth 4-manifolds and what physical insights into quantum field theory led Witten, fourteen years later, to propose the vastly simpler, but equivalent Seiberg-Witten equations as an alternative? We do not presume to answer these questions here, but only to promote an atmosphere in which both mathematicians and physicists recognize the need for answers. More succinctly, we shall endeavor to provide an exposition of elementary topology and geometry that keeps one eye on the physics in which our concepts either arose independently or have been found to lead to a deeper understanding of the phenomena.

Chapter 1 provides a synopsis of the geometrical background we assume of our readers (manifolds, Lie groups, bundles, connections, etc.). While all of this material is discussed in detail in [**N4**], most of it is standard and it will not matter where the background has been acquired. There follows a rather long chapter on physics. The discussion here is informal and heuristic and often anticipates topological and geometrical concepts that will be introduced precisely only much later. We begin by describing a general mathematical framework for the classical gauge theories of physics and then discuss in some detail a number of specific examples. These include classical electromagnetic theory and Dirac monopoles, the Klein-Gordon and Dirac equations and $SU(2)$ Yang-Mills-Higgs theory. The real purpose here is to witness such concepts as de Rham cohomology, Chern classes and spinor structures arise of their own accord in meaningful physics.

The mathematical development resumes in Chapter 3 where we collect some basic technical tools and then study various types of frame bundles. Minkowski spacetime, and then more general spacetime manifolds are introduced and some concrete examples are discussed (the Einstein-de Sitter spacetime, de Sitter spacetime and the Einstein cylinder). With this machinery we can define precisely the notion of a spinor structure, seen in Chapter 2 to be the device required to model spin $\frac{1}{2}$ particles in the presence of gravity. Finding the topological obstruction to the existence of such a structure (the 2^{nd} Stiefel-Whitney class) will have to wait until Chapter 6.

Chapter 4 contains a more or less standard exposition of multilinear algebra, differential forms, integration and Stokes' Theorem, but with rather more attention paid to vector-valued forms than is customary. In particular, there is a detailed discussion of tensorial forms on principal bundles and their covariant exterior derivatives in the presence of a connection. It is these derivatives that appear in the field equations of physics.

de Rham cohomology is the subject of Chapter 5. Explicit calculations are based on the Mayer-Vietoris sequence and we introduce the Brouwer degree of a map between two compact, connected, orientable n-manifolds by showing that such a manifold has 1-dimensional n^{th} cohomology. We prove that the degree is an integer by showing how to calculate it from a critical value of the map. It is such a degree that gives rise to the topological quantum number of the Higgs field in $SU(2)$ Yang-Mills-Higgs theory. The chapter concludes with a discussion of the Hopf invariant and its explicit calculation for the complex Hopf map.

The notion of a characteristic class arises on several occasions in the physical considerations of Chapter 2 (the magnetic charge of a Dirac monopole, topological charge of an instanton, and the obstruction to the existence of a spinor structure on a spacetime manifold) and Chapter 6 takes up the subject in earnest. We construct the Chern-Weil homomorphism and from it the Chern classes of a principal bundle. We prove that $U(1)$-bundles over S^2 are characterized (up to equivalence) by their 1^{st} Chern class and that $SU(2)$-bundles over S^4 are similarly characterized by the 2^{nd} Chern class. These results complete the identification of magnetic charge and instanton number with purely topological objects. The chapter concludes with the construction of the $\mathbb{Z}_2$-Čech cohomology of a smooth manifold from a simple cover. We build the 1^{st} and 2^{nd} Stiefel-Whitney classes for a semi-Riemannian manifold and prove that the former is the obstruction to orientability, while the latter, for a spacetime, is the obstruction to the existence of a spinor structure.

Seiberg-Witten gauge theory is in some sense analogous to the spin $\frac{1}{2}$-electrodynamics discussed in Chapter 2, but sprang from quite different soil and its significance is of a very different sort. Although much of this story lies in greater depths than we have reached in the main body of the text, the profound significance of the subject for both mathematics and physics would seem to make some attempt to tell at least part of it a moral imperative. The Appendix, which is a much expanded version of the Appendix that appeared in the first edition of this work, is a modest attempt to relate as much of the story as we can with the machinery we have available and should be considered a continuation of Appendix B in [**N4**].

There are 228 Exercises in the book. Each is an integral part of the development and has been included (rather than the equivalent term "clearly") to encourage active participation on the part of the reader.

Gregory L. Naber
2010

Acknowledgments

The California State University at Chico and, in particular, its College of Natural Sciences and Department of Mathematics have provided much support over the years without which this and many similar projects would not have been possible. For this I wish to express my sincere gratitude. Special thanks are also due to my friends at San Bernardino State University, both mathematicians and physicists, who have listened patiently and contributed much to lectures in which I attempted to sort out the final form of the presentation. Finally, and as always, my greatest debt is to my wife, Debora. She has devoted much of her life to typing these manuscripts of mine and the rest to seeing to it that I have the sort of life in which it is possible to write them. I can't imagine doing this job without her.

Contents

1
Geometrical Background

In this preliminary chapter we have gathered together a brief synopsis of those items from differential geometry upon which the main development of the text will be built. Most of the material is standard so that more detailed expositions are readily available, although, when references seem called for, we will tend to favor [**N4**].

1.1 Smooth Manifolds and Maps

A **topological manifold** of **dimension** n is a second countable, Haus-dorff topological space X with the property that, for each $p \in X$, there exists an open set U in X containing p and a homeomorphism φ of U onto an open subset $\varphi(U)$ in ^{n}. The pair (U, φ) is called a **chart** at p. If (U_1, φ_1) and (U_2, φ_2) are two charts with $U_1 \cap U_2 \neq \emptyset$, then the **overlap functions** $\varphi_1 \circ \varphi_2^{-1} : \varphi_2(U_1 \cap U_2) \longrightarrow \varphi_1(U_1 \cap U_2)$ and $\varphi_2 \circ \varphi_1^{-1} : \varphi_1(U_1 \cap U_2) \longrightarrow \varphi_2(U_1 \cap U_2)$ are inverse homeomorphisms. Two charts (U_1, φ_1) and (U_2, φ_2) are $\boldsymbol{C^\infty}$**-related** if either $U_1 \cap U_2 = \emptyset$, or $U_1 \cap U_2 \neq \emptyset$ and the overlap functions are C^∞ (i.e., have continuous partial derivatives of all orders and types). An **atlas** for X is a collection $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in \mathcal{A}}$ of charts, any two of which are C^∞-related and with $\bigcup_{\alpha \in \mathcal{A}} U_\alpha = X$. A chart is **admissible** to an atlas if it is C^∞-related to every chart in the atlas and the atlas is said to be **maximal** if it contains every chart that is admissible to it. Every atlas is contained in a unique maximal atlas (Theorem 5.3.1, [**N4**]). A maximal atlas is called a **differentiable structure** for X. A topological manifold together with some differentiable structure is called a **differentiable** (or **smooth**, or $\boldsymbol{C^\infty}$) **manifold** (examples are forthcoming). The integer n is called the **dimension** of X and is denoted $\dim X$.

Remark: Some topological manifolds admit no differentiable structures at all, while others admit many (see Appendix B of [**N4**]).

Let X and Y be smooth manifolds of dimension n and m, respectively, and let $f : X \longrightarrow Y$ be a continuous map. Let (U, φ) be a chart on X and (V, ψ) a chart on Y with $U \cap f^{-1}(V) \neq \emptyset$. Then $\psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \longrightarrow \psi(U \cap f^{-1}(V))$ is called the **coordinate expression** for f relative to these two charts. f is said to be **smooth** (or $\boldsymbol{C^\infty}$) if its coordinate expressions are C^∞ for all such charts in some atlases for X and Y. A smooth bijection with a smooth inverse is a **diffeomorphism** and, if a diffeomorphism of X onto Y exists, we say that X and Y are **diffeomorphic**. Providing the

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0_1,

real line with its standard differentiable structure (determined by the atlas consisting of the single chart $(\ , id))$ we denote by $C^\infty(X)$ the set of all smooth, real-valued functions on X and provide it with the obvious (pointwise) structure of a commutative algebra with identity.

A **tangent vector** at $p \in X$ is a real-valued function $\boldsymbol{v} : C^\infty(X) \longrightarrow$ that is linear and satisfies the **Leibnitz Product Rule** $\boldsymbol{v}(fg) = f(p)\boldsymbol{v}(g) + \boldsymbol{v}(f)g(p)$ for all $f, g \in C^\infty(X)$. The collection of all such is denoted $T_p(X)$, called the **tangent space** to X at p and provided with the natural pointwise structure of a real vector space. The dimension of $T_p(X)$ as a vector space over is the same as the dimension of X as a manifold. Indeed, if (U, φ) is a chart at p in X with coordinate functions $x^i, i = 1, \ldots, n$ $(\varphi(p) = (x^1(p), \ldots, x^n(p)))$, then the linear maps $\partial_i|_p = \frac{\partial}{\partial x^i}|_p \colon C^\infty(X) \longrightarrow$ defined by $\frac{\partial}{\partial x^i}|_p(f) = D_i(f \circ \varphi^{-1})(\varphi(p))$ are in $T_p(X)$ and $\{\frac{\partial}{\partial x^1}|_p, \ldots, \frac{\partial}{\partial x^n}|_p\}$ is a basis for $T_p(X)$ (Theorem 5.5.3, [**N4**]). Any $\boldsymbol{v} \in T_p(X)$ can be uniquely written as $\boldsymbol{v} = \boldsymbol{v}(x^i)\frac{\partial}{\partial x^i}|_p$ (summation convention). If (a, b) is an open interval in provided with its standard differentiable structure (determined by the atlas consisting of the single chart $((a, b), \iota)$, where $\iota : (a, b) \hookrightarrow$ is the inclusion map), then a smooth map $\alpha : (a, b) \longrightarrow X$ is a **smooth curve** in X. Fix $t_0 \in (a, b)$ and let $p = \alpha(t_0)$. The **velocity vector** of α at t_0 is the map $\alpha'(t_0) : C^\infty(X) \longrightarrow$ defined by $(\alpha'(t_0))(f) = D_1(f \circ \alpha)(t_0)$ for each f in $C^\infty(X)$. Then $\alpha'(t_0)$ is in $T_p(X)$ and, indeed, every element of $T_p(X)$ is the velocity vector of some smooth curve in X through p (Corollary 5.5.6, [**N4**]).

If $f : X \longrightarrow Y$ is a smooth map and $p \in X$, then the **derivative** of f at p is the linear map $f_{*p} : T_p(X) \longrightarrow T_{f(p)}(Y)$ defined as follows: For each $\boldsymbol{v} \in T_p(X)$, $f_{*p}(\boldsymbol{v}) : C^\infty(Y) \longrightarrow$ is given by $(f_{*p}(\boldsymbol{v}))(g) = \boldsymbol{v}(g \circ f)$ for all $g \in C^\infty(Y)$. If $\boldsymbol{v} = \alpha'(t_0)$ for some smooth curve α, then $f_{*p}(\boldsymbol{v}) = f_{*p}(\alpha'(t_0)) = (f \circ \alpha)'(t_0)$. f is said to be an **immersion at** p if f_{*p} is one-to-one and an **immersion** if this is true at each $p \in X$. f is a **submersion at** p if f_{*p} is onto and a **submersion** if this is true at each $p \in X$. An immersion that is also a homeomorphism onto its image is an **imbedding**. A point $q \in Y$ is a **regular value** of f if, for every $p \in f^{-1}(q)$, f is a submersion at p (this is the case, in particular, if $f^{-1}(q) = \emptyset$); otherwise, q is a **critical value** of f.

If X' is an open subspace of X and $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in \mathcal{A}}$ is an atlas for X, then $\{(U_\alpha \cap X', \varphi_\alpha | U_\alpha \cap X') : \alpha \in \mathcal{A}, U_\alpha \cap X' \neq \emptyset\}$ is an atlas for X' and, with the differentiable structure determined by this atlas, X' is called an **open submanifold** of X. Note that $\dim X' = \dim X$. More generally, if $\dim X = n$ and $1 \leq k \leq n$ is an integer, then a topological subspace X' of X is called a ***k*-dimensional submanifold** of X if, for each $p \in X'$, there exists a chart (U, φ) in the differentiable structure for X such that $\varphi(U \cap X') = \{(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) \in \varphi(U) : x^{k+1} = \cdots = x^n = 0\}$. For each such (U, φ) one obtains a chart $(U \cap X', \varphi')$ for X', where φ' is $\varphi | U \cap X'$ followed by the projection of $\ ^n = \ ^k \times \ ^{n-k}$ onto $\ ^k$. The collection of all such $(U \cap X', \varphi')$ is an atlas for X' and so determines a differentiable structure for X'. A **0-dimensional submanifold** of X is a discrete subspace of X. Restrictions of C^∞ maps on X to submanifolds are C^∞ with respect to this submanifold differentiable structure. According to the Inverse Function Theorem, if X and Y have the same dimension and $f : X \longrightarrow Y$ is smooth, then $f_{*p} : T_p(X) \longrightarrow T_{f(p)}(Y)$ is an isomorphism if and only if f is a local diffeomorphism at p, i.e., if and only if there exist open neighborhoods V of p and W of $f(p)$ such that $f|V$ is a diffeomorphism of V onto W (Corollary 5.5.8, [**N4**]).

If $f : X \longrightarrow Y$ is an imbedding, then its image $f(X)$ is a submanifold of Y and, regarded as a map of X onto $f(X)$, f is a diffeomorphism (Corollary 5.6.6, [**N4**]). On the other hand, if $f : X \longrightarrow Y$ is any smooth map and $q \in Y$ is a regular value of f (so that, in particular, $\dim X \geq \dim Y$), then $f^{-1}(q)$ is either empty or a submanifold of X of dimension $\dim X - \dim Y$ (Corollary 5.6.7, [**N4**]).

We now pause to describe a collection of examples of smooth manifolds that will play a central role in all that follows. Several methods of constructing examples are available. One can, of course, explicitly specify a second countable, Hausdorff space X and an atlas $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in \mathcal{A}}$ for it. Alternatively, one can produce smooth manifolds as submanifolds of known examples, images of imbeddings, or inverse images of regular values of smooth maps. One other technique (forming products) will be discussed after we have some manifolds available with which to form them.

1. (Standard differentiable structure on ^{n}) $X = {}^n$ with the atlas consisting of a single global chart $({}^n, id)$. A chart (U, φ) is in the differentiable structure determined by this atlas if and only if $\varphi : U \longrightarrow \varphi(U)$ is a diffeomorphism.

2. (A nonstandard differentiable structure on) $X =$ with the atlas consisting of a single global chart $(\ , \varphi)$, where $\varphi : \ \longrightarrow \ $ is given by $\varphi(x) = x^3$. The resulting differentiable structure on is not the same as the standard structure because $\varphi^{-1} : \ \longrightarrow \ $ is given by $\varphi^{-1}(y) = \sqrt[3]{y}$ and this is not C^∞ on (in the standard sense). These two manifolds (standard and nonstandard) have different differentiable structures, but are, nevertheless, diffeomorphic (Exercise, or see page 243 of [**N4**]).

3. (Natural differentiable structures on real vector spaces) Let $X = \mathcal{V}$ be a real vector space of dimension n. Select a basis $\{e_1, \ldots, e_n\}$ for $\mathcal{V}$ and let $\{e^1, \ldots, e^n\}$ be the dual basis for $\mathcal{V}^*$. Define $\varphi : \mathcal{V} \longrightarrow {}^n$ as follows: For each $v = v^a e_a \in \mathcal{V},\ \varphi(v) = (e^1(v), \ldots, e^n(v)) = (v^1, \ldots, v^n)$. Then φ is an isomorphism. Provide $\mathcal{V}$ with the topology for which φ is a homeomorphism (this is independent of the choice of basis). Now provide $\mathcal{V}$ with the differentiable structure determined by the single global chart $(\mathcal{V}, \varphi)$ (also independent of the choice of basis). Note that when $\mathcal{V} = {}^n$ this reduces to the standard differentiable structure. Also note that, for each $p \in \mathcal{V}$, the tangent space $T_p(\mathcal{V})$ can be canonically identified with $\mathcal{V}$: For each $v \in \mathcal{V}$, define $\boldsymbol{v}_p \in T_p(\mathcal{V})$ by $\boldsymbol{v}_p = \alpha'(0)$, where $\alpha : \ \longrightarrow \mathcal{V}$ is given by $\alpha(t) = p + tv$. Then $v \longrightarrow \boldsymbol{v}_p$ is the **canonical isomorphism** of $\mathcal{V}$ onto $T_p(\mathcal{V})$.

4. (Standard differentiable structure on the n-sphere S^n) $X = S^n = \{(x^1, \dots, x^n, x^{n+1}) \in {}^{n+1} : (x^1)^2 + \cdots + (x^n)^2 + (x^{n+1})^2 = 1\}$ with its topology as a subspace of ${}^{n+1}$. We define two stereographic projection charts on S^n. Let $N = (0, \dots, 0, 1)$, $S = (0, \dots, 0, -1)$, $U_S = S^n - \{N\}$ and $U_N = S^n - \{S\}$. Now define $\varphi_S : U_S \longrightarrow {}^n$ and $\varphi_N : U_N \longrightarrow {}^n$ by $\varphi_S(x^1, \dots, x^n, x^{n+1}) = (1 - x^{n+1})^{-1}(x^1, \dots, x^n)$ and $\varphi_N(x^1, \dots, x^n,\ x^{n+1}) = (1 + x^{n+1})^{-1}(x^1, \dots, x^n)$. Then $\varphi_S^{-1} : {}^n \longrightarrow U_S$ and $\varphi_N^{-1} : {}^n \longrightarrow U_N$ are given by $\varphi_S^{-1}(y) = \varphi_S^{-1}(y^1, \dots, y^n) = (1 + \|y\|^2)^{-1}(2y^1, \dots, 2y^n, \|y\|^2 - 1)$ and $\varphi_N^{-1}(y) = \varphi_N^{-1}(y^1, \dots, y^n) = (1 + \|y\|^2)^{-1}(2y^1, \dots, 2y^n, -\|y\|^2 + 1)$. Thus, on $\varphi_N(U_N \cap U_S) = \varphi_S(U_N \cap U_S) = {}^n - \{0\}$, $\varphi_S \circ \varphi_N^{-1}(y) = \varphi_N \circ \varphi_S^{-1}(y) = \|y\|^{-2} y$ and these are C^∞. Thus, $\{(U_S,\ \varphi_S), (U_N, \varphi_N)\}$ is an atlas for S^n and so determines a differentiable structure for S^n.

Remarks: The structures of S^3 and S^4 are substantially elucidated through the use of quaternions. One identifies ^{4} with the real vector space of 2×2 complex matrices of the form $\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$ with the squared norm of such a matrix taken to be its determinant (page 36, [**N4**]). Matrix multiplication then corresponds to quaternion multiplication on ^{4}. S^3 is thus the group of unit quaternions and can be identified with the special unitary group $SU(2)$ (Theorem 1.1.4, [**N4**]). Moreover, the overlap functions for the stereographic projection charts (U_S, φ_S) and (U_N, φ_N) on S^4 can be written $\varphi_S \circ \varphi_N^{-1}(y) = \varphi_N \circ \varphi_S^{-1}(y) = \bar{y}^{-1}$ for all $y \in {}^4 - \{0\} = \ - \{0\}$.

5. (Projective Spaces) Let $= \ ,\ $, or . Regard ^{n} as a right module over (a vector space if $=$ or). Let $0 = (0, \dots, 0) \in {}^n$ and define the standard bilinear form $< , >: {}^n \times {}^n \longrightarrow$ on ^{n} by $< \xi, \tau > = \bar{\xi}^1 \tau^1 + \cdots + \bar{\xi}^n \tau^n$ for all $\xi, \tau \in {}^n$. Let $S = \{\xi \in {}^n : < \xi, \xi > = 1\}$.

^{n} has the topology of ^{n}, ^{2n}, or ^{4n} depending on whether $=$, , or , so S is homeomorphic to S^{n-1}, S^{2n-1}, or S^{4n-1}. Define an equivalence relation $\sim$ on S as follows: $\tau \sim \xi$ if and only if there exists an $a \in$ with $|a| = 1$ such that $\tau = \xi a$. The equivalence class containing ξ is $[\xi] = [\xi^1, \dots, \xi^n] = \{(\xi^1 a, \dots, \xi^n a) : a \in\ , |a| = 1\}$. Each $[\xi]$ is homeomorphic to S^0, S^1, or S^3 (pages 50–51, [**N4**]). The quotient space $S/\sim$ is the (real, complex, or quaternionic) projective space ${}^{n-1}$. All of these are Hausdorff (page 51, [**N4**]), second countable (Exercise 1.3.2, [**N4**]) and, being continuous images of spheres, compact and connected.

Let $\mathcal{P} : S \longrightarrow {}^{n-1}$ be the quotient map ($\mathcal{P}(\xi) = [\xi]$). For each $k = 1, \dots, n$, let $U_k = \{[\xi] \in {}^{n-1} : \xi^k \neq 0\}$. Then $\mathcal{P}^{-1}(U_k) = \{\xi \in S : \xi^k \neq 0\}$. Define $\varphi_k : U_k \longrightarrow {}^{n-1}$ ($= {}^{n-1}, {}^{2n-2}$, or ${}^{4n-4}$) by $\varphi_k([\xi]) = \varphi_k([\xi^1, \dots, \xi^k, \dots, \xi^n]) = (\xi^1(\xi^k)^{-1}, \dots, \widehat{1}, \dots, \xi^n(\xi^k)^{-1})$ (where $\widehat{\ }$ means "deleted"). Then $\varphi_k^{-1} : {}^{n-1} \longrightarrow U_k$ is given by $\varphi_k^{-1}(y) = \varphi_k^{-1}(y^1, \dots, y^{n-1}) = [y^1, \dots, 1, \dots, y^{n-1}]$ (with the 1 in the k^{th} slot). These are inverse homeomorphisms (pages 51–52, [**N4**]) and the overlap functions $\varphi_k \circ \varphi_j^{-1} : \varphi_j(U_k \cap U_j) \longrightarrow \varphi_k(U_k \cap U_j)$ are C^∞. For example,

$\varphi_2 \circ \varphi_1^{-1}(y^2, y^3, \ldots, y^n) = \varphi_2([1, y^2, y^3, \ldots, y^n]) = ((y^2)^{-1}, y^3(y^2)^{-1}, \ldots, y^n(y^2)^{-1})$. Thus $\{(U_k, \varphi_k)\}_{k=1,\ldots,n}$ is an atlas for $\quad^{n-1}$ and so determines a differentiable structure.

Remark: When $n = 2$ we have $\quad^1 \cong S^1$, $\quad^1 \cong S^2$ and $\quad^1 \cong S^4$ (pages 53–54, [**N4**]).

6. (Classical Groups) Let $\quad = \quad, \quad$, or $\quad$ and let $\mathcal{GL}(n, \quad)$ be the set of all $n \times n$ matrices with entries in $\quad$ (the reason for the peculiar notation will emerge in Section 1.2). Topologically we identify $\mathcal{GL}(n, \quad) = \quad^{n^2}$, $\mathcal{GL}(n, \quad) = \quad^{2n^2}$ and $\mathcal{GL}(n, \quad) = \quad^{4n^2}$. Let $GL(n, \quad)$ denote the set of all invertible elements of $\mathcal{GL}(n, \quad)$. Then $GL(n, \quad)$ is an open submanifold of $\mathcal{GL}(n, \quad)$ (pages 41–43, [**N4**]). Let $U(n, \quad) = \{A \in GL(n, \quad) : A^{-1} = \bar{A}^T\}$.

$$
\begin{aligned}
\quad = \quad &: U(n, \quad) = O(n) = \text{orthogonal group of order } n \\
\quad = \quad &: U(n, \quad) = U(n) = \text{unitary group of order } n \\
\quad = \quad &: U(n, \quad) = Sp(n) = \text{symplectic group of order } n
\end{aligned}
$$

$U(n, \quad)$ is a submanifold of $GL(n, \quad)$, but this is not obvious. For example, $O(n)$ is a submanifold of $GL(n, \quad)$ because it is the inverse image of a regular value under a smooth map of $GL(n, \quad)$ into a Euclidean space (identify the set $S(n, \quad)$ of symmetric $n \times n$ real matrices with $\quad^{n(n+1)/2}$, define $f : GL(n, \quad) \longrightarrow S(n, \quad)$ by $f(A) = AA^T$ and note that $O(n) = f^{-1}(id)$). For details, see [**N4**] (pages 256–257 for $O(n)$, and Exercise 5.8.12 for $U(n)$ and $Sp(n)$). Next let

$$
\begin{aligned}
SO(n) &= \{A \in O(n) : \det A = 1\} = \text{special orthogonal group of order } n \\
SU(n) &= \{A \in U(n) : \det A = 1\} = \text{special unitary group of order } n
\end{aligned}
$$

$SO(n)$ and $SU(n)$ are submanifolds of $O(n)$ and $U(n)$, respectively.

Remarks: $SO(2) \cong U(1), SO(3) \cong \quad^3$ (pages 395 and 399, [**N4**]), and $SU(2) \cong Sp(1) \cong S^3$.

7. (Products) Let X and Y be smooth manifolds of dimension n and m, respectively. If (U, φ) is a chart on X and (V, ψ) is a chart on Y, then $(U \times V, \varphi \times \psi)$ is a chart on the topological space $X \times Y$ (the product map $\varphi \times \psi : U \times V \longrightarrow \quad^n \times \quad^m = \quad^{n+m}$ is given by $(\varphi \times \psi)(p, q) = (\varphi(p), \psi(q))$). The set of all such charts is an atlas for $X \times Y$ determining the product manifold structure on $X \times Y$. The process obviously extends to larger (finite) products. Important examples are the tori $S^1 \times S^1$, $S^1 \times S^1 \times S^1, \ldots$ and certain products of classical groups, e.g., $SU(2) \times U(1)$ (electroweak gauge group), $SU(3) \times SU(2) \times U(1)$ (standard model of elementary particles), etc.

A **vector field** on a smooth manifold X is a map $\boldsymbol{V}$ that assigns to each $p \in X$ a tangent vector $\boldsymbol{V}(p) = \boldsymbol{V}_p$ in $T_p(X)$. If (U, φ) is a chart on X with coordinate functions $x^1, \ldots, x^n$, then $p \longrightarrow \frac{\partial}{\partial x^i}|_p$ is a vector field on

U for each $i = 1, \ldots, n$. For any $\boldsymbol{V}$ and any $p \in U$, $\boldsymbol{V}(p) = \boldsymbol{V}_p(x^i) \frac{\partial}{\partial x^i}|_p$ and the functions $V^i : U \longrightarrow$ defined by $V^i(p) = \boldsymbol{V}_p(x^i)$ are called the **components** of $\boldsymbol{V}$ relative to (U, φ). $\boldsymbol{V}$ is **continuous**, **smooth**, etc. if its components are continuous, smooth, etc. for all charts in some atlas for X. The set of all smooth vector fields on X is denoted $\mathcal{X}(X)$ and has the obvious structure of a module over $C^\infty(X)$. Any $\boldsymbol{V} \in \mathcal{X}(X)$ acts on any $f \in C^\infty(X)$ to give a $\boldsymbol{V}f = \boldsymbol{V}(f)$ in $C^\infty(X)$ defined by $(\boldsymbol{V}f)(p) = \boldsymbol{V}_p(f)$ for all $p \in X$. This operator on $C^\infty(X)$ is a derivation (i.e., $\boldsymbol{V}(af + bg) = a\boldsymbol{V}f + b\boldsymbol{V}g$ and $\boldsymbol{V}(fg) = f\boldsymbol{V}g + (\boldsymbol{V}f)g$ for all $a, b \in$ and $f, g \in C^\infty(X)$). Conversely, any derivation $\mathcal{D} : C^\infty(X) \longrightarrow C^\infty(X)$ gives rise to a smooth vector field $\boldsymbol{V}$ on X defined by $\boldsymbol{V}(p)(f) = \mathcal{D}(f)(p)$. Thus, vector fields may be identified with derivations. As an application we define the **Lie bracket** $[\boldsymbol{V}, \boldsymbol{W}]$ of two smooth vector fields $\boldsymbol{V}$ and $\boldsymbol{W}$ on X to be the derivation/vector field defined by $[\boldsymbol{V}, \boldsymbol{W}](f) = \boldsymbol{V}(\boldsymbol{W}f) - \boldsymbol{W}(\boldsymbol{V}f)$. We record some useful properties of the Lie bracket (proofs on pages 263–264, [**N4**]):

$$[\boldsymbol{W}, \boldsymbol{V}] = -[\boldsymbol{V}, \boldsymbol{W}]$$

$$[a_1\boldsymbol{V}_1 + a_2\boldsymbol{V}_2, \boldsymbol{W}] = a_1[\boldsymbol{V}_1, \boldsymbol{W}] + a_2[\boldsymbol{V}_2, \boldsymbol{W}]$$

$$[f\boldsymbol{V}, g\boldsymbol{W}] = fg[\boldsymbol{V}, \boldsymbol{W}] + f(\boldsymbol{V}g)\boldsymbol{W} - g(\boldsymbol{W}f)\boldsymbol{V}$$

$$[\boldsymbol{V}_1, [\boldsymbol{V}_2, \boldsymbol{V}_3]] + [\boldsymbol{V}_3, [\boldsymbol{V}_1, \boldsymbol{V}_2]] + [\boldsymbol{V}_2, [\boldsymbol{V}_3, \boldsymbol{V}_1]] = 0$$

$$[\boldsymbol{V}, \boldsymbol{W}] = \left(V^j \frac{\partial W^i}{\partial x^j} - W^j \frac{\partial V^i}{\partial x^j} \right) \frac{\partial}{\partial x^i}$$

for all $\boldsymbol{V}, \boldsymbol{W}, \boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3 \in \mathcal{X}(X), a_1, a_2 \in$, $f, g \in C^\infty(X)$ and, for the last, any chart on X. An **integral curve** for $\boldsymbol{V} \in \mathcal{X}(X)$ is a smooth curve $\alpha : (a, b) \longrightarrow X$ whose velocity vector $\alpha'(t_0)$ at each t_0 in (a, b) coincides with the vector assigned to $\alpha(t_0)$ by $\boldsymbol{V}$, i.e., $\alpha'(t_0) = \boldsymbol{V}(\alpha(t_0))$. These always exist, at least locally (Theorem 5.7.2, [**N4**]).

Let $\mathcal{V}$ be an n-dimensional real vector space and $\{e_1, \ldots, e_n\}$ and $\{\hat{e}_1, \ldots, \hat{e}_n\}$ two ordered bases for $\mathcal{V}$. Then there exists a unique nonsingular $n \times n$ real matrix $(A^i{}_j)_{i,j=1,\ldots,n}$ such that $\hat{e}_j = A^i{}_j e_i$ for $j = 1, \ldots, n$. Since $\det(A^i{}_j) \neq 0$ we can define an equivalence relation $\sim$ on the set of all ordered bases for $\mathcal{V}$ by $\{\hat{e}_1, \ldots, \hat{e}_n\} \sim \{e_1, \ldots, e_n\}$ if and only if $\det(A^i{}_j) > 0$ and conclude that there are precisely two equivalence classes, each of which is called an **orientation** for $\mathcal{V}$. The equivalence class containing $\{e_1, \ldots, e_n\}$ is denoted $[e_1, \ldots, e_n]$.

Now let X be an n-dimensional smooth manifold and U an open subset of X. An **orientation** on U is a function μ that assigns to each $p \in U$ an orientation μ_p for $T_p(X)$ and satisfies the following smoothness condition: For each $p_0 \in U$ there is an open neighborhood W of p_0 in X with $W \subseteq U$ and smooth vector fields $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$ on W with $\{\boldsymbol{V}_1(p), \ldots, \boldsymbol{V}_n(p)\} \in \mu_p$ for each $p \in W$. For example, if (U, φ) is a chart with coordinate functions $x^1, \ldots, x^n$, then $p \longrightarrow \left[\frac{\partial}{\partial x^1}|_p, \ldots, \frac{\partial}{\partial x^n}|_p\right]$ is an orientation on U. A manifold for which an orientation exists on all of X is said to be **orientable** and such an X is said to be **oriented** by any specific choice of an orientation μ on X. A connected, orientable manifold admits precisely two orientations (Theorem 5.10.2, [**N4**]). If one of the orientations is μ, then the other is denoted $-\mu$ and called the **opposite orientation**. If X is orientable with orientation μ, then a chart (U, φ) with coordinate functions $x^1, \ldots, x^n$ is said to be **consistent with** μ if $\left\{ \frac{\partial}{\partial x^1}|_p, \ldots, \frac{\partial}{\partial x^n}|_p \right\} \in \mu_p$ for each $p \in U$. If (U, φ) and (V, ψ) are charts with coordinate functions $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$, respectively, both consistent with μ, and for which $U \cap V \neq \emptyset$, then the Jacobian of the coordinate transformation $(y^1, \ldots, y^n) = \psi \circ \varphi^{-1}(x^1, \ldots, x^n)$ must have positive determinant on $\varphi(U \cap V)$. It follows that an orientable manifold has an **oriented atlas**, i.e., an atlas whose overlap functions all have Jacobians with positive determinant. The converse is also true, i.e., a manifold is orientable if and only if it admits an oriented atlas (Exercise 5.10.9, [**N4**]). With the exception of ${}^{n-1}$ for n odd, all of the manifolds introduced in Examples 1–6 above are orientable (pages 307–309, [**N4**]). Suppose X and Y are manifolds with orientations μ and ν, respectively, and that $f : X \longrightarrow Y$ is a diffeomorphism of X onto Y. Then, for each $p \in X, f_{*p} : T_p(X) \longrightarrow T_{f(p)}(Y)$ is an isomorphism and so carries every basis for $T_p(X)$ onto a basis for $T_{f(p)}(Y)$. f is said to be **orientation preserving** if, for each $p \in X$, f_{*p} carries every basis in μ_p onto a basis in $\nu_{f(p)}$. If X is connected, then a diffeomorphism $f : X \longrightarrow Y$ is either orientation preserving or **orientation reversing** in the sense that, for each $p \in X$, it carries every basis in μ_p onto a basis in $-\nu_{f(p)}$.

The dual $T_p^*(X)$ of $T_p(X)$ is called the **cotangent space** to X at p and its elements are called **covectors** at p. A (**real-valued**) **1-form** on X is a map $\boldsymbol{\Theta}$ that assigns to each $p \in X$ a covector $\boldsymbol{\Theta}(p) = \boldsymbol{\Theta}_p$ in $T_p^*(X)$. For example, if $f \in C^\infty(X)$ we define its **exterior derivative** (or **differential**) df as follows: For each $p \in X, df(p) = df_p$ is the element of $T_p^*(X)$ whose value at $\boldsymbol{v} \in T_p(X)$ is $df(p)(\boldsymbol{v}) = df_p(\boldsymbol{v}) = \boldsymbol{v}(f)$. Then df is a 1-form on X. In particular, if (U, φ) is a chart on X with coordinate functions $x^1, \ldots, x^n$, then each dx^i is a 1-form on U and, at each $p \in U, \{dx^1{}_p, \ldots, dx^n{}_p\}$ is the basis for $T_p^*(X)$ dual to $\left\{ \frac{\partial}{\partial x^1}|_p, \ldots, \frac{\partial}{\partial x^n}|_p \right\}$. If $\boldsymbol{\Theta}$ is any 1-form, then, at each $p \in U, \boldsymbol{\Theta}(p) = \boldsymbol{\Theta}_p = \Theta_i(p) dx^i{}_p$, where $\Theta_i(p) = \boldsymbol{\Theta}(p)(\frac{\partial}{\partial x^i}|_p)$ (Exercise 5.5.14, [**N4**]). The functions Θ_i are the components of $\boldsymbol{\Theta}$ relative to (U, φ) and $\boldsymbol{\Theta}$ is said to be continuous, smooth, etc. if its components are continuous, smooth, etc. for all charts in some atlas for X. The set $\mathcal{X}^*(X)$ of all smooth 1-forms on X has the structure of a module over $C^\infty(X)$. Moreover, any $\boldsymbol{\Theta} \in \mathcal{X}^*(X)$ gives rise to a $C^\infty(X)$-module homomorphism of $\mathcal{X}(X)$ to $C^\infty(X)$ defined as follows: For every $\boldsymbol{V} \in \mathcal{X}(X)$, $\boldsymbol{\Theta}(\boldsymbol{V}) = \boldsymbol{\Theta}\boldsymbol{V}$ is given by $\boldsymbol{\Theta}(\boldsymbol{V})(p) = (\boldsymbol{\Theta}\boldsymbol{V})(p) = \boldsymbol{\Theta}_p(\boldsymbol{V}_p)$. Thus, relative to any chart, $\boldsymbol{\Theta}(\boldsymbol{V})(p) = \Theta_i(p)V^i(p)$. Conversely, every $C^\infty(X)$-module homomorphism $A : \mathcal{X}(X) \longrightarrow C^\infty(X)$ determines a unique 1-form $\boldsymbol{\Theta}$ with $\boldsymbol{\Theta}(\boldsymbol{V}) = A(\boldsymbol{V})$ for every $\boldsymbol{V} \in \mathcal{X}(X)$ (pages 265–266), [**N4**]). If the set of such homomorphisms is given its natural $C^\infty(X)$-module structure, then this correspondence is an isomorphism (Exercise 5.7.11, [**N4**]) so we will not distinguish between the 1-form and the homomorphism.

If $F : X \longrightarrow Y$ is smooth and $\boldsymbol{\Theta}$ is a 1-form on Y, then the **pullback** $F^*\boldsymbol{\Theta} \in \mathcal{X}^*(X)$ is defined by $(F^*\boldsymbol{\Theta})_p(\boldsymbol{v}) = \boldsymbol{\Theta}_{F(p)}(F_{*p}(\boldsymbol{v}))$ for each $p \in X$ and each $\boldsymbol{v} \in T_p(X)$. In coordinates, if (U, φ) is a chart on X with coordinate functions $x^1, \ldots, x^n$ and (V, ψ) is a chart on Y with coordinate functions $y^1, \ldots, y^m$ and if $\varphi(U) \subseteq V$, then $\boldsymbol{\Theta} = \Theta_i dy^i \Longrightarrow$

$$\begin{aligned} F^*\boldsymbol{\Theta} &= \frac{\partial F^i}{\partial x^j}(\Theta_i \circ F)dx^j \\ &= \Theta_i\ (F^1(x^1, \ldots, x^n), \ldots, F^m(x^1, \ldots, x^n))d(F^i(x^1, \ldots, x^n)), \end{aligned}$$

where $\psi \circ F \circ \varphi^{-1}(x^1, \ldots, x^n) = (F^1\ (x^1, \ldots, x^n), \ldots, F^m(x^1, \ldots, x^n))$. As a special case, if $g \in C^\infty(Y)$, then $F^*(dg) = d(g \circ F)$. Smooth real- valued functions are often called **0-forms** and, if one defines the pullback of a 0-form $g \in C^\infty(Y)$ by $F : X \longrightarrow Y$ to be $F^*g = g \circ F \in C^\infty(X)$, then this last result reads $F^*(dg) = d(F^*g)$, i.e., pullback commutes with the exterior derivative on 0-forms. Note that the exterior derivative operator d carries 0-forms to 1-forms. For future reference we record its basic properties as follows:

$$\begin{aligned} &d(af_1 + bf_2) = adf_1 + bdf_2 && (a, b \in \quad \text{and } f_1, f_2 \in C^\infty(X)) \\ &d(f_1 f_2) = f_1 df_2 + f_2 df_1 && (f_1, f_2 \in C^\infty(X)) \\ &F^*(dg) = d(F^*g) && (F : X \longrightarrow Y \text{ smooth and } g \in C^\infty(Y)) \end{aligned}$$

and, if (U, φ) is a chart on X with coordinate functions $x^1, \ldots, x^n$,

$$df = \frac{\partial f}{\partial x^i} dx^i \qquad (f \in C^\infty(X)).$$

If X' is a submanifold of X and $\iota : X' \hookrightarrow X$ is the inclusion map, then the **restriction** of $\boldsymbol{\Theta} \in \mathcal{X}^*(X)$ to X' is defined to be $\iota^*\boldsymbol{\Theta}$. Finally, if $F : X \longrightarrow Y$ and $G : Y \longrightarrow Z$ are smooth, then

$$(G \circ F)^* = F^* \circ G^*$$

(Exercise 5.7.15, [**N4**]).

If $\mathcal{V}$ is a finite dimensional real vector space, then its dual $\mathcal{V}^*$ (real-valued linear maps on $\mathcal{V}$) is often denoted $\mathcal{J}^1(\mathcal{V})$ and called the space of **covariant tensors of rank one** on $\mathcal{V}$. We denote by $\mathcal{J}^2(\mathcal{V})$ the set of all real-valued bilinear maps on $\mathcal{V} \times \mathcal{V}$ and refer to its elements as **covariant tensors of rank two** on $\mathcal{V}$. If $\alpha, \beta \in \mathcal{J}^1(\mathcal{V})$ we define their **tensor product** $\alpha \otimes \beta \in \mathcal{J}^2(\mathcal{V})$ by $(\alpha \otimes \beta)(v, w) = \alpha(v)\beta(w)$. If $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{V}$ and $\{e^1, \ldots, e^n\}$ is its dual basis for $\mathcal{J}^1(\mathcal{V})$, then $\{e^i \otimes e^j \colon i, j = 1, \ldots, n\}$ is a basis for $\mathcal{J}^2(\mathcal{V})$ and every $A \in \mathcal{J}^2(\mathcal{V})$ can be written uniquely as

$$A = A_{ij} e^i \otimes e^j = A\ (e_i,\ e_j)\ e^i \otimes e^j$$

(Lemma 5.11.1, [**N4**]). An $A \in \mathcal{J}^2(\mathcal{V})$ is said to **symmetric** if $A(w,v) = A(v,w)$ for all $v, w \in \mathcal{V}$, **skew-symmetric** if $A(w,v) = -A(v,w)$ for all $v, w \in \mathcal{V}$, **nondegenerate** if $A(v,w) = 0$ for all $v \in \mathcal{V}$ implies $w = 0$ and **positive** (respectively, **negative**) **definite** if $A(v,v) \geq 0$ (respectively, $A(v,v) \leq 0$) for all $v \in \mathcal{V}$ with $A(v,v) = 0$ only for $v = 0$. A nondegenerate, symmetric element of $\mathcal{J}^2(\mathcal{V})$ is called an **inner product** on $\mathcal{V}$ (some sources reserve this terminology for a nondegenerate, symmetric bilinear form that is also positive definite, but the more general notion we have introduced will be important in the context of relativity). The set of all skew-symmetric elements of $\mathcal{J}^2(\mathcal{V})$ is denoted $\Lambda^2(\mathcal{V})$. For all $\alpha, \beta \in \mathcal{J}^1(\mathcal{V})$ we define the wedge product $\alpha \wedge \beta \in \Lambda^2(\mathcal{V})$ by $\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha$. If $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{V}$ and $\{e^1, \ldots, e^n\}$ is its dual basis, then $\{e^i \wedge e^j : 1 \leq i < j \leq n\}$ is a basis for $\Lambda^2(\mathcal{V})$ and each Ω in $\Lambda^2(\mathcal{V})$ can be uniquely written as

$$\Omega = \sum_{i<j} \Omega_{ij} e^i \wedge e^j = \frac{1}{2} \Omega_{ij} e^i \wedge e^j,$$

where $\Omega_{ij} = \Omega(e^i, e^j)$ (Lemma 5.11.2, [**N4**]).

If X is a smooth manifold, then a **covariant tensor field** of **rank two** on X is a map $\boldsymbol{A}$ that assigns to each $p \in X$ an element $\boldsymbol{A}(p) = \boldsymbol{A}_p \in \mathcal{J}^2(T_p(X))$. If (U, φ) is a chart on X with coordinate functions $x^1, \ldots, x^n$, then $\boldsymbol{A}(p) = \boldsymbol{A}_p = \boldsymbol{A}_p \left(\frac{\partial}{\partial x^i}|_p, \frac{\partial}{\partial x^j}|_p \right) dx^i{}_p \otimes dx^j{}_p$. The functions $A_{ij} : U \longrightarrow$ defined by $A_{ij}(p) = \boldsymbol{A}_p \left(\frac{\partial}{\partial x^i}|_p, \frac{\partial}{\partial x^j}|_p \right)$ are the **components** of $\boldsymbol{A}$ relative to (U, φ) and $\boldsymbol{A}$ is said to be continuous, smooth, etc. if its components are continuous, smooth, etc. for all charts in some atlas for X. The set of all smooth $\boldsymbol{A}$ is denoted $\mathcal{J}^2(X)$ and has the structure of a $C^\infty(X)$-module. A $\boldsymbol{g} \in \mathcal{J}^2(X)$ which, at each $p \in X$, is an inner product on $T_p(X)$ is called a **metric** tensor on X. If each $\boldsymbol{g}(p)$ is a positive definite inner product, then $\boldsymbol{g}$ is a **Riemannian** metric; otherwise, $\boldsymbol{g}$ is **semi-Riemannian**. An $\boldsymbol{\Omega} \in \mathcal{J}^2(X)$ which, at each $p \in X$, is skew-symmetric is called a **2-form** on X and the set of all such is denoted $\Lambda^2(X)$. If $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 \in \mathcal{X}^*(X)$, then $\boldsymbol{\Theta}_1 \otimes \boldsymbol{\Theta}_2 \in \mathcal{J}^2(X)$ and $\boldsymbol{\Theta}_1 \wedge \boldsymbol{\Theta}_2 \in \Lambda^2(X)$ are defined pointwise. If $F : X \longrightarrow Y$ is smooth and $\boldsymbol{A} \in \mathcal{J}^2(Y)$, then the **pullback** $F^*\boldsymbol{A} \in \mathcal{J}^2(X)$ is defined, at each $p \in X$, by $(F^*\boldsymbol{A})_p(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{A}_{F(p)}(F_{*p}(\boldsymbol{v}), F_{*p}(\boldsymbol{w}))$ for all $\boldsymbol{v}, \boldsymbol{w} \in T_p(X)$. The pullback of a metric (Riemannian or semi-Riemannian) is a metric of the same type and the pullback, of a 2-form is another 2-form. Restrictions are, as for 1-forms, pullbacks by the inclusion map. An element $\boldsymbol{A}$ of $\mathcal{J}^2(X)$ gives rise to a $C^\infty(X)$-bilinear map $A : \mathcal{X}(X) \times \mathcal{X}(X) \longrightarrow C^\infty(X)$ defined as follows: For all $\boldsymbol{V}, \boldsymbol{W} \in \mathcal{X}(X)$ define $A(\boldsymbol{V}, \boldsymbol{W}) \in C^\infty(X)$ by $A(\boldsymbol{V}, \boldsymbol{W})(p) = \boldsymbol{A}_p(\boldsymbol{V}_p, \boldsymbol{W}_p)$ for all $p \in X$. Conversely, any $C^\infty(X)$-bilinear map $A : \mathcal{X}(X) \times \mathcal{X}(X) \longrightarrow C^\infty(X)$ gives rise to an $\boldsymbol{A} \in \mathcal{J}^2(X)$ whose value at each $p \in X$ is the bilinear map $\boldsymbol{A}_p : T_p(X) \times T_p(X) \longrightarrow$ defined as follows: For $\boldsymbol{v}, \boldsymbol{w} \in T_p(X)$ select $\boldsymbol{V}, \boldsymbol{W} \in \mathcal{X}(X)$ with $\boldsymbol{V}(p) = \boldsymbol{v}$ and $\boldsymbol{W}(p) = \boldsymbol{w}$ (Exercise 5.7.10, [**N4**]) and set $\boldsymbol{A}_p(\boldsymbol{v}_p, \boldsymbol{w}_p) = A(\boldsymbol{V}, \boldsymbol{W})(p)$. This correspondence is one-to-one and preserves the natural algebraic structures (Exercise 5.11.19, [**N4**]), so we may identify the two notions (covariant tensor field of rank two and $C^\infty(X)$-bilinear map of $\mathcal{X}(X) \times \mathcal{X}(X)$ to $C^\infty(X)$). In particular, if $\boldsymbol{\Theta}$ is a 1-form (thought of as a map from $\mathcal{X}(X)$ to $C^\infty(X)$), we define its **exterior derivative** $d\boldsymbol{\Theta} \in \Lambda^2(X)$ (thought of as a bilinear map from $\mathcal{X}(X) \times \mathcal{X}(X)$ to $C^\infty(X)$) by

$$d\boldsymbol{\Theta}(\boldsymbol{V}, \boldsymbol{W}) = \boldsymbol{V}(\boldsymbol{\Theta}\boldsymbol{W}) - \boldsymbol{W}(\boldsymbol{\Theta}\boldsymbol{V}) - \boldsymbol{\Theta}([\boldsymbol{V}, \boldsymbol{W}]).$$

We record some properties of the exterior derivative operator $d : \Lambda^1(X) \longrightarrow \Lambda^2(X)$ (proved on pages 320–322, [**N4**]).

$$d(a\boldsymbol{\Theta}_1 + b\boldsymbol{\Theta}_2) = a d\boldsymbol{\Theta}_1 + b d\boldsymbol{\Theta}_2 \qquad (a, b \in \quad \text{and } \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 \in \mathcal{X}^*(X))$$

$$d(f\boldsymbol{\Theta}) = f\, d\boldsymbol{\Theta} + df \wedge \boldsymbol{\Theta} \qquad (f \in C^\infty(X) \text{ and } \boldsymbol{\Theta} \in \mathcal{X}^*(X))$$

$$d(df) = 0 \qquad (f \in C^\infty(X))$$

$$F^*(d\boldsymbol{\Theta}) = d(F^*\boldsymbol{\Theta}) \qquad (F : X \longrightarrow Y \text{ smooth and } \boldsymbol{\Theta} \in \mathcal{X}^*(Y))$$

and, if (U, φ) is a chart on X with coordinate functions $x^1, \ldots, x^n$,

$$d(\Theta_i dx^i) = d\Theta_i \wedge dx^i = \frac{\partial \Theta_i}{\partial x^j}\, dx^j \wedge dx^i \quad (\boldsymbol{\Theta} \in \mathcal{X}^*(X)).$$

Let X be a smooth manifold, $\mathcal{V}$ a finite dimensional real vector space and $\{e_1, \ldots, e_d\}$ a basis for $\mathcal{V}$. A **$\boldsymbol{\mathcal{V}}$-valued 0-form** on X is a map of X into $\mathcal{V}$. Thus, we may write $\phi = \phi^i e_i$, where the real-valued functions ϕ^i are called the **components** of ϕ relative to $\{e_1, \ldots, e_d\}$. ϕ is said to be continuous, smooth, etc. if its components are continuous, smooth, etc. relative to some (and therefore any) basis for $\mathcal{V}$. A **$\boldsymbol{\mathcal{V}}$-valued 1-form** on X is a map $\boldsymbol{\omega}$ which assigns to every $p \in X$ a linear transformation $\boldsymbol{\omega}(p) = \boldsymbol{\omega}_p$ from $T_p(X)$ to $\mathcal{V}$. Writing $\boldsymbol{\omega} = \boldsymbol{\omega}^i e_i$, where each $\boldsymbol{\omega}^i$ is an ordinary (-valued) 1-form, we say that $\boldsymbol{\omega}$ is continuous, smooth, etc. if its **components** $\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^d$ are continuous, smooth, etc. A **$\boldsymbol{\mathcal{V}}$-valued 2-form** on X is a map $\boldsymbol{\Omega}$ which assigns to each $p \in X$ a bilinear map $\boldsymbol{\Omega}(p) = \boldsymbol{\Omega}_p : T_p(X) \times T_p(X) \longrightarrow \mathcal{V}$ and may be written $\boldsymbol{\Omega} = \boldsymbol{\Omega}^i e_i$, where each $\boldsymbol{\Omega}^i$ is an ordinary 2-form. Again, $\boldsymbol{\Omega}^1, \ldots, \boldsymbol{\Omega}^d$ are the **components** of $\boldsymbol{\Omega}$ relative to $\{e_1, \ldots, e_d\}$ and $\boldsymbol{\Omega}$ is continuous, smooth, etc. if and only if its components are continuous, smooth, etc. Pullbacks of $\mathcal{V}$-valued forms are defined componentwise as are exterior derivatives of $\mathcal{V}$-valued 0-forms and 1-forms. All of this is independent of the choice of basis. Wedge products of vector-valued 1-forms are not defined unless there is given some sort of "multiplication" in the vector space in which the 1-forms take their values. Suppose, somewhat more generally, that $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional real vector spaces and that one is given a bilinear map $\rho : \mathcal{U} \times \mathcal{V} \longrightarrow \mathcal{W}$ (when $\mathcal{U} = \mathcal{V} = \mathcal{W}$ this is a bilinear pairing, or "multiplication" on $\mathcal{V}$). Now, if $\boldsymbol{\omega}$ is a $\boldsymbol{\mathcal{U}}$-valued 1-form on X and $\boldsymbol{\eta}$ is a $\boldsymbol{\mathcal{V}}$-valued 1-form on X, then we define the **$\boldsymbol{\rho}$-wedge product** $\boldsymbol{\omega} \wedge_\rho \boldsymbol{\eta}$ of $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$ by

$$(\boldsymbol{\omega} \wedge_\rho \boldsymbol{\eta})_p(\boldsymbol{v}, \boldsymbol{w}) = \rho\left(\boldsymbol{\omega}_p(\boldsymbol{v}), \boldsymbol{\eta}_p(\boldsymbol{w})\right) - \rho\left(\boldsymbol{\eta}_p(\boldsymbol{v}), \boldsymbol{\omega}_p(\boldsymbol{w})\right)$$

for all $p \in X$ and all $\boldsymbol{v}, \boldsymbol{w} \in T_p(X)$. This then is a $\mathcal{W}$-valued 2-form on X. If $\{u_1, \ldots, u_c\}$ is a basis for $\mathcal{U}$ and $\{v_1, \ldots, v_d\}$ is a basis for $\mathcal{V}$ and if we write $\boldsymbol{\omega} = \boldsymbol{\omega}^i u_i$, and $\boldsymbol{\eta} = \boldsymbol{\eta}^j v_j$, then

$$\boldsymbol{\omega} \wedge_\rho \boldsymbol{\eta} = \sum_{i=1}^{c} \sum_{j=1}^{d} \left(\boldsymbol{\omega}^i \wedge \boldsymbol{\eta}^j\right) \rho\left(u_i, v_j\right)$$

(Lemma 5.11.5, [**N4**]). For example, if $\mathcal{U} = \mathcal{V} = \mathcal{W} =$ (regarded as a 2-dimensional real vector space) and $\rho : \ \times \ \longrightarrow$ is complex multiplication, then, writing $\boldsymbol{\omega} = \boldsymbol{\omega}^1 + \boldsymbol{\omega}^2 \boldsymbol{i}$ and $\boldsymbol{\eta} = \boldsymbol{\eta}^1 + \boldsymbol{\eta}^2 \boldsymbol{i}$, one finds that

$$\begin{aligned} \boldsymbol{\omega} \wedge_\rho \boldsymbol{\eta} &= (\boldsymbol{\omega}^1 + \boldsymbol{\omega}^2 \boldsymbol{i}) \wedge_\rho (\boldsymbol{\eta}^1 + \boldsymbol{\eta}^2 \boldsymbol{i}) \\ &= (\boldsymbol{\omega}^1 \wedge \boldsymbol{\eta}^1 - \boldsymbol{\omega}^2 \wedge \boldsymbol{\eta}^2) + (\boldsymbol{\omega}^1 \wedge \boldsymbol{\eta}^2 + \boldsymbol{\omega}^2 \wedge \boldsymbol{\eta}^1) \boldsymbol{i}. \end{aligned}$$

Thus, $\boldsymbol{\omega} \wedge_\rho \boldsymbol{\eta}$ is obtained by multiplying the forms exactly as one multiplies complex numbers, but with real and imaginary parts "multiplied" by the ordinary wedge. A similar result holds when $\mathcal{U} = \mathcal{V} = \mathcal{W} =$ and ρ is quaternion multiplication. In cases such as these one generally drops the "ρ" and writes simply $\boldsymbol{\omega} \wedge \boldsymbol{\eta}$. Most of the important examples of vector- valued forms arise in the context of Lie groups and principal bundles, to which we now turn.

1.2 Matrix Lie Groups

A Lie group is a differentiable manifold G on which is defined a group structure for which the group multiplication $(x, y) \longrightarrow xy$ is a C^∞ map of $G \times G$ to G (it follows that group inversion $x \longrightarrow x^{-1}$ is a diffeomorphism of G onto itself; see Lemma 5.8.1, [**N4**]). The left and right translation maps $L_g,\ R_g : G \longrightarrow G$ defined by $L_g(x) = gx$ and $R_g(x) = xg$ are diffeomorphisms for each $g \in G$. The Lie groups of most interest to us are the following:

1. The nonzero real numbers, complex numbers, or quaternions with their respective multiplications.

2. The circle S^1 (thought of as the complex numbers of modulus 1 in $=$ ^{2}) with complex multiplication.

3. The 3-sphere S^3 (thought of as the unit quaternions in $=$ ^{4}) with quaternion multiplication.

4. The general linear groups $GL(n, \ ), GL(n, \ )$ and $GL(n, \ )$ with matrix multiplication.

5. Any subgroup H of a Lie group G that is also a submanifold of G is a Lie group so, in particular, all of the classical groups are Lie groups:

$$O(n),\ SO(n),\ U(n),\ SU(n),\ Sp(n).$$

6. Any (finite) product of Lie groups (with the product manifold structure and the direct product group structure) is a Lie group, e.g., $SU(2) \times U(1)$, $SU(3) \times SU(2) \times U(1)$, or any torus $S^1 \times \cdots \times S^1$.

Two Lie groups G_1 and G_2 are **isomorphic** if there is a diffeomorphism of G_1 onto G_2 that is also a group isomorphism, e.g., S^1 is isomorphic to both $U(1)$ and $SO(2)$, while S^3 is isomorphic to $SU(2)$ and $Sp(1)$. Any subgroup of some $GL(n, \ )$ that is also a submanifold is called a **matrix Lie group** and we shall henceforth restrict our attention to these. $GL(n, \ )$ can be identified with a subgroup of $GL(n, \ )$ that is also a submanifold. It is less obvious, but also true, that $GL(n, \ )$ can be identified with a subgroup of $GL(2n, \ )$ that is also a submanifold. To see this observe first that any quaternion $x^0 + x^1\boldsymbol{i} + x^2\boldsymbol{j} + x^3\boldsymbol{k}$ can be written as $z^1 + z^2\boldsymbol{j}$, where $z^1 = x^0 + x^1\boldsymbol{i}$ and $z^2 = x^2 + x^3\boldsymbol{i}$, and so identified with a pair of complex numbers. Thus, any $n \times n$ quaternionic matrix P can be written as $P = A + B\boldsymbol{j}$, where A and B are $n \times n$ complex matrices. Define a map $\phi : \mathcal{GL}\ (n, \ ) \longrightarrow \mathcal{GL}(2n, \ )$ by

$$\phi(P) = \begin{pmatrix} A & B \\ -\bar{B} & \bar{A} \end{pmatrix}.$$

Then ϕ is an algebra isomorphism that also preserves the conjugate transpose (Exercise 1.1.28, [**N4**]). In particular, $P \in Sp(n)$ if and only if $\phi(P) \in U(2n)$. Moreover, a $2n \times 2n$ complex matrix M has the form $\left(\begin{smallmatrix} A & B \\ -\bar{B} & \bar{A} \end{smallmatrix}\right)$ if and only if it satisfies $JMJ^{-1} = \bar{M}$, where $J = \left(\begin{smallmatrix} 0 & id \\ -id & 0 \end{smallmatrix}\right)$ and, if M is unitary, this is equivalent to $M^T JM = J$. Thus, we may identify

$$\begin{aligned} GL(n, \ ) &= \{M \in \mathcal{GL}(2n, \ ) : M \text{ invertible and } JMJ^{-1} = \bar{M}\} \\ Sp(n) &= \{M \in U(2n) : M^T JM \ = J\}. \end{aligned}$$

The **real** and **complex special linear groups** $SL(n, \ )$ and $SL(n, \ )$ are the subgroups of $GL(n, \ )$ and $GL(n, \ )$, respectively, consisting of those elements with determinant 1. Although the noncommutativity of   blocks any obvious notion of a determinant for quaternionic matrices, one can define the **quaternionic special linear group** $SL(n, \ )$ to be the subset of $GL(n, \ )$ consisting of all those elements P such that $\det \phi(P) = 1$. This is, indeed, a subgroup of $GL(n, \mathbb{H})$ (Exercise 1.1.30, [**N4**]). One can show directly that $GL(n, \ ), Sp(n)$ and $SL(n, \ )$ are all submanifolds of $GL(2n, \ )$, but there is also a general result to the effect that any closed subgroup of a complex general linear group is necessarily a submanifold (see [**Howe**] or, for a still more general result, [**Warn**]). We conclude then that all of the classical groups are matrix Lie groups.

A **Lie algebra** is a real vector space $\mathcal{L}$ on which is defined a bilinear operation $[\ ,\ ] : \mathcal{L} \times \mathcal{L} \longrightarrow \mathcal{L}$, called **bracket**, that is skew-symmetric ($[y, x] = -[x, y]$ for all $x, y \in \mathcal{L}$) and satisfies the Jacobi identity ($[x, [y, z]] + [z, [x, y]] + [y, [z, x]] = 0$ for all $x, y, z \in \mathcal{L}$). Two Lie algebras $\mathcal{L}_1$ and $\mathcal{L}_2$ with brackets

$[\ ,\ ]_1$ and $[\ ,\ ]_2$, respectively, are **isomorphic** if there is a linear isomorphism $T : \mathcal{L}_1 \longrightarrow \mathcal{L}_2$ of $\mathcal{L}_1$ onto $\mathcal{L}_2$ that satisfies $T([x,y]_1) = [T(x), T(y)]_2$ for all $x, y \in \mathcal{L}_1$. We record a few important examples of Lie algebras

1. Any finite dimensional real vector space $\mathcal{V}$ with the **trivial bracket** $[\ ,\ ] : \mathcal{V} \times \mathcal{V} \longrightarrow \mathcal{V}$ defined by $[x,y] = 0 \in \mathcal{V}$ for all $x, y \in \mathcal{V}$. Skew-symmetry implies that any 1-dimensional Lie algebra has trivial bracket.

2. The **commutator** of two matrices A and B is defined by $[A,B] = AB - BA$. Since $[B,A] = -[A,B]$ and $[A,[B,C]] + [C,[A,B]] + [B,[C,A]] = 0$, any set of matrices that forms a real vector space and is closed under the formation of commutators is a Lie algebra with bracket given by the commutator. Examples include $\mathcal{GL}(n, \quad), \mathcal{GL}(n, \quad)$ and $\mathcal{GL}(n, \quad)$ as well as the set of all $n \times n$ real, skew-symmetric ($A^T = -A$) matrices, the set of all $n \times n$ complex, skew-Hermitian ($\bar{A}^T = -A$) matrices and the set of all $n \times n$ complex matrices that are both skew-Hermitian and tracefree ($\text{trace}(A) = 0$).

3. The set Im   of pure imaginary complex numbers $a\boldsymbol{i}$ with trivial bracket is a 1-dimensional Lie algebra.

4. The set Im   of pure imaginary quaternions $x = a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}$ with bracket $[x,y] = xy - yx = 2\,\text{Im}\,(xy)$ is a 3-dimensional Lie algebra. This Lie algebra is, in fact, isomorphic to the Lie algebra consisting of $\ ^3$ and the bracket $[x,y] = x \times y$ given by the familiar cross product on $\ ^3$.

Every Lie group G has associated with it a Lie algebra $\mathcal{G}$ (called the **Lie algebra of** G) that can be described in either of the following two equivalent ways:

(i) A vector field $\boldsymbol{V}$ on G is said to be **left-invariant** if $(L_g)_* \circ \boldsymbol{V} = \boldsymbol{V} \circ L_g$ for every $g \in G$, i.e., if $(L_g)_{*h}(\boldsymbol{V}(h)) = \boldsymbol{V}(gh)$ for all $g, h \in G$. Such vector fields are necessarily smooth (Theorem 5.8.2, [**N4**]). The set $\mathcal{G}$ of all left-invariant vector fields on G is a linear subspace of $\mathcal{X}(G)$ and is closed under the formation of Lie brackets (Theorem 5.8.4, [**N4**]). Since the Lie bracket is skew-symmetric and satisfies the Jacobi identity it provides $\mathcal{G}$ with the structure of a Lie algebra.

(ii) If e denotes the identity element in G ($e = id$ for matrix Lie groups), then the map $\boldsymbol{V} \longrightarrow \boldsymbol{V}(e)$ is a linear isomorphism of $\mathcal{G}$ (as defined in (i)) onto $T_e(G)$. Define a bracket $[\ ,]$ on $T_e(G)$ as follows: For $\boldsymbol{v}, \boldsymbol{w} \in T_e(G)$ there exist unique left-invariant vector fields $\boldsymbol{V}, \boldsymbol{W} \in \mathcal{G}$ with $\boldsymbol{V}(e) = \boldsymbol{v}$ and $\boldsymbol{W}(e) = \boldsymbol{w}$. Set $[\boldsymbol{v}, \boldsymbol{w}] = [\boldsymbol{V}, \boldsymbol{W}](e)$. With this, $T_e(G)$ is a Lie algebra isomorphic to $\mathcal{G}$.

For matrix Lie groups G, $T_{id}(G)$ can be identified with a set of matrices (velocity vectors at $t = 0$ to smooth curves $t \longrightarrow (a_{ij}(t))$ with $(a_{ij}(0)) = id$, computed entrywise). One can show (Section 5.8, [**N4**]) that this set of matrices is always closed under the formation of commutators and that the

Lie algebra consisting of $T_{id}(G)$ with commutator bracket is isomorphic to the Lie algebra $\mathcal{G}$ of G as defined in (ii). Thus, the problem of finding the Lie algebra $\mathcal{G}$ of a matrix Lie group G reduces to identifying the set of matrices that arise as velocity vectors to smooth curves in G through id (and defining the bracket on this set of matrices to be the commutator). We record a number of important examples (for the proofs see pages 278–284, [**N4**]).

1. For = , , or , the Lie algebra of $GL(n, \ )$ is the set $\mathcal{GL}(n, \ )$ of all $n \times n$ matrices with entries in .

2. The Lie algebra $o(n)$ of the orthogonal group $O(n)$ is the set of all $n \times n$ real, skew-symmetric matrices.

3. The Lie algebra $so\,(n)$ of the special orthogonal group $SO(n)$ coincides with the Lie algebra $o(n)$ of $O(n)$ (because $SO(n)$ is just the connected component of $O(n)$ containing id so that $T_{id}(SO(n)) = T_{id}(O(n))$).

4. The Lie algebra $u(n)$ of the unitary group $U(n)$ is the set of all $n \times n$ complex, skew-Hermitian matrices.

5. The Lie algebra $su(n)$ of the special unitary group $SU(n)$ is the set of all $n \times n$ complex matrices that are skew-Hermitian and tracefree.

6. Identifying $Sp(n)$ with the matrix Lie group $\{U \in U(2n) : U^T J U = J\}$, the Lie algebra $sp(n)$ of $Sp(n)$ is given by

$$\begin{aligned} sp(n) &= \{M \in \mathcal{GL}(2n, \ ) : \bar{M}^T = -M \text{ and } JM + M^T J = 0\} \\ &= \left\{ \begin{pmatrix} A & B \\ -\bar{B} & \bar{A} \end{pmatrix} \in \mathcal{GL}\ (2n, \ ) : A \in u(n), B^T = B \right\}. \end{aligned}$$

Isomorphic Lie groups have isomorphic Lie algebras. For example, since $U(1)$ and $SO(2)$ are isomorphic, so are $u(1)$ and $so(2)$ and, since $u(1)$ is the algebra of 1×1 skew-Hermitian matrices, both can be naturally identified with the Lie algebra Im . Similarly, $SU(2)$ and $Sp(1)$ are isomorphic and therefore so are their Lie algebras $su(2)$ and $sp(1)$. But $sp(1)$ is naturally identfied with the Lie algebra Im of pure imaginary quaternions which, in turn, is isomorphic to ^{3} with the cross product as bracket. However, non-isomorphic Lie groups can have isomorphic Lie algebras, e.g., $O(n)$ and $SO(n)$ in #3 above. A less trivial example consists of $SU(2)$ and $SO(3)$. These are certainly not isomorphic as Lie groups. Indeed, they are not even homeomorphic since $SU(2) \cong S^3$ (Theorem 1.1.4, [**N4**]) and S^3 is simply connected (page 119, [**N4**]), whereas $SO(3) \cong \ ^3$ (page 399, [**N4**]) and $\pi_1(\ ^3) \cong \mathbb{Z}_2$ (Theorem 2.4.5, [**N4**]). To see that the Lie algebras $su(2)$ and $so(3)$ are isomorphic we introduce the notion of a Lie group's "structure constants." Let $\mathcal{G}$ be an arbitrary Lie algebra. Select a basis $\{e_1, \ldots, e_n\}$ for $\mathcal{G}$. For all $i, j = 1, \ldots, n$, $[e_i, e_j]$ is in $\mathcal{G}$ so there exist unique constants C_{ij}^k, $k = 1, \ldots, n$, such that $[e_i, e_j] = C_{ij}^k e_k$. The constants C_{ij}^k, $i, j, k = 1, \ldots, n$ are called the **structure constants** of $\mathcal{G}$ relative to the basis $\{e_1, \ldots, e_n\}$. Two

Lie algebras are clearly isomorphic if and only if there exist bases relative to which the structure constants are the same. Now, an obvious basis for $so(3)$ (3×3 real, skew-symmetric matrices) consists of

$$\tau_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and a simple calculation shows that $[\tau_i, \tau_j] = \sum_{k=1}^{3} \epsilon_{ijk} \tau_k$, where ϵ_{ijk} is the Levi-Civita symbol (1 if ijk is an even permutation of 123, -1 if ijk is an odd permutation of 123, and 0 otherwise). The usual basis for $su(2)$ consists of $T_k = -\frac{1}{2} \boldsymbol{i} \sigma_k$, where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\boldsymbol{i} \\ \boldsymbol{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the **Pauli spin matrices** (page 396, [**N4**]) and here again one quickly verifies that $[T_i,\ T_j] = \sum_{k=1}^{3} \epsilon_{ijk} T_k$. Thus, $so(3)$ and $su(2)$ are isomorphic.

For any $A \in \mathcal{GL}(n, \ \ )$ we define $\exp(A) = e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$. The series converges absolutely and uniformly on every bounded region in $\mathcal{GL}(n, \ \ ) = \ ^{n^2} = \ ^{2n^2}$. Every $\exp(A)$ is invertible (because $\det(e^A) = e^{\mathrm{trace}(A)}$) so exp is a C^∞ map of $\mathcal{GL}(n, \ \ )$ to $GL(n, \ \ )$. If G is any matrix group, then its Lie algebra $\mathcal{G}$ can be identified with a subalgebra of $\mathcal{GL}(n, \ \ )$ for some n and the restriction of exp to $\mathcal{G}$ maps into G (see Theorem 5.8.6, [**N4**], for the cases of interest here). The map $\exp : \mathcal{G} \longrightarrow G$ has a derivative at $0 \in \mathcal{G}$ and, if $T_0(\mathcal{G})$ is identified with $\mathcal{G}$ by the canonical isomorphism, $\exp_{*o} : \mathcal{G} \longrightarrow \mathcal{G}$ is the identity map (Lemma 5.8.5 (5), [**N4**]). It follows from the Inverse Function Theorem that exp is a diffeomorphism of some neighborhood of $0 \in \mathcal{G}$ onto a neighborhood of $\exp(0) = id$ in G, i.e., on some neighborhood of id in G, $\exp^{-1}$ is a chart for G.

A 1-form $\boldsymbol{\Theta}$ on a Lie group G (real, or vector-valued) is said to be **left invariant** if $(L_g)^* \boldsymbol{\Theta} = \boldsymbol{\Theta}$ for all $g \in G$, i.e., if, for all $g, h \in G$, $\boldsymbol{\Theta}(h) = (L_g)^*(\boldsymbol{\Theta}(gh))$, or, equivalently, $\boldsymbol{\Theta}(gh) = (L_{g^{-1}})^*(\boldsymbol{\Theta}(h))$. This is the case if and only if $\boldsymbol{\Theta}(g) = (L_{g^{-1}})^*(\boldsymbol{\Theta}(id))$ for every $g \in G$. Any such 1-form is necessarily smooth (page 285, [**N4**]) and is uniquely determined by its value at the identity. In particular, any covector at id in G uniquely determines a left invariant 1-form taking that value at id. A particularly important example of a left invariant 1-form on G is the **Cartan** (**canonical**) **1-form** which takes values in the vector space $\mathcal{G}$ and is defined as follows: For each $g \in G, \boldsymbol{\Theta}(g) = \boldsymbol{\Theta}_g : T_g(G) \longrightarrow \mathcal{G} = T_{id}(G)$ is given by

$$\boldsymbol{\Theta}(g)(\boldsymbol{v}) = \boldsymbol{\Theta}_g(\boldsymbol{v}) = (L_{g^{-1}})_{*g}(\boldsymbol{v}).$$

Equivalently, $\boldsymbol{\Theta}$ is uniquely determined by the requirement that $\boldsymbol{\Theta}_g(\boldsymbol{A}(g)) = \boldsymbol{A}(id)$ for every left invariant vector field $\boldsymbol{A}$ on G. Let $\{e_1, \ldots, e_n\}$ be a basis for $\mathcal{G}$ and let $\{\boldsymbol{\Theta}^1, \ldots, \boldsymbol{\Theta}^n\}$ be the unique left invariant real-valued 1-forms on G for which $\{\boldsymbol{\Theta}^1(id), \ldots, \boldsymbol{\Theta}^n(id)\}$ is the basis dual to $\{e_1, \ldots, e_n\}$. Then the Cartan 1-form $\boldsymbol{\Theta}$ is given by

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}^1 e_1 + \cdots + \boldsymbol{\Theta}^n e_n$$

(Lemma 5.9.1, [**N4**]). The **Maurer-Cartan equations** for $\boldsymbol{\Theta}$ assert that

$$d\boldsymbol{\Theta}^k = -\frac{1}{2} C^k_{ij} \boldsymbol{\Theta}^i \wedge \boldsymbol{\Theta}^j, \quad k = 1, \ldots, n,$$

where C^k_{ij} are the structure constants for $\mathcal{G}$ relative to $\{e_1, \ldots, e_n\}$ ((5.11.3), [**N4**]). If H is a subgroup of G that is also a submanifold and $\iota : H \hookrightarrow G$ is the inclusion map, then the Cartan 1-form $\boldsymbol{\Theta}_H$ of H is the restriction to H of the Cartan 1-form $\boldsymbol{\Theta}_G$ of G, i.e.,

$$\boldsymbol{\Theta}_H = \iota^* \boldsymbol{\Theta}_G$$

((5.9.6), [**N4**]). Now we exhibit a number of specific examples of Cartan 1-forms for the groups of particular interest to us (details are available on pages 292–296, [**N4**]).

1. $G = GL(n, \ )$, $\mathcal{G} = \mathcal{GL}(n, \ )$: $GL(n, \ )$ is an open submanifold of $\ ^{n^2}$ with standard coordinate (entry) functions x^{ij}, $i, j = 1, \ldots, n$. $\left\{ \frac{\partial}{\partial x^{ij}}\big|_{id} \right\}_{i,j=1,\ldots,n}$ is a basis for $\mathcal{GL}(n, \ )$. The corresponding dual basis is $\left\{ dx^{ij}(id) \right\}_{i,j=1,\ldots,n}$. The -valued 1-forms $\boldsymbol{\Theta}^{ij}$ defined by $\boldsymbol{\Theta}^{ij}(g) = \sum_{k=1}^n x^{ik}\left(g^{-1}\right) dx^{kj}(g)$ for all $g \in GL(n, \ )$ are left invariant and satisfy $\boldsymbol{\Theta}^{ij}(id) = dx^{ij}(id)$ so $\boldsymbol{\Theta}$ is given by

$$\boldsymbol{\Theta}(g) = \boldsymbol{\Theta}^{ij}(g) \frac{\partial}{\partial x^{ij}}\bigg|_{id} = \left(\sum_{k=1}^n x^{ik}(g^{-1}) dx^{kj}(g) \right) \frac{\partial}{\partial x^{ij}}\bigg|_{id}.$$

For calculations it is most convenient to identify $\boldsymbol{\Theta}(g)$ with the matrix of components $\left(\sum_{k=1}^n x^{ik}(g^{-1}) dx^{kj}(g) \right)_{i,j=1,\ldots,n}$. Note that this is the formal matrix product of g^{-1} and

$$dx(g) = \begin{pmatrix} dx^{11}(g) & \ldots & dx^{1n}(g) \\ \vdots & & \vdots \\ dx^{n1}(g) & \ldots & dx^{nn}(g) \end{pmatrix}$$

so one can write

$$\boldsymbol{\Theta}(g) = g^{-1} dx(g).$$

To evaluate $\boldsymbol{\Theta}(g)$ at any $\boldsymbol{v} \in T_g(G)$ compute

$$dx(g)(\boldsymbol{v}) = \left(dx^{ij}\,(g)\,(\boldsymbol{v})\right)_{i,j=1,\ldots,n} = (v^{ij})_{i,j=1,\ldots,n},$$

where $\boldsymbol{v} = v^{ij}\frac{\partial}{\partial x^{ij}}|_g$, and form the product $g^{-1}(v^{ij})_{i,j=1,\ldots,n}$, the result being a matrix in $\mathcal{GL}(n,\ \ )$. Taking this computational ploy one step further we will often identify $\boldsymbol{\Theta}$ with a matrix of -valued 1-forms. We illustrate with $GL(2,\ \ )$: For each $g = \left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in GL(2,\ \ )$, $g^{-1} = (\alpha\delta - \beta\gamma)^{-1}\left(\begin{smallmatrix}\delta & -\beta\\ -\gamma & \alpha\end{smallmatrix}\right)$ and we compute the product

$$\begin{aligned} g^{-1}dx(g) &= (\alpha\delta-\beta\gamma)^{-1}\begin{pmatrix}\delta & -\beta\\ -\gamma & \alpha\end{pmatrix}\begin{pmatrix}dx^{11}(g) & dx^{12}(g)\\ dx^{21}(g) & dx^{22}(g)\end{pmatrix}\\ &= (\alpha\delta-\beta\gamma)^{-1}\begin{pmatrix}\delta dx^{11}(g) - \beta dx^{21}(g) & \delta dx^{12}(g) - \beta dx^{22}(g)\\ -\gamma dx^{11}(g) + \alpha dx^{21}(g) & -\gamma dx^{12}(g) + \alpha dx^{22}(g)\end{pmatrix}.\end{aligned}$$

Since $\alpha = x^{11}(g), \beta = x^{12}(g), \gamma = x^{21}(g)$ and $\delta = x^{22}(g)$ we may regard this as the value at g of the matrix of ordinary 1-forms given by

$$\begin{aligned}&\boldsymbol{\Theta}_{GL(2,\ \ )} =\\ &(x^{11}x^{22} - x^{12}x^{21})^{-1}\begin{pmatrix}x^{22}dx^{11} - x^{12}dx^{21} & x^{22}dx^{12} - x^{12}dx^{22}\\ -x^{21}dx^{11} + x^{11}dx^{21} & -x^{21}dx^{12} + x^{11}dx^{22}\end{pmatrix}\end{aligned}$$

2. $G = GL(n,\ \ )$, $\mathcal{G} = \mathcal{GL}(n,\ \ )$: One proceeds just as in #1 above, identifying $G = GL(n,\ \ )$ with an open submanifold of $\ ^{n^2} = \ ^{2n^2}$ with standard coordinates $\{x^{11}, y^{11}, \ldots, x^{nn}, y^{nn}\}$, where $z^{ij} = x^{ij} + \boldsymbol{i}y^{ij}$. Again, one finds that, for any $g \in G, \boldsymbol{\Theta}(g)$ can be identified with $g^{-1}dz(g)$, where dz is the matrix of complex-valued 1-forms

$$dz = \begin{pmatrix}dz^{11} & \ldots & dz^{1n}\\ \vdots & & \vdots\\ dz^{n1} & \ldots & dz^{nn}\end{pmatrix} = \begin{pmatrix}dx^{11} + \boldsymbol{i}\,dy^{11} & \ldots & dx^{1n} + \boldsymbol{i}\,dy^{1n}\\ \vdots & & \vdots\\ dx^{n1} + \boldsymbol{i}\,dy^{n1} & \ldots & dx^{nn} + \boldsymbol{i}\,dy^{nn}\end{pmatrix}$$

(complex-valued 1-forms are just vector-valued 1-forms with values in $\ = \ ^2$ and we have used the basis $\{1, \boldsymbol{i}\}$ for over). Thus, if $\boldsymbol{v} \in T_g(G)$ (thought of as an $n \times n$ matrix of complex numbers obtained by differentiating entrywise a smooth curve in $GL(n,\ \ )$ through g), then dz^{ij} picks out its ij-entry v^{ij} and $\boldsymbol{\Theta}(g)(\boldsymbol{v}) = g^{-1}(v^{ij})_{i,j=1,\ldots,n}$. For $n = 2$ one can write $\boldsymbol{\Theta}$ as a matrix of complex-valued 1-forms just as we did for $GL(2,\ \ )$ in #1:

$$\begin{aligned}&\boldsymbol{\Theta}_{GL(2,\mathbb{C})} =\\ &(z^{11}z^{22} - z^{12}z^{21})^{-1}\begin{pmatrix}z^{22}dz^{11} - z^{12}dz^{21} & z^{22}dz^{12} - z^{12}dz^{22}\\ -z^{21}dz^{11} + z^{11}dz^{21} & -z^{21}dz^{12} + z^{11}dz^{22}\end{pmatrix}\end{aligned}$$

3. $G = GL(n,\ ), \mathcal{G} = \mathcal{GL}(n,\ )$: The result is the same as for $GL(n,\ )$ and $GL(n,\ )$ in #1 and #2: For every $g \in GL(n,\ )$,

$$\boldsymbol{\Theta}(g) = g^{-1}dq(g),$$

where

$$dq = \begin{pmatrix} dq^{11} & \dots & dq^{1n} \\ \vdots & & \vdots \\ dq^{n1} & \dots & dq^{nn} \end{pmatrix}$$

and $dq^{ij} = dx^{ij} + \boldsymbol{i}dy^{ij} + \boldsymbol{j}du^{ij} + \boldsymbol{k}dv^{ij},\ i,j = 1,\dots,n$. This time, however, there is no simple formula for g^{-1} when $n = 2$ since we lack a determinant function for quaternionic matrices. However, when $n = 1$ and we identify a 1×1 quaternionic matrix $g = (q)$ with its sole entry $q, GL(1,\ )$ is just the group $\ - \{0\}$ of nonzero quaternions and $g^{-1}dq$ is just the quaternion product

$$g^{-1}dq = q^{-1}dq = \frac{1}{|q|^2}\bar{q}dq.$$

Identifying tangent vectors $\boldsymbol{v} \in T_q(\ - \{0\})$ with quaternions $v \in\ $ via the canonical isomorphism, $q^{-1}dq(\boldsymbol{v}) = \frac{1}{|q|^2}\bar{q}v$.

4. $G = SO(2) \cong U(1), \mathcal{G} = so(2) \cong u(1) \cong \mathrm{Im}\ $: First regard G as $SO(2) \subseteq GL(2,\ )$. Then $\boldsymbol{\Theta}_{SO(2)}$ is the restriction of $\boldsymbol{\Theta}_{GL(2,\ )}$ from #1 to $SO(2)$. One finds (page 294, [**N4**]) that

$$\begin{aligned}\boldsymbol{\Theta}_{SO(2)} &= \begin{pmatrix} 0 & x^{22}dx^{12} - x^{12}dx^{22} \\ -x^{22}dx^{12} + x^{12}dx^{22} & 0 \end{pmatrix} \\ &= (-x^{22}dx^{12} + x^{12}dx^{22}) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.\end{aligned}$$

Now, the standard identification of $SO(2)$ and $U(1)$ is $\left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right) \longrightarrow (e^{\boldsymbol{i}\theta})$ and this induces (by differentiation at id) the identification $\theta_0\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \longrightarrow \theta_0\boldsymbol{i}$ of $so(2)$ and $u(1) \cong \mathrm{Im}\ $. Thus,

$$\boldsymbol{\Theta}_{U(1)} = (-x^{22}dx^{12} + x^{12}dx^{22})\boldsymbol{i}$$

on $U(1) = \{(x^{12}, x^{22}) \in\ ^2 : (x^{12})^2 + (x^{22})^2 = 1\}$.

5. $G = SU(2) \cong Sp(1), \mathcal{G} = su(2) \cong sp(1) \cong \mathrm{Im}\ $: First regard G as $SU(2) \subseteq GL(2,\ )$. Then $\boldsymbol{\Theta}_{SU(2)}$ is the restriction of $\boldsymbol{\Theta}_{GL(2,\mathbb{C})}$ from #2 to $SU(2)$. One finds (Exercise 5.9.5, [**N4**]) that

$$\boldsymbol{\Theta}_{SU(2)} = \begin{pmatrix} \bar{z}^{11}dz^{11} + z^{12}d\bar{z}^{12} & \bar{z}^{11}dz^{12} - z^{12}d\bar{z}^{11} \\ -z^{11}d\bar{z}^{12} + \bar{z}^{12}dz^{11} & z^{11}d\bar{z}^{11} + \bar{z}^{12}dz^{12} \end{pmatrix}$$

on $SU(2)$. Now think of G as $Sp(1)$ (identified with the unit quaternions). Then $\boldsymbol{\Theta}_{Sp(1)}$ is the restriction of $\boldsymbol{\Theta}_{GL(1,\mathbb{H})}$ in #3 and (because $|q| = 1$ on $Sp(1)$) this is

$$\boldsymbol{\Theta}_{Sp(1)} = q^{-1}dq = \bar{q}\,dq$$

on $Sp(1)$. Notice that, by writing $q = z^{11} + z^{12}\boldsymbol{j}$, where $z^{11} = x + y\boldsymbol{i}$ and $z^{12} = u + v\boldsymbol{i}$, we have

$$\begin{aligned}\boldsymbol{\Theta}_{Sp(1)} &= (\bar{z}^{11} - \bar{z}^{12}\boldsymbol{j})(dz^{11} + dz^{12}\boldsymbol{j}) \\ &= (\bar{z}^{11}dz^{11} + z^{12}d\bar{z}^{12}) + (\bar{z}^{11}dz^{12} - z^{12}d\bar{z}^{11})\boldsymbol{j}\end{aligned}$$

on $Sp(1)$. The standard identification of $Sp(1)$ with $SU(2)$ is

$$A + B\boldsymbol{j} \longrightarrow \begin{pmatrix} A & B \\ -\bar{B} & \bar{A} \end{pmatrix}$$

and this carries $\boldsymbol{\Theta}_{Sp(1)}$ onto $\boldsymbol{\Theta}_{SU(2)}$.

Now, let G be a matrix Lie group, $\mathcal{V}$ a finite dimensional vector space and $GL(\mathcal{V})$ the group of nonsingular linear transformations on $\mathcal{V}$. A **representation of G on $\mathcal{V}$** is a group homomorphism of G into $GL(\mathcal{V})$. Choosing a basis for $\mathcal{V}$ one can regard any representation of G on $\mathcal{V}$ as a map of G into some general linear group and we say that the representation is continuous, smooth, etc., if this corresponding matrix-valued map is continuous, smooth, etc. Observe that this definition clearly does not depend on the choice of basis for $\mathcal{V}$. We will occasionally relax our terminology a bit and refer to a homomorphism of G into a general linear group as a representation of G. An example of considerable importance arises as follows. For each $g \in G$ define a map

$$Ad_g : G \longrightarrow G$$

by $Ad_g(h) = ghg^{-1}$ for all $h \in G$. Thus, $Ad_g = L_g \circ R_{g^{-1}} = R_{g^{-1}} \circ L_g$ is a diffeomorphism of G onto G. Furthermore, $Ad_g(id) = id$ so the derivative of Ad_g at the identity carries $\mathcal{G}$ isomorphically onto $\mathcal{G}$. We denote this map

$$ad_g : \mathcal{G} \longrightarrow \mathcal{G}.$$

Thus, $ad_g = (Ad_g)_{*id} = (L_g)_{*g^{-1}} \circ (R_{g^{-1}})_{*id} = (R_{g^{-1}})_{*g} \circ (L_g)_{*id}$ and one can show (Lemma 5.8.7, [**N4**]) that

$$ad_g(A) = g\,Ag^{-1}$$

for any $A \in \mathcal{G}$. The assignment $g \longrightarrow ad_g$ is a homomorphism of G into $GL(\mathcal{G})$ and is called the **adjoint representation** of G on $\mathcal{G}$. The significance of the adjoint representation may not be immediately apparent, but will become clear when we discuss connections on principal bundles. On an even more concrete level, it is shown in Appendix A to [**N4**] that $SO(3)$ can be identified with $SU(2)/\pm 1$ and that, when $su(2)$ is identified with $\ ^3$, the adjoint representation of $SU(2)$ on $su\,(2)$ is just the natural representation of $SO(3)$ on $\ ^3$ by rotation.

We outline a few general facts about Killing forms, although we will require only special cases (all that we say is proved in [**Helg**]). If G is a matrix Lie group with Lie algebra $\mathcal{G}$ and if A and B are two fixed elements of $\mathcal{G}$, then we can define a linear transformation $K_{AB} : \mathcal{G} \longrightarrow \mathcal{G}$ by $K_{AB}(X) = [A, [B, X]]$ for all $X \in \mathcal{G}$. Then the trace of this linear transformation is a real number (because $\mathcal{G}$ is a real vector space) and the map $K : \mathcal{G} \longrightarrow$ defined by $K(A, B) = \text{trace}\,(K_{AB})$ is a symmetric, bilinear form on $\mathcal{G}$ called the **Killing form** of $\mathcal{G}$. K is $\boldsymbol{ad}(\boldsymbol{G})$**-invariant** in the sense that, for any $g \in G$, $K(ad_g(A), ad_g(B)) = K(A, B)$ for all $A, B \in \mathcal{G}$. The Lie group G (or its Lie algebra $\mathcal{G}$) is said to be **semisimple** if the Killing form K is nondegenerate.

Remark: There is an algebraic characterization of semisimplicity due to Cartan. An **ideal** in a Lie algebra $\mathcal{G}$ is a linear subspace $\mathcal{H}$ of $\mathcal{G}$ with the property that $[A, B] \in \mathcal{H}$ whenever $A \in \mathcal{H}$ and $B \in \mathcal{G}$. In particular, an ideal is itself a Lie algebra under the same bracket operation. A **proper ideal** is one that is neither the zero subspace nor the entire Lie algebra. Then $\mathcal{G}$ is semisimple if and only if it can be written as a direct sum of ideals, each of which (as a Lie algebra) has no proper ideals.

If G is connected and semisimple, then, by a theorem of Weyl, the Killing form K is negative definite if and only if G is compact. In this case, one obtains a positive definite, $ad(G)$-invariant inner product $\langle\, ,\rangle$ on $\mathcal{G}$ by setting $\langle A, B\rangle = -K(A, B)$. A Riemannian metric $\boldsymbol{g}$ on G is then obtained by left translation, i.e., by defining, for each $a \in G$ and all $\boldsymbol{v}, \boldsymbol{w} \in T_a(G)$, $\boldsymbol{g}_a(\boldsymbol{v}, \boldsymbol{w}) = \langle (L_{a^{-1}})_{*a}(\boldsymbol{v}), (L_{a^{-1}})_{*a}(\boldsymbol{w})\rangle$. This metric is **left invariant**, i.e., $L_a^*\boldsymbol{g} = \boldsymbol{g}$ for all $a \in G$, because $(L_a^*\boldsymbol{g})_b(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{g}_{ab}((L_a)_{*b}(\boldsymbol{v}), (L_a)_{*b}(\boldsymbol{w})) = \langle (L_{(ab)^{-1}})_{*ab}((L_a)_{*b}(\boldsymbol{v})), (L_{(ab)^{-1}})_{*ab}((L_a)_{*b}\,(\boldsymbol{w}))\rangle = \langle (L_{b^{-1}})_{*b}(\boldsymbol{v}), (L_{b^{-1}})_{*b}(\boldsymbol{w})\rangle = \boldsymbol{g}_b(\boldsymbol{v}, \boldsymbol{w})$. It is also **right invariant**, i.e., $R_a^*\,\boldsymbol{g} = \boldsymbol{g}$ for all $a \in G$, because

$$
\begin{aligned}
(R_a^*\boldsymbol{g})_b(\boldsymbol{v}, \boldsymbol{w}) &= \boldsymbol{g}_{ba}((R_a)_{*b}(\boldsymbol{v}), (R_a)_{*b}(\boldsymbol{w})) \\
&= \langle (L_{(ba)^{-1}})_{*ba}((R_a)_{*b}(\boldsymbol{v})), (L_{(ba)^{-1}})_{*ba}((R_a)_{*b}(\boldsymbol{w}))\rangle \\
&= \langle (L_{a^{-1}b^{-1}} \circ R_a)_{*b}(\boldsymbol{v}), (L_{a^{-1}b^{-1}} \circ R_a)_{*b}(\boldsymbol{w})\rangle \\
&= \langle (Ad_{a^{-1}} \circ L_{b^{-1}})_{*b}(\boldsymbol{v}), (Ad_{a^{-1}} \circ L_{b^{-1}})_{*b}(\boldsymbol{w})\rangle \\
&= \langle ad_{a^{-1}}((L_{b^{-1}})_{*b}(\boldsymbol{v})), ad_{a^{-1}}((L_{b^{-1}})_{*b}(\boldsymbol{w}))\rangle \\
&= \langle (L_{b^{-1}})_{*b}(\boldsymbol{v}), (L_{b^{-1}})_{*b}(\boldsymbol{w})\rangle \\
&= \boldsymbol{g}_b(\boldsymbol{v}, \boldsymbol{w}).
\end{aligned}
$$

A metric on a Lie group that is both left invariant and right invariant is said to be **bi-invariant** and we have just shown that the Killing form on a compact, connected, semisimple Lie group G gives rise to a bi-invariant Riemannian metric on G. All of this applies, in particular, to $G = SU(2)$ which is surely compact and connected and also happens to be semisimple ($su\ (2) \cong so\ (3) \cong sp\ (1) \cong \mathrm{Im}$ is the only 3-dimensional, semisimple Lie algebra). Routine, but tedious calculations show that, if $su\ (2)$ is regarded as the algebra of 2×2 complex, skew-Hermitian, tracefree matrices, then $\langle A, B \rangle = -K(A,B) = -\mathrm{trace}\,(AB)$. On the other hand, with the standard identification

$$\begin{pmatrix} a_1\boldsymbol{i} & a_2 + a_3\boldsymbol{i} \\ -a_2 + a_3\boldsymbol{i} & -a_1\boldsymbol{i} \end{pmatrix} \longrightarrow a_1\boldsymbol{i} + a_2\boldsymbol{j} + a_3\boldsymbol{k} \longrightarrow (a_1, a_2, a_3)$$

of $su(2)$ with $\mathrm{Im} \cong {}^3$, $\langle\ ,\ \rangle$ is just twice the usual inner product on ^{3}.

Next suppose that G is a Lie group and P is a differentiable manifold. A **smooth right action** of G on P is a C^∞ map $\sigma : P \times G \longrightarrow P$ which satisfies

1. $\sigma(p, e) = p$ for all $p \in P$ (e is the identity element in G), and
2. $\sigma(p,\, g_1 g_2) = \sigma(\sigma(p,\, g_1), g_2)$ for all $g_1,\, g_2 \in G$ and all $p \in P$.

One generally writes $\sigma(p, g) = p \cdot g$ and thinks of g as "acting on" p to produce $p \cdot g \in P$. Then the defining properties assume the form

1. $p \cdot e = p$ for all $p \in P$, and
2. $p \cdot (g_1 g_2) = (p \cdot g_1) \cdot g_2$ for all $g_1, g_2 \in G$ and all $p \in P$.

For each fixed $g \in G$ we define $\sigma_g : P \longrightarrow P$ by $\sigma_g(p) = p \cdot g$. Then σ_g is a diffeomorphism of P onto P with inverse $\sigma_{g^{-1}}$. Similarly, a **smooth left action** of G on P is a smooth map $\rho : G \times P \longrightarrow P, \rho(g,p) = g \cdot p$, that satisfies

1. $e \cdot p = p$ for all $p \in P$, and
2. $(g_1 g_2) \cdot p = g_1 \cdot (g_2 \cdot p)$ for all $g_1, g_2 \in G$ and all $p \in P$.

The maps $\rho_g : P \longrightarrow P$ defined by $\rho_g(p) = g \cdot p$ are all diffeomorphisms.

Remark: If $(g,p) \longrightarrow g \cdot p$ is a left action, then $(p,g) \longrightarrow p \odot g = g^{-1} \cdot p$ is a right action and, if $(p,g) \longrightarrow p \cdot g$ is a right action, then $(g,p) \longrightarrow g \odot p = p \cdot g^{-1}$ is a left action. Which sort of action one chooses to deal with is generally a matter of personal taste, although some actions appear more natural in one guise than another. For example, any smooth representation $\rho : G \longrightarrow GL(\mathcal{V})$ of G on the vector space $\mathcal{V}$ gives rise to an action of G on the manifold $\mathcal{V}$ ($(g,v) \longrightarrow (\rho(g))(v)$) which, because of the way matrix multiplication is defined, it is more natural to view as acting on the left. We will formulate the remaining definitions in terms of right actions and leave the obvious modifications required for left actions to your imagination.

A right action σ of G on P is said to be **effective** if $p \cdot g = p$ for all $p \in P$ implies $g = e$, i.e., if $\sigma_g = id_P$ if and only if $g = e$. The action is said to be **free** if $p \cdot g = p$ for some $p \in P$ implies $g = e$, i.e., if σ_g has a fixed point if and only if $g = e$. A free action is effective, but the converse is false in general. The action is said to be **transitive** if, given any two points p_1 and p_2 in P, there exists a $g \in G$ such that $p_2 = p_1 \cdot g$. Given any $p \in P$ we define the **orbit** of p under σ to be the subset $p \cdot G = \{p \cdot g : g \in G\}$ of P and its **isotropy subgroup** (or **stabilizer**) to be the subset $G_p = \{g \in G : p \cdot g = p\}$ of G. G_p is a closed subgroup of G for any $p \in P$ (Exercise 1.6.10, [**N4**]). An action is free if every isotropy subgroup is trivial and transitive if there is precisely one orbit. We pause now to describe a number of important examples.

1. Any Lie group G acts on itself by right multiplication. That is, defining $\sigma : G \times G \longrightarrow G$ by $\sigma(p, g) = pg$ for all $p, g \in G$ gives a smooth right action of G on G that is clearly free ($pg = p$ implies $g = e$) and transitive ($p_2 = p_1(p_1^{-1}p_2)$).

2. Any Lie group G acts on itself by conjugation. That is, defining $\sigma : G \times G \longrightarrow G$ by $\sigma(p, g) = g^{-1}pg$ for all $p, g \in G$ gives a smooth right action of G on G that is neither free nor transitive (unless $G = \{e\}$). Note that $G_e = G$ and, in general, G_g is the centralizer of $g \in G$. The action is effective if and only if the center of G is trivial. The orbits are the conjugacy classes of G. The corresponding left action is $\rho(g, p) = g \cdot p = gpg^{-1} = Ad_g(p)$.

3. $GL(n,\ )$ acts on $\ ^n$ (on the left) by $A \cdot v = Av$, where elements of $\ ^n$ are written as column matrices and the product on the right is matrix multiplication. This action is effective, but neither free nor transitive because $GL(n,\ )$ fixes $0 \in\ ^n$. The same action on $\ ^n - \{0\}$ is transitive.

4. $O(n)$ and $SO(n)$ both act transitively on S^{n-1} by $A \cdot v = Av$ (pages 90–91, [**N4**]). The isotropy subgroup of the north pole $e_n = (0, \ldots, 0, 1) \in S^{n-1}$ under the action of $O(n)$ (respectively, $SO(n)$) on S^{n-1} is isomorphic to $O(n-1)$ (respectively, $SO(n-1)$) (page 91, [**N4**]). In the same way (pages 91–92, [**N4**]), $U(n)$ and $SU(n)$ act transitively on S^{2n-1} and $Sp(n)$ acts transitively on S^{4n-1}. The isotropy subgroups at the north pole are isomorphic to $U(n-1)$, $SU(n-1)$ and $Sp(n-1)$, respectively.

Remark: There is a general result (Theorem 1.6.6, [**N4**]) which asserts that, if G is compact, $(g, p) \longrightarrow g \cdot p$ is a transitive left action of G on P and G_{p_0} is the isotropy subgroup of some fixed $p_0 \in P$ under this action, then the quotient group G/G_{p_0} is homeomorphic to P. Example #4 above then gives the following homeomorphisms:

$$
\begin{aligned}
S^{n-1} &\cong O(n)/O(n-1) \cong SO(n)/SO(n-1) \\
S^{2n-1} &\cong U(n)/U(n-1) \cong SU(n)/SU(n-1) \\
S^{4n-1} &\cong Sp(n)/Sp(n-1).
\end{aligned}
$$

5. S^{n-1} is the set of all $(x^1, \ldots, x^n) \in {}^n$ with $(x^1)^2 + \cdots + (x^n)^2 = 1$. Let $G = \mathbb{Z}_2$ be the discrete subgroup $\{-1, 1\}$ of the multiplicative group of nonzero real numbers. Define $\sigma : S^{n-1} \times \mathbb{Z}_2 \longrightarrow S^{n-1}$ by $\sigma(p, g) = p \cdot g = (x^1, \ldots, x^n) \cdot g = (x^1 g, \ldots, x^n g)$. Then σ is a smooth right action of $\mathbb{Z}_2$ on S^{n-1}, the orbits of which are pairs $\{p, -p\}$ of antipodal points in S^{n-1}. The orbit space (i.e., the quotient space obtained from S^{n-1} by identifying each orbit to a point) is therefore ${}^{n-1}$. Note that, when $n = 2$, this is just S^1 ((1.2.7), [**N4**]).

6. Identify S^{2n-1} with the set of all $(z^1, \ldots, z^n) \in {}^n$ satisfying $|z^1|^2 + \cdots + |z^n|^2 = 1$ and $U(1)$ with the group of all complex numbers of modulus 1. Define $\sigma : S^{2n-1} \times U(1) \longrightarrow S^{2n-1}$ by $\sigma(p, g) = p \cdot g = (z^1, \ldots, z^n) \cdot g = (z^1 g, \ldots, z^n g)$. Then σ is a smooth right action of $U(1)$ on S^{2n-1}, the orbits of which are submanifolds of S^{2n-1} diffeomorphic to $U(1)$, i.e., to S^1 (pages 51 and 260, [**N4**]). The orbit space is ${}^{n-1}$ (pages 51–52, [**N4**]). Note that, when $n = 2$, this is just S^2 ((1.2.8), [**N4**]).

7. Identify S^{4n-1} with the set of all $(q^1, \ldots, q^n) \in {}^n$ satisfying $|q^1|^2 + \cdots + |q^n|^2 = 1$ and $Sp(1) \cong SU(2)$ with the group of unit quaternions. Define $\sigma : S^{4n-1} \times Sp(1) \longrightarrow S^{4n-1}$ by $\sigma(p, g) = p \cdot g = (q^1, \ldots, q^n) \cdot g = (q^1 g, \ldots, q^n g)$. Then σ is a smooth right action of $Sp(1)$ on S^{4n-1}, the orbits of which are submanifolds of S^{4n-1} diffeomorphic to $Sp(1)$, i.e., to S^3 (Exercise 1.2.4 and page 260, [**N4**]). The orbit space is ${}^{n-1}$ (pages 51–52, [**N4**]) and, when $n = 2$, this is just S^4 ((1.2.9), [**N4**]).

1.3 Principal Bundles

Let X be a differentiable manifold and G a Lie group. A **smooth principal bundle over** X with **structure group** G (or, simply, a ***G*-bundle over** X) consists of a differentiable manifold P, a smooth map $\mathcal{P} : P \longrightarrow X$ of P onto X and a smooth right action $\sigma : P \times G \longrightarrow P$, $\sigma(p, g) = p \cdot g$, of G on P such that the following conditions are satisfied:

1. σ preserves the fibers of $\mathcal{P}$, i.e., $\mathcal{P}(p \cdot g) = \mathcal{P}(p)$ for all $p \in P$ and all $g \in G$, and

2. (**Local Triviality**) For each $x_0 \in X$ there exists an open set V in X containing x_0 and a diffeomorphism $\Psi : \mathcal{P}^{-1}(V) \longrightarrow V \times G$ of the form

$$\Psi(p) = (\mathcal{P}(p),\ \psi(p)),$$

where $\psi : \mathcal{P}^{-1}(V) \longrightarrow G$ satisfies

$$\psi\ (p \cdot g) = \psi(p) g$$

for all $p \in \mathcal{P}^{-1}(V)$ and all $g \in G$.

The pair (V, Ψ) is called a **local trivialization** of the G-bundle and a family $\{(V_j, \Psi_j)\}_{j \in J}$ of such with $\bigcup_{j \in J} V_j = X$ is called a **trivializing cover** of X. Depending on how much of this structure is clear from the context we may refer to $\mathcal{P} : P \longrightarrow X$, or even just P itself, as a principal G-bundle over X and indicate this diagrammatically by writing $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ or $G \hookrightarrow P \longrightarrow X$. For each $p \in P$, the fiber of $\mathcal{P}$ above $\mathcal{P}(p)$ coincides with the orbit of a p under σ (Lemma 4.1.1, [**N4**]) and is a submanifold of P diffeomorphic to G (page 260, [**N4**]). We begin by recording a few examples.

1. The **trivial *G*-bundle** over X consists of the product manifold $P = X \times G$, the projection $\mathcal{P} : X \times G \longrightarrow X$ onto the first factor and the action $\sigma((x,h),g) = (x,h) \cdot g = (x,hg)$. In this case one takes V in #2 above to be all of X and Ψ to be the identity map on $\mathcal{P}^{-1}(V) = \mathcal{P}^{-1}(X) = X \times G$.

2. Let $P = S^{n-1}$, $G = \mathbb{Z}_2 = \{-1, 1\}$ and $\sigma : S^{n-1} \times \mathbb{Z}_2 \longrightarrow S^{n-1}$ the right action $\sigma(p,g) = p \cdot g = (x^1, \ldots, x^n) \cdot g = (x^1 g, \ldots, x^n g)$ described in Example #5, page 27. The orbit space is $X = {}^{n-1}$ and we let $\mathcal{P} : S^{n-1} \longrightarrow {}^{n-1}$ be the quotient map, i.e., $\mathcal{P}(p) = \mathcal{P}(x^1, \ldots, x^n) = [x^1, \ldots, x^n]$ (see Example #5, page 5). Then $\mathcal{P}(p \cdot g) = \mathcal{P}(\pm p) = [p] = \mathcal{P}(p)$ for all $p \in S^{n-1}$ and $g \in \mathbb{Z}_2$. For each $k = 1, \ldots, n$, let $V_k = \{[p] = [x^1, \ldots, x^n] \in {}^{n-1} : x^k \neq 0\}$. Then $\mathcal{P}^{-1}(V_k) = \{p = (x^1, \ldots, x^n) \in S^{n-1} : x^k \neq 0\}$ and we define $\Psi_k : \mathcal{P}^{-1}(V_k) \longrightarrow V_k \times \mathbb{Z}_2$ by $\Psi_k(p) = \Psi_k(x^1, \ldots, x^n) = ([p], x^k/|x^k|)$. Then Ψ_k is a diffeomorphism of the form $\Psi_k(p) = (\mathcal{P}(p), \psi_k(p))$, where $\psi_k(p) = \psi_k(x^1, \ldots, x^n) = x^k/|x^k|$ and $\psi_k(p \cdot g) = (x^k g)/|x^k g| = (x^k g)/|x^k| = \psi_k(p) g$. Thus, $\{(V_k, \Psi_k)\}_{k=1,\ldots,n}$ is a trivializing cover of ${}^{n-1}$ and

$$\mathbb{Z}_2 \hookrightarrow S^{n-1} \xrightarrow{\mathcal{P}} {}^{n-1}$$

is a principal $\mathbb{Z}_2$-bundle over ${}^{n-1}$.

Remark: Since $\mathcal{P}^{-1}(V_k)$ is a disjoint union of two open hemispheres on S^{n-1} ($x^k > 0$ and $x^k < 0$) each of which is mapped homeomorphically onto V_k by $\mathcal{P}, \mathcal{P} : S^{n-1} \longrightarrow {}^{n-1}$ is actually a covering space (page 81, [**N4**]). This is *not* true of the complex and quaternionic analogues to which we now turn.

3. Let $P = S^{2n-1}, G = U(1)$ and $\sigma : S^{2n-1} \times U(1) \longrightarrow S^{2n-1}$ the right action $\sigma(p,g) = p \cdot g = (z^1, \ldots, z^n) \cdot g = (z^1 g, \ldots, z^n g)$ described in Example #6, page 28. The orbit space is $X = {}^{n-1}$ and we let $\mathcal{P} : S^{2n-1} \longrightarrow {}^{n-1}$ be the quotient map. Then $\mathcal{P}(p \cdot g) = \mathcal{P}(p)$ for all $p \in S^{2n-1}$ and $g \in U(1)$. For each $k = 1, \ldots, n$, let $V_k = \{[p] = [z^1, \ldots, z^n] \in {}^{n-1} : z^k \neq 0\}$ and define $\Psi_k : \mathcal{P}^{-1}(V_k) \longrightarrow V_k \times U(1)$ by $\Psi_k(p) = \Psi_k(z^1, \ldots, z^n) = ([p], z^k/|z^k|)$. Then, $\{(V_k, \Psi_k)\}_{k=1,\ldots,n}$ is a trivializing cover of ${}^{n-1}$ so

$$U(1) \hookrightarrow S^{2n-1} \xrightarrow{\mathcal{P}} {}^{n-1}$$

is a principal $U(1)$-bundle over ${}^{n-1}$.

Remark: When $n = 2$ one can identify $\ ^1$ with the 2-sphere (see the Remark on page 6) and thereby obtain a $U(1)$-bundle over S^2 generally known as the **complex Hopf bundle**. There are, in fact, two natural ways of identifying $\ ^1$ with S^2 and these yield bundles which are not "equivalent" in a sense soon to be made precise. Since we will require both of these bundles we shall write out their descriptions explicitly. We use the charts (U_1, φ_1) and (U_2, φ_2) on $\ ^1$ (Example #5, page 5) and the stereographic projection chart (U_S, φ_S) on S^2, (Example #4, page 4). On its domain, $\varphi_S^{-1} \circ \varphi_2$ is a diffeomorphism into S^2 given by $\varphi_S^{-1} \circ \varphi_2([z^1, z^2]) = \varphi_S^{-1}(z^1(z^2)^{-1}) = \varphi_S^{-1}(\frac{z^1}{z^2}) = (z^1\bar{z}^2 + \bar{z}^1 z^2, -\boldsymbol{i}\, z^1\bar{z}^2 + \boldsymbol{i}\,\bar{z}^1 z^2, |z^1|^2 - |z^2|^2)$. But this last formula does not require $z^2 \neq 0$ and, in fact, defines a diffeomorphism $[z^1, z^2] \longrightarrow (z^1\bar{z}^2 + \bar{z}^1 z^2, -\boldsymbol{i}\, z^1\bar{z}^2 + \boldsymbol{i}\,\bar{z}^1 z^2, |z^1|^2 - |z^2|^2)$ of all of $\ ^1$ onto S^2. Composing with the projection $\mathcal{P}$ of S^3 onto $\ ^1$ we obtain a map

$$\mathcal{P}_1 : S^3 \longrightarrow S^2$$

given by

$$\mathcal{P}_1\ (z^1, z^2) = (z^1\bar{z}^2 + \bar{z}^1 z^2, -\boldsymbol{i}\, z^1\bar{z}^2 + \boldsymbol{i}\,\bar{z}^1 z^2, |z^1|^2 - |z^2|^2)$$

and this provides a concrete realization of the bundle $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}} \ ^1$ where $\ ^1$ is identified with S^2:

$$U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_1} S^2$$

Similarly, $\varphi_S^{-1} \circ \varphi_1([z^1, z^2]) = \varphi_S^{-1}\left(\frac{z^2}{z^1}\right) = (\bar{z}^1 z^2 + z^1\bar{z}^2, -\boldsymbol{i}\,\bar{z}^1 z^2 + \boldsymbol{i}\, z^1\bar{z}^2, |z^1|^2 - |z^2|^2)$ actually determines a global diffeomorphism of $\ ^1$ onto S^2 and thus a map

$$\mathcal{P}_{-1} : S^3 \longrightarrow S^2$$

given by

$$\mathcal{P}_{-1}(z^1, z^2) = (\bar{z}^1 z^2 + z^1\bar{z}^2, -\boldsymbol{i}\,\bar{z}^1 z^2 + \boldsymbol{i}\, z^1\bar{z}^2, |z^1|^2 - |z^2|^2)$$

and therefore another bundle

$$U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_{-1}} S^2.$$

Notice that $\mathcal{P}_{-1}(z^1, z^2)$ differs from $\mathcal{P}_1(z^1, z^2)$ only in the sign of the second coordinate so $\mathcal{P}_{-1}$ is $\mathcal{P}_1$ followed by reflection across the "*xz*-plane."

4. Let $P = S^{4n-1}, G = Sp(1)$ and $\sigma : S^{4n-1} \times Sp(1) \longrightarrow S^{4n-1}$ the right action $\sigma(p, g) = p \cdot g = (q^1, \ldots, q^n) \cdot g = (q^1 g, \ldots, q^n g)$ described in Example #7, page 28. The orbit space is $X = \ ^{n-1}$ and we let $\mathcal{P} : S^{4n-1} \longrightarrow \ ^{n-1}$ be the quotient map. Then $\mathcal{P}(p \cdot g) = \mathcal{P}(p)$ for all $p \in S^{4n-1}$ and $g \in Sp(1)$. For each $k = 1, \ldots, n$, let $V_k = \{[p] = [q^1, \ldots, q^n] \in \ ^{n-1} : q^k \neq 0\}$ and define $\Psi_k : \mathcal{P}^{-1}(V_k) \longrightarrow V_k \times Sp(1)$ by $\Psi_k(p) = \Psi_k(q^1, \ldots, q^n) = ([p], q^k/|q^k|)$.

Then, $\{(V_k, \Psi_k)\}_{k=1,\ldots,n}$ is a trivializing cover of $\quad^{n-1}$ so

$$Sp(1) \hookrightarrow S^{4n-1} \xrightarrow{\mathcal{P}} \quad^{n-1}$$

is a principal $Sp(1)$-bundle over $\quad^{n-1}$.

Remark: Just as in the complex case (page 30), when $n = 2$ one can identify $\quad^{1}$ with S^4 in two natural ways and thereby obtain two **quaternionic Hopf bundles**

$$Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}_1} S^4 \text{ and } Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}_{-1}} S^4.$$

5. Let $\mathcal{P} : P \longrightarrow X$ be any principal G-bundle over X (with right action σ) and suppose X' is a submanifold of X. Let $P' = \mathcal{P}^{-1}(X'), \mathcal{P}' = \mathcal{P}|P'$ and $\sigma' = \sigma|P' \times G$. For each local trivialization (V, Ψ) of $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ with $V \cap X' \neq \emptyset$ set $V' = V \cap X'$ and $\Psi' = \Psi|(\mathcal{P}')^{-1}(V')$. Then $\mathcal{P}'$: $P' \longrightarrow X'$ is a principal G-bundle over X', called the **restriction** of $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ to X', and each (V', Ψ') is a local trivialization for $G \hookrightarrow P' \xrightarrow{\mathcal{P}'} X'$ (Exercise 4.1.3, [**N4**]).

Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a principal G-bundle over X and fix a trivializing cover $\{(V_j, \Psi_j)\}_{j \in J}$ of X. Write each Ψ_j as $\Psi_j(p) = (\mathcal{P}(p), \psi_j(p))$ for all $p \in \mathcal{P}^{-1}(V_j)$, where $\psi_j(p \cdot g) = \psi_j(p)g$ for all $p \in \mathcal{P}^{-1}(V_j)$ and $g \in G$. Suppose $i, j \in J$ and $V_i \cap V_j \neq \emptyset$. Then, for any $x \in V_i \cap V_j$, $\psi_j(p)(\psi_i(p))^{-1}$ takes the same value for every $p \in \mathcal{P}^{-1}(x)$. Thus, we may define

$$g_{ji} : V_i \cap V_j \longrightarrow G$$

by

$$g_{ji}(x) = (\psi_j(p))\ (\psi_i(p))^{-1}$$

for any $p \in \mathcal{P}^{-1}(x)$. These maps are smooth and are called the **transition functions** for $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ corresponding to $\{(V_j, \Psi_j)\}_{j \in J}$. They satisfy all of the following conditions:

$$\begin{aligned} g_{ii}(x) &= e \\ g_{ij}(x) &= (g_{ji}(x))^{-1} \\ g_{kj}(x)g_{ji}(x) &= g_{ki}(x) \qquad (\textbf{cocycle condition}) \end{aligned}$$

whenever $i, j, k \in J$ and $V_i \cap V_j \cap V_k \neq \emptyset$. A (**local**) **cross-section** of $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ defined on an open set $V \subseteq X$ is a smooth map $s : V \longrightarrow \mathcal{P}^{-1}(V)$ that satisfies $\mathcal{P} \circ s = id_V$, i.e., it is a smooth selection of an element from each fiber above V. A cross-section s on V gives rise to a trivialization (V, Ψ), where $\Psi : \mathcal{P}^{-1}(V) \longrightarrow V \times G$ is given by $\Psi(s(x) \cdot g) = (x, g)$ (page 221, [**N4**]). Conversely, a trivialization (V, Ψ)

gives rise to a cross-section $s : V \longrightarrow \mathcal{P}^{-1}(V)$ defined by $s(x) = \Psi^{-1}(x, e)$ and this correspondence between trivializations and cross-sections is bijective (pages 220–221, [**N4**]).

As concrete illustrations of these last few notions we consider the bundle $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad ^1$ (the $n = 2$ case of Example #4, page 31) and its trivializing cover $\{(V_k, \Psi_k)\}_{k=1,2}$. Then $V_1 = \{x = [q^1, q^2] \in \quad ^1 : q^1 \neq 0\}, V_2 = \{x = [q^1, q^2] \in \quad ^1 : q^2 \neq 0\}$, $\psi_1(p) = \psi_1(q^1, q^2) = q^1/|q^1|$ and $\psi_2(q^1, q^2) = q^2/|q^2|$ so the transition functions $g_{12}, g_{21} : V_1 \cap V_2 \longrightarrow Sp(1)$ are given by

$$g_{12}(x) = g_{12}([q^1, q^2]) = (q^1/|q^1|)(q^2/|q^2|)^{-1}, \quad \text{and}$$
$$g_{21}(x) = g_{21}([q^1, q^2]) = (q^2/|q^2|)(q^1/|q^1|)^{-1}.$$

One can check that the inverses $\Psi_k^{-1} : V_k \times Sp(1) \longrightarrow \mathcal{P}^{-1}(V_k), k = 1, 2$, are given by $\Psi_1^{-1}([q^1, q^2], y) = (|q^1|y, q^2(q^1/|q^1|)^{-1}y) \in S^7 \subseteq \quad ^2$ and $\Psi_2^{-1}([q^1, q^2], y) = (q^1(q^2/|q^2|)^{-1}y, |q^2|y) \in S^7 \subseteq \quad ^2$ so the associated cross-sections $s_k : V_k \longrightarrow \mathcal{P}^{-1}(V_k), k = 1, 2$, are

$$s_1(x) = s_1([q^1, q^2]) = (|q^1|, q^2(q^1/|q^1|)^{-1}), \quad \text{and}$$
$$s_2(x) = s_2([q^1, q^2]) = (q^1(q^2/|q^2|)^{-1}, |q^2|).$$

For future reference we note that V_1 and V_2 are also the standard coordinate neighborhoods on $\quad ^1$ (called U_1 and U_2 in Example #5, page 5) and that the corresponding diffeomorphisms $\varphi_k : V_k \longrightarrow \quad = \quad ^4$ are

$$\varphi_1(x) = \varphi_1([q^1, q^2]) = q^2(q^1)^{-1} \quad \text{and}$$
$$\varphi_2(x) = \varphi_2([q^1, q^2]) = q^1(q^2)^{-1}.$$

Their inverses are $\varphi_1^{-1}(q) = [1, q]$ and $\varphi_2^{-1}(q) = [q, 1]$ so the overlap maps are

$$\varphi_2 \circ \varphi_1^{-1}(q) = q^{-1} = \varphi_1 \circ \varphi_2^{-1}(q)$$

for all $q \in \quad - \{0\}$. All of this is the same for $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}} \quad ^1$.

Let $G \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} X_1$ and $G \hookrightarrow P_2 \xrightarrow{\mathcal{P}_2} X_2$ be two principal G-bundles and, for convenience, denote the actions of G on P_1 and P_2 by the same dot $\cdot$. A (**principal**) **bundle map** from P_1 to P_2 is a smooth map $f : P_1 \longrightarrow P_2$ which satisfies $f(p \cdot g) = f(p) \cdot g$ for all $p \in P_1$ and $g \in G$. Such an f carries each fiber of $\mathcal{P}_1$ diffeomorphically onto some fiber of $\mathcal{P}_2$ and therefore induces a smooth map $\bar{f} : X_1 \longrightarrow X_2$ defined by $\mathcal{P}_2 \circ f = \bar{f} \circ \mathcal{P}_1$. If $X_1 = X_2 = X$, then a bundle map $f : P_1 \longrightarrow P_2$ is called an **equivalence** if it is a diffeomorphism and induces the identity map on X, i.e., $\bar{f} = id_X$. In this case the bundles $G \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} X$ and $G \hookrightarrow P_2 \xrightarrow{\mathcal{P}_2} X$ are said to be **equivalent**. It then follows that $f^{-1} : P_2 \longrightarrow P_1$ is also an equivalence. If $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is a single principal G-bundle over X, then an equivalence $f : P \longrightarrow P$ is called an **automorphism** of the bundle. A principal G-bundle over X is said to be **trivial** if it is equivalent to the trivial G-bundle $G \hookrightarrow X \times G \longrightarrow X$ over X. This is the case if and only if the bundle has a **global trivialization** (i.e., if and only if it is possible to take $V = X$ in #2 of the definition on page 28) and this, in turn, is the case if and only if it has a **global cross-section** $s : X \longrightarrow P$. Determining whether or not a given principal bundle is trivial (i.e., admits a global cross-section) is generally not a simple matter. Any principal G-bundle over a contractible manifold X (e.g., ^{n}) is necessarily trivial. Notice finally that there is a difference between "trivial" and "trivialized". The former asserts the existence of a global cross-section, while the latter amounts to a specific choice of such a cross-section/trivialization.

We record now the smooth versions of a number of basic results that are proved in the topological context (i.e., for C^0-principal bundles) in Chapter 4 of [**N4**]. Extending these to smooth (i.e., C^∞) principal bundles involves some technical issues that will be discussed in Section 3.2.

The Reconstruction Theorem: *Let X be a smooth manifold, G a Lie group and $\{V_j\}_{j\in J}$ an open cover of X. Suppose that, for each $i, j \in J$ with $V_i \cap V_j \neq \emptyset$, there is given a smooth map $g_{ji} : V_i \cap V_j \longrightarrow G$ and that these maps have the property that, if $V_i \cap V_j \cap V_k \neq \emptyset$, then*

$$g_{kj}(x)g_{ji}(x) = g_{ki}(x)$$

for all $x \in V_i \cap V_j \cap V_k$. Then there exists a principal G-bundle over X, unique up to equivalence, which has the $V_j, j \in J$, as trivializing neighborhoods and the g_{ji} as corresponding transition functions.

The Classification Theorem: *Let G be a connected Lie group. Then the set of equivalence classes of principal G-bundles over $S^n, n \geq 2$, is in one-to-one correspondence with the elements of the homotopy group $\pi_{n-1}(G)$.* Thus, for example, the principal $U(1)$-bundles over S^2 are in one-to-one correspondence with the elements of $\pi_1(U(1)) \cong \pi_1(S^1) \cong \mathbb{Z}$ (the integers). Similarly, the $Sp(1)$-bundles over S^4 are in one-to-one correspondence with the elements of $\pi_3(Sp(1)) \cong \pi_3(S^3) \cong \mathbb{Z}$. We will find that the theory of characteristic classes provides a natural means of associating an integer (Chern number) with a bundle of either of these types.

For any smooth principal bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ the fibers $\mathcal{P}^{-1}(x), x \in X$, are all submanifolds of P diffeomorphic to G (page 260, [**N4**]) so, for each $p \in P$, $T_p(P)$ contains a subspace isomorphic to the Lie algebra $\mathcal{G}$ of G (all tangent vectors at p to smooth curves in the fiber containing p). We call this the **vertical subspace** of $T_p(P)$ and denote it $\mathrm{Vert}_p(P)$. The elements of $\mathrm{Vert}_p(P)$ are called **vertical vectors** at p. The action σ of G on P provides a natural means of identifying each $\text{Vert}_p(P)$ with $\mathcal{G}$. To see this, fix an element $A \in \mathcal{G}$ (thought of as a set of matrices). We associate with A a vector field $A^{\#}$ on P, called the **fundamental vector field** on P determined by A, as follows: For each $p \in P$ the map $\sigma_p : G \longrightarrow P$ defined by $\sigma_p(g) = \sigma(p,g) = p \cdot g$ is smooth and so has a derivative $(\sigma_p)_{*id}$ at the identity. Then,

$$A^{\#}(p) = (\sigma_p)_{*id}(A) = \frac{d}{dt}(p \cdot \exp(tA))|_{t=0}$$

(page 287, [**N4**]). The mapping $A \longrightarrow A^{\#}(p)$ is an isomorphism of $\mathcal{G}$ onto $\text{Vert}_p(P)$ (Corollary 5.8.9, [**N4**]). Furthermore, for any $A,\ B \in \mathcal{G}$,

$$[A,B]^{\#} = [A^{\#}, B^{\#}]$$

(Theorem 5.8.8, [**N4**]) and, for all $g \in G$,

$$(\sigma_g)_*(A^{\#}) = \left(ad_{g^{-1}}(A)\right)^{\#}.$$

1.4 Connections and Curvature

A **connection (gauge field)** on a principal bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ with action σ is a smooth $\mathcal{G}$-valued 1-form $\boldsymbol{\omega}$ on P which satisfies

1. $(\sigma_g)^*\boldsymbol{\omega} = ad_{g^{-1}} \circ \boldsymbol{\omega}$ for all $g \in G$, i.e., for all $g \in G, p \in P$ and $\boldsymbol{v} \in T_{p\cdot g^{-1}}(P)$,
$$\boldsymbol{\omega}_p\left((\sigma_g)_{*p\cdot g^{-1}}(\boldsymbol{v})\right) = g^{-1}\boldsymbol{\omega}_{p\cdot g^{-1}}(\boldsymbol{v})g,$$

2. $\boldsymbol{\omega}(A^{\#}) = A$ for all $A \in \mathcal{G}$, i.e., for all $g \in G$ and $p \in P$,
$$\boldsymbol{\omega}_p(A^{\#}(p)) = A.$$

A local cross-section $s : V \longrightarrow \mathcal{P}^{-1}(V)$ of the bundle is called a **local gauge** and the pullback $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{\omega}$ of $\boldsymbol{\omega}$ to $V \subseteq X$ by s is called a **local gauge potential** (in gauge s). If $\{(V_j, \Psi_j)\}_{j\in J}$ is a trivializing cover of X and $s_j : V_j \longrightarrow \mathcal{P}^{-1}(V_j)$ is the cross-section associated with (V_j, Ψ_j) (page 32), then the family $\{\boldsymbol{\mathcal{A}}_j = s_j^*\boldsymbol{\omega}\}_{j\in J}$ of local gauge potentials satisfies

$$\boldsymbol{\mathcal{A}}_j = ad_{g_{ij}^{-1}} \circ \boldsymbol{\mathcal{A}}_i + g_{ij}{}^*\boldsymbol{\Theta}$$

for all $i,j \in J$ with $V_i \cap V_j \neq \emptyset$, where g_{ij} is the corresponding transition function and $\boldsymbol{\Theta}$ is the Cartan 1-form for G (Lemma 5.9.2, [**N4**]). Conversely, given a trivializing cover $\{(V_j, \Psi_j)\}_{j\in J}$ for some principal G-bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ and a $\mathcal{G}$-valued 1-form $\boldsymbol{\mathcal{A}}_j$ on V_j for each $j \in J$ with $\boldsymbol{\mathcal{A}}_j$ and $\boldsymbol{\mathcal{A}}_i$ related in this way whenever $V_i \cap V_j \neq \emptyset$, there exists a unique connection form $\boldsymbol{\omega}$ on P such that $\boldsymbol{\mathcal{A}}_j = s_j^*\boldsymbol{\omega}$ for each $j \in J$ (Theorem 6.1.1, [**N4**]). The $\mathcal{G}$-valued 1-forms ${g_{ij}}^*\boldsymbol{\Theta}$ on $V_i \cap V_j$ are readily computable when G is a matrix group for then one can show that, for any $x \in V_i \cap V_j$ and any $\boldsymbol{v} \in T_x(X), ({g_{ij}}^*\boldsymbol{\Theta})_x(\boldsymbol{v}) = (g_{ij}(x))^{-1}\, dg_{ij}\,(x)(\boldsymbol{v})$, where dg_{ij} is the entrywise differential of $g_{ij} : V_j \cap V_i \longrightarrow G$ (pages 305, [**N4**]). The relationship between the gauge potentials $\boldsymbol{\mathcal{A}}_j$ and $\boldsymbol{\mathcal{A}}_i$ can then be written

$$\boldsymbol{\mathcal{A}}_j = {g_{ij}}^{-1}\boldsymbol{\mathcal{A}}_i\, g_{ij} + {g_{ij}}^{-1}dg_{ij}.$$

Given a connection form $\boldsymbol{\omega}$ on P one defines, for each $p \in P$, the **horizontal subspace** $\mathrm{Hor}_p(P)$ of $T_p(P)$ by $\mathrm{Hor}_p(P) = \{\boldsymbol{v} \in T_p(P) : \boldsymbol{\omega}_p(\boldsymbol{v}) = 0\}$. Then

$$T_p(P) = \mathrm{Hor}_p(P) \oplus \mathrm{Vert}_p(P)$$

((6.1.2), [**N4**]) so every $\boldsymbol{v} \in T_p(P)$ can be written uniquely as $\boldsymbol{v} = \boldsymbol{v}^H + \boldsymbol{v}^V$, where $\boldsymbol{v}^H \in \mathrm{Hor}_p(P)$ and $\boldsymbol{v}^V \in \mathrm{Vert}_p(P)$. Similarly, a smooth vector field $\boldsymbol{V}$ on P can be written $\boldsymbol{V} = \boldsymbol{V}^H + \boldsymbol{V}^{\boldsymbol{V}}$, where $\boldsymbol{V}^H$ and $\boldsymbol{V}^V$ are smooth and, for each $p \in P$, $\boldsymbol{V}^H(p) \in \mathrm{Hor}_p(P)$ and $\boldsymbol{V}^V(p) \in \mathrm{Vert}_p\,(P)$ (page 347, [**N4**]). If $p \in P$ and $\mathcal{P}(p) = x$, then $\mathcal{P}_{*p}$ carries $\mathrm{Hor}_p(P)$ isomorphically onto $T_x(X)$ (Exercise 6.1.7, [**N4**]). Furthermore, the horizontal subspaces are invariant under the action of G on P in the sense that

$$(\sigma_g)_{*p}\,(\mathrm{Hor}_p(P)) = \mathrm{Hor}_{p\cdot g}(P)$$

for all $p \in P$ and $g \in G$ ((6.1.3), [**N4**]). If $\dim X = n$, then the assignment $p \longrightarrow \mathrm{Hor}_p(P)$ is an example of a smooth n-dimensional distribution on P (page 333, [**N4**]) and, moreover, any smooth n-dimensional distribution $p \longrightarrow \mathcal{D}\,(p)$ on P that satisfies $T_p(P) \cong \mathcal{D}(p) \oplus \mathrm{Vert}_p(P)$ and $(\sigma_g)_{*p}(\mathcal{D}(p)) = \mathcal{D}(p \cdot g)$ for all $p \in P$ and $g \in G$ is the distribution of horizontal subspaces for some connection form on P (Exercise 6.1.9, [**N4**]). One often sees a connection on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ defined in any one of the three equivalent ways described above ($\mathcal{G}$-valued 1-form, collection of local gauge potentials, or distribution of horizontal subspaces).

If $G \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} X$ and $G \hookrightarrow P_2 \xrightarrow{\mathcal{P}_2} X$ are two principal G-bundles over $X, f : P_1 \longrightarrow P_2$ is a bundle map and $\boldsymbol{\omega}$ is a connection form on $G \hookrightarrow P_2 \xrightarrow{\mathcal{P}_2} X$, then $f^*\boldsymbol{\omega}$ is a connection form on $G \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} X$ (Theorem 6.1.3, [**N4**]). This is true, in particular, for an automorphism of a single principal bundle. If $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are two connection forms on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$, and if there exists an automorphism $f : P \longrightarrow P$ such that $\boldsymbol{\omega}_2 = f^*\boldsymbol{\omega}_1$, then $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are said to be **gauge equivalent** (the origin of the terminology is in physics and will be discussed shortly). Notice that, if $s : V \longrightarrow \mathcal{P}^{-1}(V)$ is a cross-section and $\boldsymbol{\mathcal{A}}_i = s^*\boldsymbol{\omega}_i$ is the corresponding gauge potential of $\boldsymbol{\omega}_i, i = 1, 2$, then $\boldsymbol{\mathcal{A}}_2 = s^*\boldsymbol{\omega}_2 = s^*(f^*\boldsymbol{\omega}_1) = (f \circ s)^*\boldsymbol{\omega}_1$. Since $s' = f \circ s : V \longrightarrow \mathcal{P}^{-1}(V)$ is also a cross-section, we conclude that $\boldsymbol{\mathcal{A}}_1 = s^*\boldsymbol{\omega}_1$ and $\boldsymbol{\mathcal{A}}_2 = (s')^*\boldsymbol{\omega}_1$ are both gauge potentials for the *same* connection $\boldsymbol{\omega}_1$ (by different cross-sections).

A connection also gives rise to a "path lifting procedure" from X to P and thereby notions of "parallel translation" and "holonomy." Specifically, if $\boldsymbol{\omega}$ is a connection form on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ and $\alpha : [0,1] \longrightarrow X$ is a smooth curve in X with $\alpha(0) = x_0$, then for each $p_0 \in \mathcal{P}^{-1}(x_0)$ there exists a unique smooth curve $\tilde{\alpha}_{p_0} : [0,1] \longrightarrow P$ such that

1. $\tilde{\alpha}_{p_0}(0) = p_0$,
2. $\mathcal{P} \circ \tilde{\alpha}_{p_0}(t) = \alpha(t)$ for all $t \in [0,1]$, and
3. $\tilde{\alpha}'_{p_0}(t) \in \mathrm{Hor}_{\tilde{\alpha}_{p_0}(t)}(P)$ for all $t \in [0,1]$

(Theorem 6.1.4, [**N4**]). In particular, if $\alpha(1) = x_1$, then $\tilde{\alpha}_{p_0}(1) \in \mathcal{P}^{-1}(x_1)$ and we may define a map $\tau_\alpha : \mathcal{P}^{-1}(x_0) \longrightarrow \mathcal{P}^{-1}(x_1)$, called the **parallel translation along** α determined by $\boldsymbol{\omega}$, by $\tau_\alpha(p_0) = \tilde{\alpha}_{p_0}(1)$ for each $p_0 \in \mathcal{P}^{-1}(x_0)$. If $x_1 = x_0$ (i.e., if α is a loop in X), then $\tau_\alpha : \mathcal{P}^{-1}(x_0) \longrightarrow \mathcal{P}^{-1}(x_0)$. Since G acts transitively on the fibers of $\mathcal{P}$ (Lemma 4.1.1, [**N4**]), for each $p_0 \in \mathcal{P}^{-1}(x_0)$ there is a unique $g \in G$ such that $\tau_\alpha(p_0) = p_0 \cdot g$. Holding p_0 fixed and allowing α to vary over all smooth loops at x_0 in X one obtains a subset $\mathcal{H}(p_0)$ of G consisting of all those $g \in G$ such that p_0 is parallel translated to $p_0 \cdot g$ over some smooth loop at x_0. $\mathcal{H}(p_0)$ is, in fact, a subgroup of G (Exercise 6.1.24, [**N4**]) called the **holonomy group** of $\boldsymbol{\omega}$ at p_0. We record now a number of important examples of connections on principal bundles.

1. (Flat connections on trivial bundles) Let X be any smooth manifold, G any matrix Lie group, and $G \hookrightarrow X \times G \xrightarrow{\mathcal{P}} X$ the corresponding trivial G-bundle (Example #1, page 29). Let $\boldsymbol{\Theta}$ be the Cartan 1-form on G and define a $\mathcal{G}$-valued 1-form $\boldsymbol{\omega}$ on $X \times G$ by $\boldsymbol{\omega} = \pi^*\boldsymbol{\Theta}$, where $\pi : X \times G \longrightarrow G$ is the projection onto G. Then $\boldsymbol{\omega}$ is a connection form on $X \times G$ whose horizontal subspace $\mathrm{Hor}_{(x,g)}(X \times G)$ at any $(x,g) \in X \times G$ is the tangent space to the submanifold $X \times \{g\}$ at (x,g) (Exercise 6.2.12, [**N4**]).

Remark: The reason these connections are called "flat" will emerge when we discuss the "curvature" of a connection.

2. (Natural connection on the complex Hopf bundle) We consider the $U(1)$-bundle

$$U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}} \quad {}^1$$

(the $n = 2$ case of Example #3, page 30). Regard S^3 as the submanifold of ^{2} consisting of those (z^1, z^2) with $|z^1|^2 + |z^2|^2 = 1$ and identify the Lie algebra of $U(1)$ with the algebra Im of pure imaginary complex numbers. Define an Im -valued 1-form $\tilde{\boldsymbol{\omega}}$ on ^{2} by $\tilde{\boldsymbol{\omega}} = \boldsymbol{i}\,\mathrm{Im}\,(\bar{z}^1 dz^1 + \bar{z}^2 dz^2)$. Thus, for each

$p = (p^1, p^2) \in \;^2$ and $\boldsymbol{v} = (v^1, v^2) \in T_p(\;^2) \cong T_{p^1}(\;) \oplus T_{p^2}(\;) \cong \; \oplus \;$, $\tilde{\boldsymbol{\omega}}_p(\boldsymbol{v}) = \boldsymbol{i}\,\mathrm{Im}(\bar{p}^1 v^1 + \bar{p}^2 v^2)$. Now let $\boldsymbol{\omega}$ be the restriction of $\tilde{\boldsymbol{\omega}}$ to S^3 (i.e., $\boldsymbol{\omega} = \iota^* \tilde{\boldsymbol{\omega}}$, where $\iota : S^3 \hookrightarrow \;^2$ is the inclusion map). Then $\boldsymbol{\omega}$ is a connection form on $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}} \;^1$(the proof is exactly the same as that for the quaternionic case on pages 297–299, [**N4**]). For each $p = (p^1, p^2) \in S^3$, $\mathrm{Vert}_p(S^3)$ is the tangent space to the fiber of $\mathcal{P}$ containing p and this is a 1-dimensional subspace of $T_p(S^3)$ since this fiber is diffeomorphic to S^1. The horizontal subspace $\mathrm{Hor}_p(S^3)$ determined by $\boldsymbol{\omega}$ is just that part of the real orthogonal complement of $\mathrm{Vert}_p(S^3)$ in $\;^2 = \;^4$ that lies in $T_p(S^3)$ (the proof is identical to that in the quaternionic case on page 334, [**N4**]). One could compute gauge potentials $s_k^* \boldsymbol{\omega}$, $k = 1, 2$, for the cross-sections corresponding to the standard trivializations $(V_k,\ \Psi_k)$, $k = 1, 2$, and the results would be entirely analogous to those in the quaternionic case derived on pages 297–303, [**N4**]. However, in order to facilitate comparison with the physics (Dirac monopoles) to be discussed in the next section we record instead the corresponding results when $\;^1$ is identified with S^2 (see the Remark following Example #3, page 30). First consider $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_1} S^2$. Here the diffeomorphism $[z^1,\, z^2] \longrightarrow (z^1 \bar{z}^2 + \bar{z}^1 z^2,\ -\boldsymbol{i} z^1 \bar{z}^2 + \boldsymbol{i} \bar{z}^1 z^2,\ |z^1|^2 - |z^2|^2)$ of $\;^1$ onto S^2 identifies V_1 and V_2 with U_N and U_S, respectively, and so it is on these sets that we have standard cross-sections $s_N : U_N \longrightarrow \mathcal{P}_1^{-1}(U_N)$ and $s_S : U_S \longrightarrow \mathcal{P}_1^{-1}(U_S)$. Our interest is in the gauge potentials $\boldsymbol{\mathcal{A}}_N = s_N^* \boldsymbol{\omega}$ and $\boldsymbol{\mathcal{A}}_S = s_S^* \boldsymbol{\omega}$ and we wish to describe them in terms of standard spherical coordinates ϕ and θ on S^2 (page 238, [**N4**]), i.e., we want $(s_N \circ \varphi^{-1})^* \boldsymbol{\omega}$ and $(s_S \circ \varphi^{-1})^* \boldsymbol{\omega}$, where φ is a spherical coordinate chart on $S^2 (\varphi^{-1}(\phi,\ \theta) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi))$. These are computed by writing $\boldsymbol{\omega} = \iota^* \tilde{\boldsymbol{\omega}} = \iota^*(\boldsymbol{i}\,\mathrm{Im}(\bar{z}^1\, dz^1 + \bar{z}^2\, dz^2)) = \boldsymbol{i}\,\iota^*(\mathrm{Im}(\bar{z}^1 dz^1 + \bar{z}^2 dz^2)) = \boldsymbol{i}\iota^*(-x^2 dx^1 + x^1 dx^2 - x^4 dx^3 + x^3 dx^4)$, where $z^1 = x^1 + x^2 \boldsymbol{i}$ and $z^2 = x^3 + x^4 \boldsymbol{i}$ so that, for example,

$$(s_N \circ \varphi^{-1})^* \boldsymbol{\omega} = \boldsymbol{i} (\iota \circ s_N \circ \varphi^{-1})^* (-x^2 dx^1 + x^1 dx^2 - x^4 dx^3 + x^3 dx^4).$$

But

$$\begin{aligned} (\iota \circ s_N \circ \varphi^{-1})(\phi,\ \theta) &= \iota \circ s_N (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi) \\ &= \left(\cos\frac{\phi}{2}, 0, \sin\frac{\phi}{2}\cos\theta, -\sin\frac{\phi}{2}\sin\theta \right) \end{aligned}$$

(page 269, [**N4**]). A simple calculation with the coordinate formula for pullbacks on page 9 gives

$$(s_N \circ \varphi^{-1})^* \boldsymbol{\omega} \;= -\frac{1}{2} \boldsymbol{i} (1 - \cos\phi)\, d\theta$$

(page 270, [**N4**]). Similarly,

$$(s_S \circ \varphi^{-1})^* \boldsymbol{\omega} \;= \frac{1}{2} \boldsymbol{i} (1 + \cos\phi)\, d\theta.$$

Analogous calculations for $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_{-1}} S^2$ give results that differ from these only by a sign.

3. (Natural connection on the quaternionic Hopf bundle) We consider the $Sp(1)$-bundle

$$Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad {}^1$$

(the $n = 2$ case of Example #4, page 31). Regard S^7 as the submanifold of $\quad^2$ consisting of those (q^1, q^2) with $|q^1|^2 + |q^2|^2 = 1$ and identify the Lie algebra of $Sp(1)$ with the algebra Im of pure imaginary quaternions. Define an Im -valued 1-form $\tilde{\boldsymbol{\omega}}$ on $\quad^2$ by $\tilde{\boldsymbol{\omega}} = \mathrm{Im}\,(\bar{q}^1\, dq^1 + \bar{q}^2\, dq^2)$ and let $\boldsymbol{\omega}$ be the restriction of $\tilde{\boldsymbol{\omega}}$ to S^7. Then $\boldsymbol{\omega}$ is a connection form on $Sp\,(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad{}^1$ ((5.9.10) and (5.9.11), [**N4**]). For each $p \in S^7$, $\mathrm{Hor}_p(S^7)$ is that part of the real orthogonal complement of $\mathrm{Vert}_p(S^7)$ in $\quad^2 = \quad^8$ that lies in $T_p(S^7)$ (page 335, [**N4**]). The standard trivializations $\{(V_1, \Psi_1), (V_2, \Psi_2)\}$ of the bundle give cross-sections $s_1 : V_1 \longrightarrow \mathcal{P}^{-1}(V_1)$ and $s_2 : V_2 \longrightarrow \mathcal{P}^{-1}(V_2)$ and we are interested in the gauge potentials $\boldsymbol{\mathcal{A}}_1 = s_1^*\boldsymbol{\omega}$ and $\boldsymbol{\mathcal{A}}_2 = s_2^*\boldsymbol{\omega}$. Since V_1 and V_2 are also coordinate neighborhoods for the standard charts $\varphi_1 : V_1 \longrightarrow$ and $\varphi_2 : V_2 \longrightarrow$ we may compute these gauge potentials $\boldsymbol{\mathcal{A}}_1$ and $\boldsymbol{\mathcal{A}}_2$ in terms of φ_1 and φ_2 coordinates. The results are as follows (pages 297–301, [**N4**]):

$$(s_1 \circ \varphi_1^{-1})^* \boldsymbol{\omega} \ = \mathrm{Im}\left(\frac{\bar{q}}{1 + |q|^2} dq\right)$$

and

$$(s_2 \circ \varphi_2^{-1})^* \boldsymbol{\omega} \ = \mathrm{Im}\left(\frac{\bar{q}}{1 + |q|^2} dq\right).$$

Thus, for any $x \in V_1$ and any $\boldsymbol{X} \in T_x(\quad^1)$,

$$\begin{aligned}(s_1^*\boldsymbol{\omega}\ )_x(\boldsymbol{X}) &= ((s_1 \circ \varphi_1^{-1}))^*\boldsymbol{\omega})_{\varphi_1(x)}\,((\varphi_1)_{*x}\boldsymbol{X})\\ &= \mathrm{Im}\left(\frac{\overline{\varphi_1(x)}v}{1 + |\varphi_1(x)|^2}\right),\end{aligned}$$

where $v = dq\,((\varphi_1)_{*x}(\boldsymbol{X}))$ and, for any $x \in V_2$ and any $\boldsymbol{X} \in T_x(\quad^1)$,

$$\begin{aligned}(s_2^*\boldsymbol{\omega}\ )_x(\boldsymbol{X}) &= \left((s_2 \circ \varphi_2^{-1})^*\boldsymbol{\omega}\right)_{\varphi_2(x)}\,((\varphi_2)_{*x}\boldsymbol{X})\\ &= \mathrm{Im}\left(\frac{\overline{\varphi_2(x)}w}{1 + |\varphi_2(x)|^2}\right),\end{aligned}$$

where $w = dq\,((\varphi_2)_{*x}(\boldsymbol{X}))$. Notice that, on $\quad - \{0\}$,

$$\mathrm{Im}\left(\frac{\bar{q}}{1 + |q|^2}\, dq\right) = \frac{|q|^2}{1 + |q|^2}\,\mathrm{Im}\,(q^{-1}dq).$$

To compare the two gauge potentials $\boldsymbol{\mathcal{A}}_1$ and $\boldsymbol{\mathcal{A}}_2$ one expresses both of them in terms of the same coordinates on $V_1 \cap V_2$. For example, for any $x \in V_1 \cap V_2$ and any $\boldsymbol{X} \in T_x(\quad{}^1)$,

$$(s_2^*\boldsymbol{\omega})_x\ (\boldsymbol{X}) = \operatorname{Im}\left(\frac{\varphi_1(x)\bar{v}}{|\varphi_1(x)|^2(1+|\varphi_1(x)|^2)}\right),$$

where $v = dq((\varphi_1)_{*x}(\boldsymbol{X}))$ (pages 301–303, [**N4**]). Thus, on $V_1 \cap V_2$,

$$(s_2 \circ \varphi_1^{-1})^*\boldsymbol{\omega} = \operatorname{Im}\left(\frac{q}{|q|^2(1+|q|^2)}\, d\bar{q}\right) = \frac{1}{1+|q|^2}\operatorname{Im}\left(\bar{q}^{-1}\, d\bar{q}\right).$$

The gauge potentials $\boldsymbol{\mathcal{A}}_1 = s_1^*\boldsymbol{\omega}$ and $\boldsymbol{\mathcal{A}}_2 = s_2^*\boldsymbol{\omega}$ arose first in the physics literature [**BPST**] as solutions to the Yang-Mills equations (Section 6.3, [**N4**]). There they were called **pseudoparticles**. Today it is more common to refer to them (or the natural connection $\boldsymbol{\omega}$ from which they arose and which they uniquely determine) as **instantons**.

4. (More instantons on $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad{}^1$) Let $\boldsymbol{\omega}$ denote the natural connection on $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad{}^1$ described in the previous example. For any bundle map $f : S^7 \longrightarrow S^7$, $f^*\boldsymbol{\omega}$ is also a connection on $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad{}^1$ (page 37). By judiciously selecting bundle maps f (pages 336–341, [**N4**]) one can produce, for each pair $(\lambda, n) \in (0,\infty) \times \quad$, a connection form $\boldsymbol{\omega}_{\lambda,n}$ on $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad{}^1$, uniquely determined by the two gauge potentials

$$\left(s_2 \circ \varphi_2^{-1}\right)^* \boldsymbol{\omega}_{\lambda,n} = \operatorname{Im}\left(\frac{\bar{q}-\bar{n}}{\lambda^2+|q-n|^2}\, dq\right)$$

and

$$\left(s_1 \circ \varphi_1^{-1}\right)^* \boldsymbol{\omega}_{\lambda,n} = \operatorname{Im}\left(\frac{(|n|^2+\lambda^2)\,\bar{q}-n}{\lambda^2|q|^2+|1-nq|^2}\, dq\right).$$

Remark: Any connection $\boldsymbol{\eta}$ on $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad{}^1$ is uniquely determined by either one of the gauge potentials ${s_1}^*\boldsymbol{\eta}$ or ${s_2}^*\boldsymbol{\eta}$ (page 339, [**N4**]) so it is customary to specify $\boldsymbol{\omega}_{\lambda,n}$ by giving the single potential

$$\boldsymbol{\mathcal{A}}_{\lambda,n} = \operatorname{Im}\left(\frac{\bar{q}-\bar{n}}{\lambda^2+|q-n|^2} dq\right).$$

$\boldsymbol{\mathcal{A}}_{\lambda,n}$ is called the **generic BPST potential** with **center** n and **scale** λ (page 357, [**N4**]).

The **curvature** $\boldsymbol{\Omega}$ of a connection $\boldsymbol{\omega}$ on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is its covariant exterior derivative, defined by having $d\boldsymbol{\omega}$ operate only on horizontal parts, i.e., for each $p \in P$ and all $\boldsymbol{v},\ \boldsymbol{w} \in T_p(P)$ we let

$$\boldsymbol{\Omega}_p\,(\boldsymbol{v},\boldsymbol{w}) = (d\boldsymbol{\omega})_p\,(\boldsymbol{v}^H,\boldsymbol{w}^H).$$

The **Cartan Structure Equation** (Theorem 6.2.1, [**N4**]) asserts that

$$\boldsymbol{\Omega} = d\boldsymbol{\omega} + \frac{1}{2}[\boldsymbol{\omega},\boldsymbol{\omega}],$$

where we use $[\boldsymbol{\omega},\boldsymbol{\eta}]$ to denote the wedge product $\boldsymbol{\omega}\wedge_\rho\boldsymbol{\eta}$ (page 13) in which $\rho:\mathcal{G}\times\mathcal{G}\longrightarrow\mathcal{G}$ is the pairing given by the Lie bracket ($\rho(A,B)=[A,B]$). This wedge product is easy to compute for matrix groups: First define $\boldsymbol{\omega}\wedge\boldsymbol{\eta}$ to be the wedge product $\boldsymbol{\omega}\wedge_{\rho'}\boldsymbol{\eta}$, where ρ' is the pairing given by matrix multiplication ($\rho'(A,B)=AB$). We will prove in Chapter 4 that $\boldsymbol{\omega}\wedge\boldsymbol{\eta}$ is just the matrix of ordinary -, -, or -valued 2-forms obtained by regarding $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$ as matrices of 1-forms and forming their matrix product with entries multiplied by the ordinary wedge product. Moreover, $[\boldsymbol{\omega},\boldsymbol{\eta}]=\boldsymbol{\omega}\wedge\boldsymbol{\eta}+\boldsymbol{\eta}\wedge\boldsymbol{\omega}$ so $[\boldsymbol{\omega},\boldsymbol{\omega}]=2\boldsymbol{\omega}\wedge\boldsymbol{\omega}$ and

$$\boldsymbol{\Omega} = d\boldsymbol{\omega} + \boldsymbol{\omega}\wedge\boldsymbol{\omega}.$$

If $s:V\longrightarrow\mathcal{P}^{-1}(V)$ is a local cross-section, then the pullback $s^*\boldsymbol{\Omega}$ is called the **local field strength** (in gauge s) and denoted $\boldsymbol{\mathcal{F}}$. Writing $\boldsymbol{\mathcal{A}}=s^*\boldsymbol{\omega}$ and $\boldsymbol{\mathcal{F}}=s^*\boldsymbol{\Omega}$ the Cartan Structure Equation becomes

$$\boldsymbol{\mathcal{F}} = d\,\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{A}}\wedge\boldsymbol{\mathcal{A}}$$

(page 350, [**N4**]). Assuming (as we may) that the domain V of s is also a coordinate neighborhood for a chart (V,φ) with coordinate functions $x^1,\ldots,x^n$, we can write $\boldsymbol{\mathcal{A}}=\mathcal{A}_\alpha dx^\alpha$ and $\boldsymbol{\mathcal{F}}=\frac{1}{2}\mathcal{F}_{\alpha\beta}dx^\alpha\wedge dx^\beta$, where the $\mathcal{A}_\alpha$ and $\mathcal{F}_{\alpha\beta}$ are $\mathcal{G}$-valued functions on V. Then

$$\mathcal{F}_{\alpha\beta} = \partial_\alpha\mathcal{A}_\beta - \partial_\beta\mathcal{A}_\alpha + [\mathcal{A}_\alpha,\ \mathcal{A}_\beta],$$

where we have written ∂_α for $\frac{\partial}{\partial x^\alpha}$ and these derivatives are computed componentwise in $\mathcal{G}$ (Exercise 6.2.9, [**N4**]). If $s_j:V_j\longrightarrow\mathcal{P}^{-1}(V_j)$ and $s_i:V_i\longrightarrow\mathcal{P}^{-1}(V_i)$ are two local cross-sections with $V_j\cap V_i\neq\emptyset$ and if $g_{ij}:V_j\cap V_i\longrightarrow G$ is the transition function relating the corresponding trivializations, then $s_j(x)=s_i(x)\cdot g_{ij}(x)$ for each $x\in V_j\cap V_i$ (Exercise 4.3.5, [**N4**]) and, whereas $\boldsymbol{\mathcal{A}}_j=g_{ij}^{-1}\boldsymbol{\mathcal{A}}_i g_{ij}+g_{ij}^{-1}dg_{ij}$, we have

$$\boldsymbol{\mathcal{F}}_j = g_{ij}^{-1}\boldsymbol{\mathcal{F}}_i g_{ij}.$$

We write out a few specific examples.

1. (Flat connections on trivial bundles) We let $G\hookrightarrow X\times G\xrightarrow{\mathcal{P}}X$ be a trivial bundle, $\boldsymbol{\Theta}$ the Cartan 1-form on G and $\pi:X\times G\longrightarrow G$ the projection. As in Example #1, page 38, $\boldsymbol{\omega}=\pi^*\boldsymbol{\Theta}$ is a connection form on $X\times G$. The Maurer-Cartan equations (page 19) yield (page 329, [**N4**]) the **equation of**

structure for G, i.e.,

$$d\boldsymbol{\Theta} + \frac{1}{2}[\boldsymbol{\Theta}, \boldsymbol{\Theta}] = 0,$$

and from this it follows that

$$d\boldsymbol{\omega} + \frac{1}{2}[\boldsymbol{\omega}, \boldsymbol{\omega}] = 0$$

(page 353, [**N4**]). Thus, flat connections $\boldsymbol{\omega}$ are "flat" because their curvature forms $\boldsymbol{\Omega}$ are identically zero.

2. (Natural connection on the complex Hopf bundle) Connections $\boldsymbol{\omega}$ on $U(1)$-bundles $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ have a number of very special properties and we begin with a general discussion of a few of these. We identify the Lie algebra $u(1)$ with the algebra Im of pure imaginary complex numbers. Since $u(1)$ is 1-dimensional, all brackets vanish so the curvature $\boldsymbol{\Omega}$ of any connection $\boldsymbol{\omega}$ coincides with its exterior derivative:

$$\boldsymbol{\Omega} = d\boldsymbol{\omega}.$$

If $s : V \longrightarrow \mathcal{P}^{-1}(V)$ is any cross-section, then we can write the gauge potential $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{\omega}$ and field strength $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ as

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= -\boldsymbol{i}\,\boldsymbol{A} \\ \boldsymbol{\mathcal{F}} &= d\,\boldsymbol{\mathcal{A}} = -\boldsymbol{i}d\boldsymbol{A} = -\boldsymbol{i}\boldsymbol{F}, \end{aligned}$$

where $\boldsymbol{A}$ and $\boldsymbol{F}$ are real-valued forms on V (the minus signs are conventional). If $s_j : V_j \longrightarrow \mathcal{P}^{-1}(V_j)$ and $s_i : V_i \longrightarrow \mathcal{P}^{-1}(V_i)$ are two local cross-sections with $V_j \cap V_i \neq \emptyset$ and if $g_{ij} : V_j \cap V_i \longrightarrow U(1)$ is the corresponding transition function, then

$$\boldsymbol{\mathcal{A}}_j = g_{ij}^{-1}\boldsymbol{\mathcal{A}}_i g_{ij} + g_{ij}^{-1}dg_{ij} = \boldsymbol{\mathcal{A}}_i + g_{ij}^{-1}\ dg_{ij}$$

and

$$\boldsymbol{\mathcal{F}}_j = g_{ij}^{-1}\boldsymbol{\mathcal{F}}_i g_{ij} = \boldsymbol{\mathcal{F}}_i$$

on $V_j \cap V_i$ because $U(1)$ is Abelian. In particular, the local field strengths, since they agree on any intersections of their domains, piece together to give a *globally defined field strength* 2-form $\boldsymbol{\mathcal{F}}$ on X. This is a peculiarity of Abelian gauge fields and generally is not true in the non-Abelian case. Also note that since g_{ij} maps into $U(1)$ it can be written as $g_{ij}(x) = e^{-\boldsymbol{i}\Lambda(x)}$ for some real-valued function Λ on $V_j \cap V_i$. Then $g_{ij}^{-1}dg_{ij} = e^{\boldsymbol{i}\Lambda}e^{-\boldsymbol{i}\Lambda}(-\boldsymbol{i}d\Lambda) = -\boldsymbol{i}d\Lambda$ so $\boldsymbol{\mathcal{A}}_j = \boldsymbol{\mathcal{A}}_i - \boldsymbol{i}d\Lambda$, i.e.,

$$\boldsymbol{A}_j = \mathbf{A}_i + d\Lambda,$$

which is the traditional form for the relationship between two "vector potentials."

Now we return to the case of the natural connection $\boldsymbol{\omega}$ on the complex Hopf bundle $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}} \quad {}^1$. Identifying $\quad {}^1$ with S^2 as indicated on page 39 we obtain $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_1} S^2$ and local gauge potentials $\boldsymbol{\mathcal{A}}_N$ and $\boldsymbol{\mathcal{A}}_S$ which, in spherical coordinates, are given by

$$\begin{aligned} \boldsymbol{\mathcal{A}}_N &= -\frac{1}{2}\boldsymbol{i}(1-\cos\phi)d\theta \\ \boldsymbol{\mathcal{A}}_S &= \ \frac{1}{2}\boldsymbol{i}(1+\cos\phi)d\theta. \end{aligned}$$

Each has the property that, on its domain, the field strength is given by

$$\boldsymbol{\mathcal{F}} = -\frac{1}{2}\boldsymbol{i}\sin\phi\ d\phi \wedge d\theta$$

(just compute $d\boldsymbol{\mathcal{A}}_N$ and $d\boldsymbol{\mathcal{A}}_S$).

Remarks: The real 1-forms $\boldsymbol{A}_N$ and $\boldsymbol{A}_S$ defined by $\boldsymbol{\mathcal{A}}_N = -\boldsymbol{i}\,\boldsymbol{A}_N$ and $\boldsymbol{\mathcal{A}}_S = -\boldsymbol{i}\,\boldsymbol{A}_S$ are the vector potentials for a Dirac magnetic monopole of strength $g = \frac{1}{2}$ (pages 19–20, [**N4**]) and the real 2-form $\boldsymbol{F}$ defined by $\boldsymbol{\mathcal{F}} = -\boldsymbol{i}\boldsymbol{F}$ is the corresponding magnetic field strength. For any integer n one can define Im -valued 1-forms on U_N and U_S by

$$\begin{aligned} \boldsymbol{\mathcal{A}}_N &= -\frac{1}{2}n\boldsymbol{i}(1-\cos\phi)d\theta, \quad \text{and} \\ \boldsymbol{\mathcal{A}}_S &= \ \frac{1}{2}n\boldsymbol{i}(1+\cos\phi)d\theta. \end{aligned}$$

We will investigate these 1-forms in some detail a bit later. For the time being we point out only that, while they are, indeed, gauge potentials for a connection whose field strength represents a Dirac monopole of strength $n/2$, this connection does not live on the Hopf bundle (unless $n = 1$), but on some other $U(1)$-bundle over S^2. That this must be the case is clear from $\boldsymbol{\mathcal{A}}_N = \boldsymbol{\mathcal{A}}_S - n\boldsymbol{i}d\theta = e^{n\theta\boldsymbol{i}}\boldsymbol{\mathcal{A}}_S e^{-n\theta\boldsymbol{i}} + e^{n\theta\boldsymbol{i}}de^{-n\theta\boldsymbol{i}}$, whereas the transition function for $U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_1} S^2$ is $g_{SN}(\phi,\ \theta) = e^{-\theta\boldsymbol{i}}$ ((0.3.8), [**N4**]). We will eventually show that the field strength 2-form $\boldsymbol{F}$ defines what is called the 1st Chern class of the bundle on which $\boldsymbol{\mathcal{A}}_N$ and $\boldsymbol{\mathcal{A}}_S$ determine a connection and that this characteristic class completely determines the principal $U(1)$-bundles over S^2. We will find also that the integral of this Chern class over S^2 is the magnetic charge n of the monopole.

3. (Instantons on the quaternionic Hopf bundle) We consider the $Sp(1)$-bundle $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \quad {}^1$ and, for each $\lambda > 0$ and $n \in$, the connection

$\boldsymbol{\omega}_{\lambda,n}$ on it determined by the potential

$$\boldsymbol{\mathcal{A}}_{\lambda,n} = \mathrm{Im}\left(\frac{\bar{q}-\bar{n}}{\lambda^2+|q-n|^2}dq\right)$$

(see the Remark on page 42). A calculation (pages 327–328, [**N4**]) gives

$$\boldsymbol{\mathcal{F}}_{\lambda,n} = d\,\boldsymbol{\mathcal{A}}_{\lambda,n} + \frac{1}{2}\left[\boldsymbol{\mathcal{A}}_{\lambda,n},\, \boldsymbol{\mathcal{A}}_{\lambda,n}\right] = \frac{\lambda^2}{(\lambda^2+|q-n|^2)^2}d\,\bar{q}\wedge dq$$

on . In Chapter 4 we will find that a natural inner product on Im , together with the usual metric on $= \ ^4$, gives rise to a norm for field strengths $\boldsymbol{\mathcal{F}}$ such as these. We will then define the Yang-Mills action of such a potential $\boldsymbol{\mathcal{A}}$ by

$$\mathcal{YM}\ (\boldsymbol{\mathcal{A}})\ = \frac{1}{2}\int_{\ ^4}\|\boldsymbol{\mathcal{F}}\|^2$$

and calculate to show that, for any $\lambda > 0$ and $n \in$,

$$\mathcal{YM}\ (\boldsymbol{\mathcal{A}}_{\lambda,n}) = 8\pi^2.$$

Remarks: All of the connections $\boldsymbol{\omega}_{\lambda,n}$ on $Sp(1) \hookrightarrow S^7 \xrightarrow{\mathcal{P}} \mathrm{H}\ ^1$ give rise to potentials $\boldsymbol{\mathcal{A}}_{\lambda,n}$ with the same Yang-Mills action $\mathcal{YM}(\boldsymbol{\mathcal{A}}_{\lambda,n})$ and this, we will find, is no accident. The number $\frac{1}{8\pi^2}\mathcal{YM}(\boldsymbol{\mathcal{A}}_{\lambda,n})$ is, in fact, a topological characteristic of the quaternionic Hopf bundle, not unlike the Euler characteristic of a surface. We will eventually show that this number is essentially the integral over $\ ^1$ of what is called the 2^{nd} Chern class of the bundle. Like the magnetic charge of a Dirac monopole, the Yang-Mills action for an instanton can be thought of as a sort of "topological charge."

1.5 Associated Bundles and Matter Fields

Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal G-bundle over X with right action $\sigma : P \times G \longrightarrow P,\ \sigma(p,\ g) = p \cdot g$. Now let F be a smooth manifold on which G acts smoothly on the left (the image of $(g,\ \xi) \in G \times F$ under this action will be written $g \cdot \xi$). Then $((p,\xi),g) \longrightarrow (p,\xi)\cdot g = (p\cdot g, g^{-1}\cdot\xi)$ is a smooth right action of G on $P \times F$ (Exercise 6.7.1, [**N4**]). We denote by $P \times_G F$ the orbit space of $P \times F$ by this action. More precisely, we define an equivalance relation $\sim$ on $P \times F$ as follows: $(p_1,\xi_1) \sim (p_2,\xi_2)$ if and only if there exists a $g \in G$ such that $(p_2,\xi_2) = (p_1,\xi_1)\cdot g$. The equivalence class containing (p,ξ) is $[p,\xi] = \{(p\cdot g, g^{-1}\cdot\xi) : g \in G\}$. Then, as a set, $P \times_G F = \{[p,\xi] : (p,\xi) \in P \times F\}$ and we provide $P \times_G F$ with the quotient topology determined by $\mathcal{Q} : P \times F \longrightarrow P \times_G F, \mathcal{Q}(p,\ \xi) = [p,\xi]$. Also define $\mathcal{P}_G : P \times_G F \longrightarrow X$ by $\mathcal{P}_G([p,\xi]) = \mathcal{P}(p)$. Then $\mathcal{P}_G$ is continuous and, for any $x \in X, \mathcal{P}_G^{-1}(x) = \{[p,\xi] : \xi \in F\}$, where p is any point in $\mathcal{P}^{-1}(x)$ (Exercise 6.7.2, [**N4**]). If (V,Ψ)

is any local trivialization of $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ and $s : V \longrightarrow \mathcal{P}^{-1}(V)$ is the associated cross-section, then the map $\tilde{\Phi} : V \times F \longrightarrow P_G^{-1}(V)$ defined by $\tilde{\Phi}(x, \xi) = [s(x), \xi]$ is a homeomorphism with inverse $\tilde{\Psi} : \mathcal{P}_G^{-1}(V) \longrightarrow V \times F$ given by $\tilde{\Psi}([s(x), \xi]) = (x, \xi)$ (page 381, [**N4**]). If (V_i, Ψ_i) and (V_j, Ψ_j) are two such trivializations with $V_i \cap V_j \neq \emptyset$ and $g_{ji} : V_i \cap V_j \longrightarrow G$ is the corresponding transition function, then $\tilde{\Psi}_j \circ \tilde{\Psi}_i^{-1} : (V_i \cap V_j) \times F \longrightarrow (V_i \cap V_j) \times F$ is given by $\tilde{\Psi}_j \circ \tilde{\Psi}_i^{-1}(x, \xi) = (x,\ g_{ji}(x) \cdot \xi)$ and so is a diffeomorphism (page 382, [**N4**]). It follows that there is a unique differentiable structure on $P \times_G F$ relative to which each $\tilde{\Psi} : \mathcal{P}_G^{-1}(V) \longrightarrow V \times F$ is a diffeomorphism and that, relative to this structure, $\mathcal{P}_G : P \times_G F \longrightarrow X$ is smooth. We call

$$\mathcal{P}_G : P \times_G F \longrightarrow X$$

the **fiber bundle associated with** $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ by the given left action of G on F.

The special case of most interest to us arises as follows: Let $F = \mathcal{V}$ be a finite dimensional vector space (with its natural differentiable structure) and $\rho : G \longrightarrow GL(\mathcal{V})$ a smooth representation of G on $\mathcal{V}$. Then ρ gives rise to a smooth left action of G on $\mathcal{V}$ $((g, v) \longrightarrow g \cdot v = (\rho(g))(v))$. The fiber bundle associated with $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ by this action is denoted

$$\mathcal{P}_\rho : P \times_\rho \mathcal{V} \longrightarrow X$$

and called the **vector bundle associated with** $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ by the representation ρ. In this case each fiber $\mathcal{P}_\rho^{-1}(x) = \{[p,\ v] : v \in \mathcal{V}\}$, where p is any point in $\mathcal{P}^{-1}(x)$, is a copy of $\mathcal{V}$ and admits a natural vector space structure: $a_1[p, v_1] + a_2[p, v_2] = [p, a_1 v_1 + a_2 v_2]$ for all $a_1, a_2 \in$ and $v_1, v_2 \in \mathcal{V}$. We record a few examples of particular interest.

1. Let $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be an arbitrary principal $U(1)$-bundle and take $\mathcal{V} =$ (as a 2-dimensional real vector space). If $\rho : U(1) \longrightarrow GL(\ \ )$ is any representation of $U(1)$ on , then the associated vector bundle $\mathcal{P}_\rho : P \times_\rho \quad \longrightarrow X$ has fibers that are copies of   and is called a **complex line bundle** over X. An obvious choice for $\rho : U(1) \longrightarrow GL(\ \ )$ is obtained by taking $(\rho(g))(z) = gz$ for each $g \in U(1)$ and $z \in$ (if $g = e^{i\theta}, 0 \leq \theta < 2\pi$, then $\rho(g)$ is rotation by θ). More generally, one can define, for any integer n, a representation $\rho : U(1) \longrightarrow GL(\ \ )$ by $(\rho(g))(z) = g^n z$ and thereby an associated vector bundle over X.

2. Let $Sp(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be any principal $Sp(1)$-bundle and take $\mathcal{V} =$ (as a 4-dimensional real vector space). Then any representation $\rho : Sp(1) \longrightarrow GL(\ \ )$ (e.g., $(\rho(g))(q) = gq$, or $(\rho(g))(q) = g^n q$ for some integer n) defines a vector bundle $P \times_\rho$ with fibers isomorphic to   and called a **quaternionic line bundle** over X.

3. Let $SU(2) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be any principal $SU(2)$-bundle and take $\mathcal{V} = \ {}^2$ (as a 4-dimensional real vector space). Then any representation $\rho : SU(2) \longrightarrow GL(\ {}^2)$ defines an associated vector bundle $P \times_\rho \ {}^2$. A particularly useful choice for ρ is obtained as follows (we write the elements of ^{2} as column vectors $\left(\begin{smallmatrix} \xi^1 \\ \xi^2 \end{smallmatrix}\right)$): For each $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in SU(2)$ define

$$(\rho(g))\begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix} = g\begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix} = \begin{pmatrix} \alpha\xi^1 + \beta\xi^2 \\ \gamma\xi^1 + \delta\xi^2 \end{pmatrix}.$$

This representation of $SU(2)$ is generally denoted $D^{\frac{1}{2}}$ and we will find (in Section 2.4) that it arises naturally in Pauli's nonrelativistic theory of the electron.

4. Let G be an arbitrary matrix Lie group and $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be an arbitrary principal G-bundle. The adjoint representation $ad : G \longrightarrow GL(\mathcal{G})$ assigns to each $g \in G$ the nonsingular linear transformation ad_g on the Lie algebra $\mathcal{G}$ defined by

$$ad_g(A) = g\,A\,g^{-1}$$

(page 24). The vector bundle associated with $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ by ad is called the **adjoint bundle** of $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ and denoted

$$ad\ P = P \times_{ad} \mathcal{G}.$$

The fibers of $ad\ P$ are copies of the Lie algebra $\mathcal{G}$ of G.

Remark: The significance of the adjoint bundle, which is considerable, will emerge somewhat later. Roughly, the situation is as follows: The curvature $\boldsymbol{\Omega}$ of a connection on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is a $\mathcal{G}$-valued 2-form that is *globally defined on* P. However, the field strengths $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ on X generally do not piece together into a globally defined $\mathcal{G}$-valued 2-form on X (page 44). Physically, this is unfortunate since one would like field strengths to live naturally on space (or spacetime) and not on some abstract space of internal states. Mathematically, it is inconvenient since connections on bundles and their curvatures are often only auxiliary devices for studying the topology and geometry of the base manifold (see, for example, the discussion of Donaldson's Theorem in Appendix B of [**N4**]). In any case, we will find that by allowing 2-forms to take values, not in a single, fixed copy of $\mathcal{G}$, but in the "parametrized family of $\mathcal{G}$'s" that the adjoint bundle represents, we will be able to uniquely represent the field strength as a globally defined $ad\,P$-valued 2-form on X. More generally, we will find that many locally defined $\mathcal{V}$-valued objects of interest (e.g., wavefunctions) become globally defined when thought of as taking their values in some associated vector bundle.

Now, let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be an arbitrary principal G-bundle, F a smooth manifold on which G acts on the left and $\mathcal{P}_G : P \times_G F \longrightarrow X$ the associated fiber bundle. If V is an open subset of X, then a smooth map $\phi : \mathcal{P}^{-1}(V) \longrightarrow F$ is said to be **equivariant** (with respect to the given actions of G on P and F) if

$$\phi(p \cdot g) = g^{-1} \cdot \phi(p)$$

for all $p \in \mathcal{P}^{-1}(V)$ and $g \in G$. Given such a map one defines $s_\phi : V \longrightarrow \mathcal{P}_G^{-1}(V)$ by

$$s_\phi(x) = [p,\ \phi(p)],$$

where p is any point in $\mathcal{P}^{-1}(x)$. Then s_ϕ is smooth and satisfies $\mathcal{P}_G \circ s_\phi = id_V$. Conversely, suppose $s : V \longrightarrow \mathcal{P}_G^{-1}(V)$ is a smooth map satisfying $\mathcal{P}_G \circ s = id_V$ (called a local **cross-section** of the fiber bundle $\mathcal{P}_G : P \times_G F \longrightarrow X$). Define $\phi_s : \mathcal{P}^{-1}(V) \longrightarrow F$ as follows: Let $p \in \mathcal{P}^{-1}(V)$. Then $\mathcal{P}(p) = x$ is in V so $s(x)$ is in $\mathcal{P}_G^{-1}(V)$ and there exists a unique element $\phi_s(p) \in F$ such that

$$s(x) = [p,\ \phi_s(p)].$$

Then ϕ_s is a smooth equivariant map (page 384, [**N4**]). Moreover, this correspondence between equivariant maps $\phi : \mathcal{P}^{-1}(V) \longrightarrow F$ and cross-sections $s : V \longrightarrow \mathcal{P}_G^{-1}(V)$ of the associated fiber bundle is one-to-one and onto (Exercise 6.8.4, [**N4**]). All of this applies, in particular, to the special case of a vector bundle $\mathcal{P}_\rho : P \times_\rho \mathcal{V} \longrightarrow X$ associated to $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ by some representation $\rho : G \longrightarrow GL(\mathcal{V})$ of G on $\mathcal{V}$.

Remarks: Our interest in equivariant maps arises from the following considerations. A connection $\boldsymbol{\omega}$ on a principal bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is what the physicists call a gauge field and is thought of as something akin to a modern version of the Newtonian concept of a "force" in that particles "respond" to it by experiencing changes in their internal states. The particles coupled to (i.e., experiencing the effects of) the gauge field have wavefunctions ψ on X taking values in some real vector space $\mathcal{V}$ (e.g., , ^{2} etc.) that are obtained by solving various partial differential equations involving the gauge potentials $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{\omega}$. Since the gauge potentials $\boldsymbol{\mathcal{A}}$ are, in general, only locally defined on X, the same is true of these wavefunctions. A change of gauge ($s \longrightarrow s \cdot g$) gives rise to a new potential $\boldsymbol{\mathcal{A}}^g = ad_{g^{-1}} \circ \boldsymbol{\mathcal{A}} + g^*\boldsymbol{\Theta}$, a new field strength $\boldsymbol{\mathcal{F}}^g = ad_{g^{-1}} \circ \boldsymbol{\mathcal{F}}$ and thereby a new wavefunction $\psi^g = (\rho(g^{-1}))\psi = g^{-1} \cdot \psi$, where ρ is some representation of G on $\mathcal{V}$ that is characteristic of the particular class of particles under consideration. Now, we can identify ψ (in gauge s) with a local cross-section of the associated vector bundle $P \times_\rho \mathcal{V}$:

$$x \longrightarrow [s(x),\ \psi(x)].$$

Then, in gauge $s \cdot g$,

$$x \longrightarrow [s(x) \cdot g(x),\ g(x)^{-1} \cdot \psi(x)] = [s(x),\ \psi(x)]$$

so that the locally defined wavefunctions piece together to determine a *globally* defined cross-section of $P \times_\rho \mathcal{V}$. But a globally defined cross-section of $P \times_\rho \mathcal{V}$ correspondings to a globally defined equivariant map $\phi : P \longrightarrow \mathcal{V}$. As $\mathcal{V}$-valued maps on X, the wavefunctions are only locally defined, but all of these local wavefunctions on X can be uniquely represented by a single, globally defined $\mathcal{V}$-valued map on P (or a single, globally defined cross-section of the associated vector bundle). With this as motivation we introduce the following definitions.

Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a principal G-bundle, $\mathcal{V}$ a real vector space and $\rho : G \longrightarrow GL(\mathcal{V})$ a representation of G on $\mathcal{V}$. An equivariant, $\mathcal{V}$-valued map $\phi : P \longrightarrow \mathcal{V}$ on P is called a **matter field** (of **type** ρ) on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$. If $\mathcal{V} =$, then ϕ is called a **complex scalar field**. If $\mathcal{V} = \mathcal{G}$ and $\rho = ad$ is the adjoint representation, then ϕ is called a **Higgs field**.

A matter field ϕ is, in particular, a $\mathcal{V}$-valued 0-form on P and so has an exterior derivative $d\phi$. Assuming now that $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ has defined on it a connection form $\boldsymbol{\omega}$, we define the **covariant exterior derivative** $d^{\boldsymbol{\omega}}\phi$ of ϕ by having $d\phi$ act only on horizontal parts: For each $p \in P$ and $\boldsymbol{v} \in T_p(P)$,

$$(d^{\boldsymbol{\omega}}\phi)_p(\boldsymbol{v}) = (d\phi)_p(\boldsymbol{v}^H).$$

This is a $\mathcal{V}$-valued 1-form on P (Exercise 6.8.5, [**N4**]) and satisfies

$$\sigma_g^*(d^{\boldsymbol{\omega}}\phi) = g^{-1} \cdot d^{\boldsymbol{\omega}}\phi$$

for each $g \in G$ ((6.8.1), [**N4**]). These are the derivatives that appear in the field equations describing the quantitative response of the particle to the gauge field (pages 391–392, [**N4**]). A computational formula analogous to the Cartan Structure Equation for curvature $\boldsymbol{\Omega}$ (which is the covariant exterior derivative of the connection form $\boldsymbol{\omega}$) is obtained as follows: For any $A \in \mathcal{G}$ and $v \in \mathcal{V}$ we define $A \cdot v \in \mathcal{V}$ by

$$A \cdot v = \frac{d}{dt}(\exp(tA) \cdot v)|_{t=0} = \frac{d}{dt}(\rho(\exp(tA))(v))|_{t=0}.$$

Remarks: Two special cases are worth pointing out immediately. If G is a group of $n \times n$ matrices (with entries in $=$, , or), $\mathcal{V} = {}^n$ (thought of as column matrices) and ρ is the natural representation of G on $\mathcal{V}$ (matrix multiplication), then, identifying $\mathcal{G}$ with an algebra of matrices, $A \cdot v = Av$ (matrix multiplication). On the other hand, if $\mathcal{V} = \mathcal{G}$ and $\rho = ad$, then, for all $A, B \in \mathcal{G}$, $A \cdot B = [A,\ B]$ (see the proof of Theorem 5.8.8, [**N4**]).

Now, if ϕ is a $\mathcal{V}$-valued 0-form on P and $\boldsymbol{\omega}$ is a $\mathcal{G}$-valued 1-form on P we can define a $\mathcal{V}$-valued 1-form $\boldsymbol{\omega} \cdot \phi$ on P by

$$(\boldsymbol{\omega} \cdot \phi)_p(\boldsymbol{v}) = \boldsymbol{\omega}_p(\boldsymbol{v}) \cdot \phi(p)$$

for each $p \in P$ and $\boldsymbol{v} \in T_p(P)$. Then

$$d^{\boldsymbol{\omega}}\phi = d\phi + \boldsymbol{\omega} \cdot \phi$$

((6.8.4), [**N4**]). We conclude by writing out a few concrete examples in local coordinates (details are available on pages 388–391, [**N4**]).

1. Let $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a principal $U(1)$-bundle with connection $\boldsymbol{\omega}$ and let $\mathcal{V} = \quad$ (as a 2-dimensional real vector space). For each integer q we let $\rho_q : U(1) \longrightarrow GL(\quad)$ be the representation $\rho_q(g)(z) = g^q z$. For simplicity, we will write $\phi = \phi(x^1, \ldots, x^n)$ for a local coordinate expression of the pullback by some cross-section of a matter field. Similarly the gauge potential will be written locally as $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}(x^1, \ldots, x^n) = \mathcal{A}_\alpha(x^1, \ldots, x^n)dx^\alpha = -\boldsymbol{i}A_\alpha(x^1, \ldots, x^n)dx^\alpha$. The corresponding local coordinate expression for the pullback of $d^{\boldsymbol{\omega}}\phi$ is given by

$$(\partial_\alpha \phi + q\mathcal{A}_\alpha \phi)dx^\alpha = (\partial_\alpha - \boldsymbol{i}qA_\alpha)\phi dx^\alpha,$$

where $\partial_\alpha = \frac{\partial}{\partial x^\alpha}$.

2. Let $SU(2) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a principal $SU(2)$-bundle with connection $\boldsymbol{\omega}$ and let $\mathcal{V} = \ ^2$ (as a 4-dimensional real vector space). For each integer q we let $\rho_q : SU(2) \longrightarrow GL(\ ^2)$ be the representation $\rho_q(g)\begin{pmatrix} z^1 \\ z^2 \end{pmatrix} = g^q \begin{pmatrix} z^1 \\ z^2 \end{pmatrix}$. Writing $\phi = \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix} = \begin{pmatrix} \phi^1(x^1,\ldots,x^n) \\ \phi^2(x^1,\ldots,x^n) \end{pmatrix}$ for the local coordinate expression of the pullback by some cross-section of a matter field and $\boldsymbol{\mathcal{A}} = \mathcal{A}_\alpha(x^1, \ldots, x^n)dx^\alpha = -\boldsymbol{i}\mathcal{B}_\alpha(x^1, \ldots, x^n)dx^\alpha$ for the local gauge potential (where the $\mathcal{A}_\alpha$ are skew-Hermitian and tracefree, while the $\mathcal{B}_\alpha$ are Hermitian and tracefree; Exercise 6.8.5, [**N4**]), we have

$$\left(\partial_\alpha \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix} + q\mathcal{A}_\alpha \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix}\right) dx^\alpha = (\partial_\alpha - \boldsymbol{i}q\mathcal{B}_\alpha) \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix} dx^\alpha$$

for the corresponding local coordinate expression for the pullback of $d^{\boldsymbol{\omega}}\phi$.

2
Physical Motivation

2.1 General Framework for Classical Gauge Theories

Our objective now is to use the machinery assembled in the previous chapter, and that to be developed in subsequent chapters, to study a number of rather specific "classical gauge theories" arising in modern physics. The discussion is heuristic and informal and its intention is to indicate how the topology and geometry that are our real concern here arise naturally in meaningful physics. In order to have a context within which to place these examples we will devote this rather brief section to an explicit enumeration of the basic mathematical ingredients required to describe, at the classical level, the interaction of a particle with a gauge field.

1. A smooth, oriented, (semi-) Riemannian manifold X.

Generally, this will be space (^{3}), a spacetime (e.g., ^{1,3}; see Section 2.2), a Euclidean ("Wick rotated") version of a spacetime (e.g., ^{4}), a compactification of one of these (e.g., $S^4 = \ ^4 \cup \{\infty\}$), or an open submanifold of one of these. The particles "live" in X.

2. A finite dimensional vector space $\mathcal{V}$.

The particles have wavefunctions that take values in $\mathcal{V}$. The choice of $\mathcal{V}$ is dictated by the internal structure of the particle (e.g., phase, isospin, spin, etc.) and so $\mathcal{V}$ is called the **internal space**. Typical examples are , ^{2}, ^{4}, or the Lie algebra of some Lie group, e.g., $u(1)$, or $su(2)$. $\mathcal{V}$ is equipped with an inner product $\langle\ ,\ \rangle$ by which one computes squared norms $\|\psi\|^2$ and thereby the probabilities with which quantum mechanics deals.

3. A matrix Lie group G and a representation

$$\rho : G \longrightarrow GL(\mathcal{V})$$

of G on $\mathcal{V}$ that is orthogonal with respect to the inner product on $\mathcal{V}$:

$$\langle\, \rho(g)(v), \rho(g)(w)\,\rangle = \langle\, v, w\,\rangle.$$

This will generally be one of the classical groups (e.g., $U(1)$, $SU(2)$) or a product of these and plays a dual role.

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0_2,

(i) The inner product on $\mathcal{V}$ determines a class of orthonormal bases, or "frames" (e.g., isospin axes) and these are related by the elements of G, i.e., $g \in G$ "acts" on a frame p to give a new frame $p \cdot g$. By fixing some frame at the outset (the elements of which will correspond to certain "states" of the particle) one can identify the elements of G with the frames.

(ii) G also acts on $\mathcal{V}$ via the representation $\rho(v \longrightarrow \rho(g)(v) = g \cdot v)$ and so acts on the wavefunction ψ at each point. If $\psi(p)$ is a value of the wavefunction, described relative to the frame p, then

$$\psi\,(p \cdot g) = g^{-1} \cdot \psi\,(p)$$

is its description relative to the new frame $p \cdot g$.

4. A smooth principal G-bundle over X:

$$G \hookrightarrow P \xrightarrow{\mathcal{P}} X.$$

Typical examples are trivial bundles (e.g., $SU(2) \hookrightarrow \;{}^4 \times SU(2) \longrightarrow \;{}^4$) and Hopf bundles (e.g., $U(1) \hookrightarrow S^3 \longrightarrow S^2$, $SU(2) \hookrightarrow S^7 \longrightarrow S^4$). At each $x \in X$ the fiber $\mathcal{P}^{-1}(x)$ is a copy of G thought of as the set of all frames in the internal space $\mathcal{V}$ at x. A local cross-section $s : V \longrightarrow P$ is a smooth selection of a frame at each point in some open subset V of X (a local "gauge") relative to which wavefunctions can be described on V.

5. A connection $\boldsymbol{\omega}$ on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ with curvature $\boldsymbol{\Omega}$.

If $s : V \longrightarrow P$ is a local cross-section (local gauge), then the pullback $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{\omega}$ is the **local gauge potential** and $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ is the **local field strength**. Generally, these exist only locally since nontrivial principal bundles do not admit global cross-sections and pullbacks by different local cross-sections usually do not agree on the intersection of their domains. Particles coupled to (i.e., experiencing the effects of) the field determined by $\boldsymbol{\omega}$ have locally defined wavefunctions ψ taking values in $\mathcal{V}$ that are obtained by solving equations of motion (see #8 below) involving the local potentials $\boldsymbol{\mathcal{A}}$. A change of gauge $(s \longrightarrow s \cdot g)$ changes the wavefunction by the representation ρ $(\psi \longrightarrow g^{-1} \cdot \psi)$. These local wavefunctions piece together into a globally defined object called a **matter field** which can be described in two equivalent ways:

6. A global cross-section of the vector bundle $P \times_\rho \mathcal{V}$ associated to $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ by ρ (equivalently, a $\mathcal{V}$-valued map $\phi : P \longrightarrow \mathcal{V}$ on P that is equivariant: $\phi(p \cdot g) = g^{-1} \cdot \phi(p)$).

Physically, such a matter field has potential energy which we describe with

7. A non-negative, smooth, real-valued function

$$U : \mathcal{V} \longrightarrow$$

that is invariant under the action of G on $\mathcal{V}$:

$$U\,(g \cdot v) = U\,(v).$$

U is to be regarded as a potential function with $U \circ \phi$ describing the self-interaction energy of the matter field ϕ. Typically, this will depend only on $\|\,\phi\,\|^2$, e.g., $\frac{1}{2}m\,\|\phi\,\|^2$, or $\frac{\lambda}{8}(\|\phi\,\|^2 - 1)^2$, where m and λ are non-negative constants.

8. An action (energy) functional $A(\boldsymbol{\omega}, \phi)$, the stationary points of which are the physically significant field configurations $(\boldsymbol{\omega}, \phi)$.

Typically, this functional is of the following general form:

$$A\,(\boldsymbol{\omega}, \phi) = c \int_X \Big[\, \|F_{\boldsymbol{\omega}}\,\|^2 + c_1\,\|\,d^{\boldsymbol{\omega}}\phi\,\|^2 + c_2\,U \circ \phi \Big].$$

We will spell out in detail what each of these terms means in the concrete examples to follow. Briefly, c is a normalizing constant, c_1 and c_2 are "coupling constants," $F_{\boldsymbol{\omega}}$ is a global 2-form on X with values in the ad-joint bundle ad P which locally pulls back to the gauge field strengths $\boldsymbol{\mathcal{F}}, d^{\boldsymbol{\omega}}\phi$ is the covariant exterior derivative of the matter field ϕ (thought of as a cross-section of the associated vector bundle) and the norms arise from the metric on X and the Killing form on the Lie algebra $\mathcal{G}$ of G. Integrals of such objects over manifolds like X will be introduced in Chapter 4. The physically interesting field configurations $(\boldsymbol{\omega}, \phi)$, are those which (at least locally) minimize the value of the action functional. The Calculus of Variations provides necessary conditions (the Euler-Lagrange differential equations) that must be satisfied by such minima. The Euler-Lagrange equations for the action $A(\boldsymbol{\omega}, \phi)$, are the appropriate field equations (the "equations of motion") of our gauge theory. One can, of course, generalize the model we have described by including more than one matter field.

From the point-of-view of physics, one is generally interested only in **finite action** configurations $(\boldsymbol{\omega}, \phi)$, i.e., those for which

$$A\,(\boldsymbol{\omega}, \phi) < \infty.$$

This is assured if X is compact, but otherwise one must assume some sort of appropriate asymptotic behavior for the terms in the integrand. Such asymptotic conditions turn out to have profound topological consequences (e.g., the existence of "topological charge") and investigating this link between topology and asymptotics is one of our primary objectives.

These then are the basic ingredients required to build a classical gauge theory. There is a special case that has gotten a great deal of attention, particularly in the mathematical community. When $c_1 = c_2 = 0$ in the action $A\,(\boldsymbol{\omega}, \phi)$, (so that, effectively, there are no matter fields present), then one can think of A as depending only on $\boldsymbol{\omega}$ and, in this case, it is referred to as the

Yang-Mills action and written

$$\mathcal{YM}(\boldsymbol{\omega}) = c \int_X \| F_{\boldsymbol{\omega}} \|^2.$$

The Euler-Lagrange equations for $\mathcal{YM}$ are called the **Yang-Mills equations** and can be written

$$d^{\boldsymbol{\omega}} {}^*F_{\boldsymbol{\omega}} = 0,$$

where ${}^*F_{\boldsymbol{\omega}}$ is the Hodge dual of $F_{\boldsymbol{\omega}}$ and $d^{\boldsymbol{\omega}}$ is the covariant exterior derivative.

Remark: We have not yet defined the Hodge dual or the covariant exterior derivative in sufficient generality to cover the context in which we now find ourselves, but the generalization is easy and will be provided in Chapter 4. We will also find that, quite independently of the action, the field $F_{\boldsymbol{\omega}}$ also satisfies a purely geometrical constraint known as the **Bianchi identity**

$$d^{\boldsymbol{\omega}} F_{\boldsymbol{\omega}} = 0.$$

These last two equations lie at the heart of what is called pure **Yang-Mills theory**. The impact of this subject on low dimensional topology is discussed at some length in [**N4**]. Although this special case may not appear to be in the spirit of our announced intention here to model interactions we will find that it leads (through a process known as "dimensional reduction") directly to the particular interactions of most interest to us in Section 2.5.

2.2 Electromagnetic Fields

The prototypical example of a gauge theory is classical electrodynamics. Although our real interests lie elsewhere, this example will provide a nice warm-up in familiar territory and so we will describe it in some detail. Keep in mind that our intention in this chapter is primarily motivational so we will feel free to adopt a rather casual attitude, occasionally anticipating concepts and results that are introduced carefully only later in the text.

The arena within which electrodynamics is done is **Minkowski spacetime** ^{1,3} (assuming gravitational effects are neglected). As a differentiable manifold, ^{1,3} is just ^{4}. Rather than the usual Riemannian metric on ^{4}, however, we define on ^{1,3} the semi-Riemannian metric given, relative to standard coordinates x^0, x^1, x^2, x^3 on ^{4} by $\eta_{\alpha\beta} dx^\alpha \otimes dx^\beta$, where

$$\eta_{\alpha\beta} = \begin{cases} 1, & \alpha = \beta = 0 \\ -1, & \alpha = \beta = 1, 2, 3. \\ 0, & \alpha \neq \beta \end{cases}$$

Remark: For the moment we will require very little of the geometry of ^{1,3} and its physical significance. Simply think of the elements of ^{1,3} as "events" whose standard coordinates represent the time (x^0) and spatial (x^1, x^2, x^3) coordinates by which the event is identified in some fixed, but arbitrary inertial frame of reference. The entire history of a (point) object can then be identified with a continuous sequence of events (i.e., a curve) in ^{1,3} called its "worldline." Finally, since the differentiable structure of ^{4} is just its natural structure as a real vector space (Example #3, page 4), each tangent space $T_p(\ ^{1,3})$ is canonically identified with ^{4} itself. Since the components of the semi-Riemannian metric we have introduced are the same at every $p \in\ ^{1,3}$ one can think of ^{1,3} simply as the vector space ^{4} equipped with the **Minkowski inner product** $g(v,w) = \eta_{\alpha\beta} v^\alpha w^\beta = v^0 w^0 - v^1 w^1 - v^2 w^2 - v^3 w^3$. A vector v in ^{1,3} is said to be **spacelike**, **timelike**, or **null** if $g(v,v)$ is $< 0, > 0$, or $= 0$, respectively. The physical origin of the terminology will emerge as we procced.

We introduce a matrix

$$\eta = (\eta_{\alpha\beta}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and note that its inverse $\eta^{-1} = (\eta^{\alpha\beta})$ is, in fact, equal to η.

Now, to build our gauge theory we begin by letting X denote some open submanifold of ^{1,3} (the charges creating our electromagnetic field live in ^{1,3} and we intend to carve out their worldlines and deal only with the source free Maxwell equations on the resulting open submanifold of ^{1,3}). The gauge group G is $U(1)$ so we consider a principal $U(1)$-bundle

$$U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$$

over X and a connection $\boldsymbol{\omega}$ on it (we consider first the pure Yang-Mills theory in which matter fields are absent). Since $U(1)$ is Abelian, all brackets in the Lie algebra $u(1) = \mathrm{Im}$ are zero so the curvature $\boldsymbol{\Omega}$ of $\boldsymbol{\omega}$ is given by $\boldsymbol{\Omega} = d\boldsymbol{\omega}$. If $s : V \longrightarrow P$ is a local cross-section, then we may write the local gauge potential and field strength as

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= s^* \boldsymbol{\omega} = -\boldsymbol{i}\, \boldsymbol{A} \\ \boldsymbol{\mathcal{F}} &= s^* \boldsymbol{\Omega} = d\boldsymbol{\mathcal{A}} = -\boldsymbol{i}\, d\boldsymbol{A} = -\boldsymbol{i}\, \boldsymbol{F}, \end{aligned}$$

where $\boldsymbol{A}$ and $\boldsymbol{F}$ are real-valued forms on V (the minus sign is conventional). If $s_i : V_i \longrightarrow P$ and $s_j : V_j \longrightarrow P$ are two such local cross-sections with $V_j \cap V_i \neq \emptyset$ and if $g_{ij} : V_j \cap V_i \longrightarrow U(1)$ is the corresponding transition function, then

$$\boldsymbol{\mathcal{A}}_j = g_{ij}{}^{-1} \boldsymbol{\mathcal{A}}_i\, g_{ij} + g_{ij}{}^{-1}\, dg_{ij} = \boldsymbol{\mathcal{A}}_i + g_{ij}{}^{-1}\, dg_{ij}$$

and

$$\boldsymbol{\mathcal{F}}_j = {g_{ij}}^{-1}\,\boldsymbol{\mathcal{F}}_i\,g_{ij} = \boldsymbol{\mathcal{F}}_i$$

on $V_j \cap V_i$ because $U(1)$ is Abelian. In particular, the local field strengths, since they agree on any intersections of their domains, piece together to give a *globally defined field strength* 2-form $\boldsymbol{\mathcal{F}}$ on X. This is a peculiarity of Abelian gauge theories and one should note that, even here, the potentials $\boldsymbol{\mathcal{A}}$ do *not* agree on the intersections of their domains and so do not give rise to a globally defined object on X. Indeed, since the transition function g_{ij} is a map into $U(1)$ it can be written as

$$g_{ij}(x) = e^{-\boldsymbol{i}\Lambda_{ij}(x)}$$

so that ${g_{ij}}^{-1}dg_{ij} = -\boldsymbol{i}d\Lambda_{ij}$ and $\boldsymbol{\mathcal{A}}_j = \boldsymbol{\mathcal{A}}_i - \boldsymbol{i}d\Lambda_{ij}$. Equivalently,

$$\boldsymbol{A}_j = \boldsymbol{A}_i + d\Lambda_{ij},$$

which is the traditional form for the relationship between two "vector potentials."

Relative to standard coordinates x^0, x^1, x^2, x^3 on ^{1,3} we can write, for any $s : V \longrightarrow P$,

$$\boldsymbol{\mathcal{A}} = \mathcal{A}_\alpha dx^\alpha = -\boldsymbol{i}A_\alpha dx^\alpha$$

and

$$\boldsymbol{\mathcal{F}} = \frac{1}{2}\mathcal{F}_{\alpha\beta}dx^\alpha \wedge dx^\beta = -\frac{1}{2}\boldsymbol{i}F_{\alpha\beta}dx^\alpha \wedge dx^\beta,$$

where

$$\begin{aligned}
\mathcal{F}_{\alpha\beta} &= \partial_\alpha\,\boldsymbol{\mathcal{A}}_\beta - \partial_\beta\,\boldsymbol{\mathcal{A}}_\alpha + [\boldsymbol{\mathcal{A}}_\alpha, \boldsymbol{\mathcal{A}}_\beta] \\
&= \partial_\alpha\,\boldsymbol{\mathcal{A}}_\beta - \partial_\beta\,\boldsymbol{\mathcal{A}}_\alpha \qquad \text{(because } U(1) \text{ is Abelian)} \\
&= -\boldsymbol{i}(\partial_\alpha\,A_\beta - \partial_\beta\,A_\alpha) \\
&= -\boldsymbol{i}F_{\alpha\beta}.
\end{aligned}$$

The $F_{\alpha\beta}$ are skew-symmetric in α and β. To make some contact with the notation used in physics we define functions E^1, E^2, E^3 and B^1, B^2, B^3 by

$$F_{i0} = E_i$$

and

$$F_{ij} = \varepsilon_{ijk}\,B^k$$

where $i, j, k = 1, 2, 3$ and ε_{ijk} is the Levi-Civita symbol (1 if ijk is an even permutation of 123, -1 if ijk is an odd permutation of 123, and 0 otherwise). Thus,

$$(F_{\alpha\beta}) = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix}$$

and

$$\begin{aligned}\boldsymbol{F} = \frac{1}{2}\, F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta &= -E^1 dx^0 \wedge dx^1 - E^2 dx^0 \wedge dx^2 - E^3 dx^0 \wedge dx^3 \\ &\quad + B^3 dx^1 \wedge dx^2 - B^2 dx^1 \wedge dx^3 + B^1 dx^2 \wedge dx^3 \\ &= (E^1 dx^1 + E^2 dx^2 + E^3 dx^3) \wedge dx^0 \\ &\quad + B^3 dx^1 \wedge dx^2 + B^1 dx^2 \wedge dx^3 + B^2 dx^3 \wedge dx^1.\end{aligned}$$

One is to think of $\vec{E} = (E^1, E^2, E^3)$ and $\vec{B} = (B^1, B^2, B^3)$ as the "electric field" and the "magnetic field," respectively, that correspond to $\boldsymbol{\mathcal{F}}$ (the justification for thinking this way will appear shortly).

Next we introduce functions $F^{\alpha\beta}$ on X defined by

$$F^{\alpha\beta} = \eta^{\alpha\gamma}\eta^{\beta\delta} F_{\gamma\delta}, \qquad \alpha, \beta = 0, 1, 2, 3$$

(classically this is referred to as "raising the indices" with the Minkowski metric). Thus, for example, $F^{01} = \eta^{0\gamma}\eta^{1\delta}\, F_{\gamma\delta} = \eta^{00}\eta^{11}\, F_{01} = (1)(-1)\, F_{01} = -F_{01} = E^1$ and $F^{12} = \eta^{1\gamma}\eta^{2\delta}\, F_{\gamma\delta} = \eta^{11}\eta^{22} F_{12} = (-1)(-1)F_{12} = F_{12} = B^3$, etc., so

$$(F^{\alpha\beta}) = \begin{pmatrix} 0 & E^1 & E^2 & E^3 \\ -E^1 & 0 & B^3 & -B^2 \\ -E^2 & -B^3 & 0 & B^1 \\ -E^3 & B^2 & -B^1 & 0 \end{pmatrix}.$$

The Hodge dual ${}^*\boldsymbol{F}$ of $\boldsymbol{F}$ is defined to be the 2-form on X whose standard components are given by

$${}^*F_{\alpha\beta} = \frac{1}{2}\varepsilon_{\alpha\beta\gamma\delta}\, F^{\gamma\delta}, \quad \alpha, \beta = 0, 1, 2, 3.$$

Writing these out one finds that

$$({}^*F_{\alpha\beta}) = \begin{pmatrix} 0 & B^1 & B^2 & B^3 \\ -B^1 & 0 & E^3 & -E^2 \\ -B^2 & -E^3 & 0 & E^1 \\ -B^3 & E^2 & -E^1 & 0 \end{pmatrix}$$

so that

$$\begin{aligned}{}^*F &= \frac{1}{2}\, {}^*F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta \\ &= (-B^1 dx^1 - B^2 dx^2 - B^3 dx^3) \wedge dx^0 \\ &\quad + E^3 dx^1 \wedge dx^2 + E^1 dx^2 \wedge dx^3 + E^2 dx^3 \wedge dx^1.\end{aligned}$$

One verifies that

$${}^{**}\boldsymbol{F} = -\boldsymbol{F} \qquad (\text{on } \ {}^{1,3}).$$

We also define

$$^*\boldsymbol{\mathcal{F}} = -\boldsymbol{i}\,^*\boldsymbol{F}.$$

Finally, one can also "raise the indices" of $^*\boldsymbol{F}$ and define $^*F^{\alpha\beta} = \eta^{\alpha\gamma}\,\eta^{\beta\delta}\,^*F_{\alpha\beta}$ so that

$$(^*F_{\alpha\beta}) = \begin{pmatrix} 0 & -B^1 & -B^2 & -B^3 \\ B^1 & 0 & E^3 & -E^2 \\ B^2 & -E^3 & 0 & E^1 \\ B^3 & E^2 & -E^1 & 0 \end{pmatrix}.$$

One can think of the Hodge dual as the 2-form obtained by formally replacing $\vec{B}$ by $\vec{E}$ and $\vec{E}$ by $-\vec{B}$.

All of this apparently *ad hoc* notation will eventually be seen to fit naturally into the general scheme of things. For the time being the reader may wish to regard all of it as simply a useful bookkeeping device. For example, one has the following easily verified formulas:

$$\begin{aligned} \frac{1}{2}\,F_{\alpha\beta}\,F^{\alpha\beta} &= |\vec{B}|^2 - |\vec{E}|^2 \\ \frac{1}{4}\,F_{\alpha\beta}\,^*F^{\alpha\beta} &= \vec{E}\cdot\vec{B} \end{aligned}$$

for the two scalar invariants normally associated with $\boldsymbol{F}$ in classical electromagnetic theory.

But what is the justification for all of these references to classical electromagnetic theory? We began by looking at an arbitrary connection $\boldsymbol{\omega}$ on an arbitrary principal $U(1)$-bundle over an open submanifold of Minkowski spacetime and have deviously interjected things we have called "electric fields" and "magnetic fields." Is there any reason to believe that these objects have anything whatever to do with what physicists call "electric fields" and "magnetic fields?" The answer lies in the Yang-Mills equations. We are, after all, not really interested in arbitrary connections, but only in the stationary values of the Yang-Mills action (page 56). In our present circumstances we will find that this action can be written

$$\mathcal{YM}(\boldsymbol{\omega}) = \int_X -\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}dx^0dx^1dx^2dx^3$$

and the corresponding Yang-Mills equations are

$$d\,^*\boldsymbol{F} = 0,$$

while the Bianchi identity is

$$d\boldsymbol{F} = 0$$

(in the Abelian case, covariant exterior derivatives are just ordinary exterior derivatives). In standard coordinates these read

$$\partial_\alpha F^{\alpha\beta} = 0, \quad \beta = 0, 1, 2, 3$$

and

$$\partial_\alpha {}^*F^{\alpha\beta} = 0, \quad \beta = 0, 1, 2, 3,$$

respectively. Now, the remarkable part is that, if one writes these out in terms of the $\vec{E}$'s and $\vec{B}$'s we introduced earlier the result is

$$\vec{\nabla} \times \vec{B} - \frac{\partial \vec{E}}{\partial x^0} = \vec{0} \quad \text{and} \quad \vec{\nabla} \cdot \vec{E} = 0 \qquad (d\,{}^*\boldsymbol{F} = 0)$$

and

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial x^0} = \vec{0} \quad \text{and} \quad \vec{\nabla} \cdot \vec{B} = 0 \qquad (d\boldsymbol{F} = 0),$$

respectively, where $\vec{\nabla} = (\partial_1, \partial_2, \partial_3)$ is the usual gradient operator and the $\times$ and $\cdot$ refer to the cross product and dot product on ^{3}. These, of course, are the source free Maxwell equations in their usual guise.

Remark: One should note that the second pair of Maxwell equations corresponds to the Bianchi identity and is, in this sense, purely geometrical, i.e., due to the fact that we have chosen to model our fields as connections on principal bundles. The proper way to look at this, however, is the other way around. We are able to build a model of an electromagnetic field as a connection on a principal bundle only because of this second pair of Maxwell equations. As a matter of terminology, one says that a differential form whose exterior derivative is zero is **closed**. Thus, the source free Maxwell equations assert that both $\boldsymbol{F}$ and ${}^*\boldsymbol{F}$ are closed.

Before moving on let us make one more bit of notational contact with physics. The local gauge potentials $\boldsymbol{\mathcal{A}} = -\boldsymbol{i}\boldsymbol{A} = -\boldsymbol{i}A_\alpha dx^\alpha$ satisfy $d\boldsymbol{A} = \boldsymbol{F}$, i.e., $F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha$, which we rewrite as follows: Define $A^\alpha = \eta^{\alpha\gamma} A_\gamma$ for $\alpha = 0, 1, 2, 3$ (i.e., "raise the indices" of $\boldsymbol{A}$) and write

$$(A^0, A^1, A^2, A^3) = \left(V, \vec{A}\right),$$

where $V = A^0 = A_0$ and $\vec{A} = (A^1, A^2, A^3) = (-A_1, -A_2, -A_3)$. Then a brief calculation shows that $d\boldsymbol{A} = \boldsymbol{F}$ becomes

$$\vec{E} = -\frac{\partial \vec{A}}{\partial x^0} - \vec{\nabla} V \quad \text{and} \quad \vec{B} = \vec{\nabla} \times \vec{A}$$

which are the usual expressions from physics for the electric and magnetic fields in terms of the scalar and vector potentials.

We would now like to write out two concrete examples (Coulomb fields and Dirac monopoles) which illustrate all of this apparatus and which, more importantly, make clear the difference between magnetic charge (which is "topological") and electric charge (which is not). We will build the examples by specifying local gauge potentials and allowing these to determine (by the way they are related on the intersections of their domains) the transition functions and therefore the bundle on which the corresponding connections are defined. In order to fully appreciate the topological nature of what is going on here, however, we must briefly anticipate some material that we will treat carefully only somewhat later in the text.

Principal $U(1)$-bundles over any manifold X are classified up to equivalence by a certain de Rham cohomology class of X called the first Chern class of X. Very briefly, here's what these words mean: As mentioned earlier, a differential form $\boldsymbol{\varphi}$ whose exterior derivative $d\boldsymbol{\varphi}$ is zero is said to be closed. $\boldsymbol{\varphi}$ is said to be **exact** if it is the exterior derivative of some form $\boldsymbol{\psi}$ of degree one less ($\boldsymbol{\varphi} = d\boldsymbol{\psi}$). According to the Poincaré Lemma (Theorem 4.4.2), every exact form is closed ($d^2 = 0$), but the converse is not true. Two closed forms $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ are said to be **cohomologous** if they differ by an exact form ($\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2 = d\boldsymbol{\psi}$). For each degree k this is an equivalence relation and the set of equivalence classes, which admits a natural real vector space structure, is denoted $H^k_{\mathrm{deR}}(X)$ and called the k^{th} **de Rham cohomology group** of X.

Now let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be any principal G-bundle over X (with G a matrix Lie group). The **first Chern class** $c_1(P)$ of the bundle is an element of $H^2_{\mathrm{deR}}(X)$ defined as follows: Choose *any* connection $\boldsymbol{\omega}$ on the bundle (we will prove later that connections exist on any principal bundle). Let $\boldsymbol{\Omega}$ be the curvature of $\boldsymbol{\omega}$. For any local cross-section s let $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ be the local field strength. These generally depend on the choice of s and do not agree on the intersections of their domains (except in the Abelian case). Indeed, if s^g is another cross-section, then $\boldsymbol{\mathcal{F}}^g = g^{-1}\boldsymbol{\mathcal{F}}g$ on the intersection. Notice, however, that, since the trace of a matrix is invariant under conjugation, trace $\boldsymbol{\mathcal{F}}^g =$ trace $\boldsymbol{\mathcal{F}}$ on the intersection and so these piece together to give a globally defined 2-form on X which we will denote simply trace $\boldsymbol{\mathcal{F}}$. *A priori* trace $\boldsymbol{\mathcal{F}}$ is complex-valued, but one can show that, in fact, it takes values in Im so that $\frac{\boldsymbol{i}}{2\pi}$ trace $\boldsymbol{\mathcal{F}}$ is real-valued (the $\frac{1}{2\pi}$ actually forces the integrals of this 2-form over compact, oriented surfaces in X to be integers—not obvious, but true). It also happens that this 2-form is closed and so determines a cohomology class

$$c_1(P) = \left[\frac{\boldsymbol{i}}{2\pi}\text{trace } \boldsymbol{\mathcal{F}}\right] \in H^2_{\mathrm{deR}}(X).$$

Now, the remarkable part of all this is that this cohomology class (*not* the 2-form, but its cohomology class) does not depend on the initial choice of the connection $\boldsymbol{\omega}$ from which it arose. It is therefore a characteristic of the bundle itself and not of the connection. Indeed, $c_1(P)$ is the simplest example of what is called a **characteristic class** for the bundle. We'll encounter one more example of such a thing (the second Chern class) in Section 2.5.

Now let us specialize to the case of $U(1)$-bundles. Here we have seen that there is a globally defined field strength $\boldsymbol{\mathcal{F}}$ for any connection. Moreover, since $u(1)$ consists of 1×1 matrices, the trace just picks out the sole entry in this matrix so trace $\boldsymbol{\mathcal{F}} = -\boldsymbol{i}\boldsymbol{F}$ and therefore

$$c_1(P) = \left[\frac{\boldsymbol{i}}{2\pi}(-\boldsymbol{iF})\right] = \frac{1}{2\pi}\,[\boldsymbol{F}] \qquad\qquad (G = U(1)).$$

In particular, the first Chern class for a principal $U(1)$-bundle over an open submanifold of Minkowski spacetime is just $\frac{1}{2\pi}$ times the cohomology class of the globally defined (electromagnetic) field $\boldsymbol{F}$ on X. Since principal $U(1)$-bundles over any manifold are classified up to equivalence by their first Chern classes (Appendix E, [**FU**]), we conclude that the $U(1)$-bundle on which an electromagnetic field $\boldsymbol{F}$ is to be modeled as a connection is uniquely determined by the cohomology class of $\boldsymbol{F}$.

Now we return to our concrete examples. First, the Coulomb field, i.e., a static, purely electric field of a point charge which we assume to be located at the (x^1, x^2, x^3)-origin in $\ ^{1,3}$. Thus, the worldline of our source is the x^0-axis in $\ ^{1,3}$ so we take

$$X = \ ^{1,3} - \{(x^0, 0, 0, 0) \in \ ^{1,3} : x^0 \in \ \ \}.$$

Define $\boldsymbol{A} = A_\alpha dx^\alpha = (-n/\rho)dx^0$, where n is an integer and $\rho > 0$ with $\rho^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$ (we measure "charge" in multiples of the charge of the electron so it is an integer). A simple calculation shows that the functions $F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha,\ \alpha, \beta = 0, 1, 2, 3$, are given by

$$(F_{\alpha\beta}) = \frac{n}{\rho^3}\begin{pmatrix} 0 & -x^1 & -x^2 & -x^3 \\ x^1 & 0 & 0 & 0 \\ x^2 & 0 & 0 & 0 \\ x^3 & 0 & 0 & 0 \end{pmatrix}.$$

Thus,

$$\boldsymbol{F} = \frac{n}{\rho^3}(x^1 dx^1 + x^2 dx^2 + x^3 dx^3) \wedge dx^0$$

so

$$\vec{B} = \vec{0} \quad \text{and} \quad \vec{E} = \frac{n}{\rho^3}\vec{r}, \quad \vec{r} = (x^1,\ x^2,\ x^3).$$

This is, of course, the classical Coulomb field that we wish to describe.

The critical observation here is this: Our Coulomb potential $\boldsymbol{A} = (-n/\rho)dx^0$ is defined and satisfies $d\boldsymbol{A} = \boldsymbol{F}$ *globally* on all of X so that $\boldsymbol{F}$ is exact on X and its cohomology class $[\boldsymbol{F}] \in H^2_{\mathrm{deR}}(X)$ is zero. Thus, the $U(1)$-bundle on which $\boldsymbol{F}$ is modeled by a connection with field strength $\boldsymbol{\mathcal{F}} = -\boldsymbol{iF}$ has first Chern class zero and so must be the trivial bundle. This is true for *any* charge n so that, in particular, the electric charge of the source is not encoded in the topology of this bundle (we will find that the situation is quite different for a Dirac magnetic monopole).

Here's another way to look at this: Somewhat later we will calculate the cohomology of $X = \ ^{1,3} - \{(x^0, 0, 0, 0) \in \ ^{1,3} : x^0 \in \ \ \}$ and find that

$$H^k_{\mathrm{deR}}(X) = \begin{cases} \ \ , & k = 0, 2 \\ 0, & \text{otherwise} \end{cases}.$$

Now, the 2-form $\boldsymbol{F}$ describing the Coulomb field on X is cohomologically trivial, i.e., $[\boldsymbol{F}] = 0 \in H^2_{\mathrm{deR}}(X)$. After learning how to integrate 2-forms over 2-dimensional manifolds we will find that this implies that the integral of $\boldsymbol{F}$ over any closed, smoothly embedded surface in some $x^0 =$ constant slice must be zero. In particular, these integrals do not detect any "enclosed charge." The rest of the 2-dimensional cohomology of X is to be found in the dual ${}^*\boldsymbol{F} = \frac{1}{2}\, {}^*F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta$, where

$$({}^*F_{\alpha\beta}) = \frac{n}{\rho^3} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & x^3 & -x^2 \\ 0 & -x^3 & 0 & x^1 \\ 0 & x^2 & -x^1 & 0 \end{pmatrix}.$$

Thus,
$$ {}^*\boldsymbol{F} = \frac{n}{\rho^3}\,(x^1\, dx^2 \wedge dx^3 - x^2\, dx^1 \wedge dx^3 + x^3\, dx^1 \wedge dx^2).$$

Notice that, on the 2-sphere $\rho = 1$ in any $x^0 =$ constant slice of X, ${}^*\boldsymbol{F}$ reduces to n times $x^1\, dx^2 \wedge dx^3 - x^2\, dx^1 \wedge dx^3 + x^3\, dx^1 \wedge dx^2$ on S^2. This, as we shall see, is what is called the standard volume form of S^2 and its integral over S^2 is the area 4π of S^2 (Section 4.6). Thus, the integral of ${}^*\boldsymbol{F}$ over this sphere is $4\pi n$, which is nonzero and this implies that ${}^*\boldsymbol{F}$ cannot be cohomologically trivial. Thus, $[{}^*\boldsymbol{F}]$ generates $H^2_{\mathrm{deR}}(X)$. Furthermore, we will show that the integral of ${}^*\boldsymbol{F}$ over any 2-sphere surrounding $\{(x^0, 0, 0, 0) : x^0 \in \quad \}$ is the same so these integrals "detect" the charge n enclosed by the sphere. Over any 2-sphere that does not enclose the x^0-axis, the integral of ${}^*\boldsymbol{F}$ is zero (this will follow from "Stokes' Theorem").

The electric charge n is not "topological" because all Coulomb fields are represented by connections on the trivial bundle—the charge is not encoded in the topology of the bundle. The situation is quite different for a Dirac monopole (which is not surprising since the Hodge dual for 2-forms on X essentially interchanges "electric" and "magnetic" so, for a magnetic charge, $[\boldsymbol{F}]$ will play the role that $[{}^*\boldsymbol{F}]$ played for the Coulomb field). In more detail, we once again let $X = \quad ^{1,3} - \{(x^0, 0, 0, 0) \in \quad ^{1,3} : x^0 \in \quad \}$. For the moment we let g denote an arbitrary real number (to be thought of as the magnetic "charge" of the monopole whose worldline is the x^0-axis). We are interested in the field $\boldsymbol{F} = \frac{1}{2}\, F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta$, where

$$(F_{\alpha\beta}) = \frac{g}{\rho^3} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & x^3 & -x^2 \\ 0 & -x^3 & 0 & x^1 \\ 0 & x^2 & -x^1 & 0 \end{pmatrix}.$$

Thus,
$$\boldsymbol{F} = \frac{g}{\rho^3}\,(x^1\, dx^2 \wedge dx^3 - x^2\, dx^1 \wedge dx^3 + x^3\, dx^1 \wedge dx^2)$$

so

$$\vec{E} = \vec{0} \quad \text{and} \quad \vec{B} = \frac{g}{\rho^3}\,\vec{r}, \qquad \vec{r} = (x^1, x^2, x^3).$$

Since $\boldsymbol{F}$ is independent of x^0 this defines a 2-form on any x^0 = constant slice. Moreover, expressed in terms of standard spherical coordinates (ρ, φ, θ) on such a slice,

$$\boldsymbol{F} = g\,\sin\varphi\, d\varphi \wedge d\theta.$$

Notice that this is independent of ρ and so may be further restricted to the copy $\rho = 1$ of S^2 in, say, the $x^0 = 0$ slice of X. Now, if the monopole field on X is the field strength of some connection on a $U(1)$-bundle over X, then its restrictions would likewise be field strengths for connections on $U(1)$-bundles over the submanifolds (restriction means pullback by the inclusion map and the inclusion of a restricted bundle is a bundle map). Henceforth, we will concentrate on these restrictions.

For the field $\boldsymbol{F}$ under consideration there is no globally defined potential $\boldsymbol{A}$ satisfying $d\boldsymbol{A} = \boldsymbol{F}$ (pages 2–3 of [**N4**]), but there are the usual local potentials. Specifically, we define $\boldsymbol{A}_N$ and $\boldsymbol{A}_S$ on $U_N = S^2 - \{(0,0,0,-1)\}$ and $U_S = S^2 - \{(0,0,0,1)\}$, respectively, by

$$\boldsymbol{A}_N = g(1 - \cos\varphi)\, d\theta$$

and

$$\boldsymbol{A}_S = -g(1 + \cos\varphi)\, d\theta.$$

Then, on their respective domains, these satisfy $d\boldsymbol{A}_N = \boldsymbol{F}$ and $d\boldsymbol{A}_S = \boldsymbol{F}$. Consider now the corresponding $u(1)$-valued forms (we identify $u(1)$ with Im):

$$\begin{aligned} \boldsymbol{\mathcal{A}}_N &= -\boldsymbol{i}\boldsymbol{A}_N &&= -\boldsymbol{i}g(1 - \cos\varphi)\, d\theta \\ \boldsymbol{\mathcal{A}}_S &= -\boldsymbol{i}\boldsymbol{A}_S &&= \quad \boldsymbol{i}g\,(1 + \cos\varphi)\, d\theta \\ \boldsymbol{\mathcal{F}} &= -\boldsymbol{i}\,\boldsymbol{F} &&= -\boldsymbol{i}g\,\sin\varphi\, d\varphi \wedge d\theta \end{aligned}$$

If our monopole field is to be modeled by a connection on some principal $U(1)$-bundle over S^2, then that bundle could be trivialized over U_N and U_S and so would have a transition function $g_{SN} : U_N \cap U_S \longrightarrow U(1)$ for which $\boldsymbol{\mathcal{A}}_N = {g_{SN}}^{-1}\, \boldsymbol{\mathcal{A}}_S g_{SN} + {g_{SN}}^{-1} dg_{SN}$. But notice that $\boldsymbol{\mathcal{A}}_N - \boldsymbol{\mathcal{A}}_S = -2g\boldsymbol{i}d\theta$ so

$$\boldsymbol{\mathcal{A}}_N = \boldsymbol{\mathcal{A}}_S - 2g\boldsymbol{i}d\theta = e^{2g\theta\boldsymbol{i}}\boldsymbol{\mathcal{A}}_S e^{-2g\theta\boldsymbol{i}} + e^{2g\theta\boldsymbol{i}} d(e^{-2g\theta\boldsymbol{i}})$$

and this gives

$$g_{SN}(\varphi, \theta) = e^{-2g\theta\boldsymbol{i}}.$$

Next we observe that the only values of the constant g that are of any interest are given by

$$g = \frac{n}{2}, \qquad n \in \mathbb{Z}.$$

One can understand this in a number of ways, all of which are instructive. From the point of view of physics, it is just the Dirac quantization condition (page 7, [**N4**]). On the other hand, if $g_{SN}(\varphi, \theta) = e^{-2g\theta\boldsymbol{i}}$ is really the transition function for a principal $U(1)$-bundle $U(1) \hookrightarrow P \longrightarrow S^2$, then that bundle is characterized by its first Chern class $c_1(P)$. As we pointed out earlier, the integral of $c_1(P)$ over S^2 is an integer, called the **first Chern number** of the bundle. But a simple calculation gives

$$\int_{S^2} c_1(P) = \frac{1}{2\pi}\int_{S^2} \boldsymbol{F} = \frac{g}{2\pi}\int_{S^2} \sin\varphi d\varphi \wedge d\theta = 2\,g$$

(we assume S^2 has its standard orientation). Thus, $2g \in \mathbb{Z}$. From yet another perspective, the restriction of $g_{SN}(\varphi,\theta) = e^{-2g\theta \boldsymbol{i}}$ to the equatorial circle S^1 in $U_N \cap U_S$ (i.e., $e^{\boldsymbol{i}\theta} \longrightarrow (e^{\boldsymbol{i}\theta})^{-2g}$) would be the "characteristic map" whose homotopy type determines the bundle (page 228, [**N4**]) and this map is not even well-defined (single-valued) on S^1 unless $2g$ is an integer. However you choose to view the situation, we will henceforth restrict our attention to the following forms:

$$\begin{aligned}
\boldsymbol{\mathcal{A}}_N &= -\frac{1}{2}\, n\boldsymbol{i}\,(1-\cos\varphi)\,d\theta \\
\boldsymbol{\mathcal{A}}_S &= \;\;\frac{1}{2}\, n\boldsymbol{i}\,(1+\cos\varphi)d\theta \\
\boldsymbol{\mathcal{F}} &= -\frac{1}{2}\, n\boldsymbol{i}\sin\varphi\; d\varphi\wedge d\theta.
\end{aligned}$$

For each fixed integer n, the potentials $\boldsymbol{\mathcal{A}}_N$ and $\boldsymbol{\mathcal{A}}_S$ uniquely determine a connection $\boldsymbol{\omega}_n$ on the principal $U(1)$-bundle

$$U(1) \hookrightarrow P_n \xrightarrow{\mathcal{P}_n} S^2$$

whose transition function g_{SN} is given by

$$g_{SN}(\varphi,\theta) = e^{-n\theta\boldsymbol{i}}.$$

The globally defined field strength on S^2 is $\boldsymbol{\mathcal{F}}$ and represents the field of a Dirac monopole of "magnetic charge" n (which is the Chern number of the bundle). Since the Chern number is a topological characteristic of the bundle, magnetic charge is directly encoded into the topology of the bundle and is therefore an instance of a "topological charge."

Although it is not really necessary to do so (because the transition functions and local gauge potentials contain all of the relevant information about the bundles and the connections), it is possible to describe these Dirac monopoles more explicitly. We will outline such a description.

$$\boxed{n=1}$$

This gives the natural connection on the complex Hopf bundle

$$U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_1} S^2,$$

where $\mathcal{P}_1$ is the restriction to $S^3 \subseteq \ ^2$ of the map

$$\mathcal{P}_1\left(z^1, z^2\right) = \left(z^1\bar{z}^2 + \bar{z}^1 z^2, -\boldsymbol{i} z^1\bar{z}^2 + \boldsymbol{i}\bar{z}^1 z^2, \left|z^1\right|^2 - \left|z^2\right|^2\right).$$

$\boldsymbol{\omega}_1$ is given by

$$\boldsymbol{\omega}_1 = \boldsymbol{i}\iota^*(\mathrm{Im}(\bar{z}^1 dz^1 + \bar{z}^2 dz^2)),$$

where $\iota : S^3 \hookrightarrow \ ^2$ is the inclusion.

$$\boxed{n = -1}$$

This gives the natural connection on the alternate version of the complex Hopf bundle

$$U(1) \hookrightarrow S^3 \xrightarrow{\mathcal{P}_{-1}} S^2,$$

where $\mathcal{P}_{-1}$ is the restriction to $S^3 \subseteq \ ^2$ of the map

$$\mathcal{P}_{-1}\left(z^1, z^2\right) = \left(z^1\bar{z}^2 + \bar{z}^1 z^2, \boldsymbol{i} z^1\bar{z}^2 - \boldsymbol{i}\bar{z}^1 z^2, \left|\, z^1 \,\right|^2 - \left|\, z^2 \,\right|^2\right).$$

$\boldsymbol{\omega}_{-1}$ is given by

$$\boldsymbol{\omega}_{-1} = \boldsymbol{i}\iota^*\left(\mathrm{Im}\left(\bar{z}^1 dz^1 + \bar{z}^2 dz^2\right)\right) = \boldsymbol{\omega}_1,$$

where $\iota : S^3 \hookrightarrow \ ^2$ is the inclusion.

$$\boxed{n = 0}$$

This gives the flat connection on the trivial bundle

$$U(1) \hookrightarrow S^2 \times U(1) \xrightarrow{\mathcal{P}_0} S^2,$$

where $\mathcal{P}_0$ is the projection onto the first factor. $\boldsymbol{\omega}_0$ is given by

$$\boldsymbol{\omega}_0 = \pi^*\boldsymbol{\Theta},$$

where $\pi : S^2 \times U(1) \longrightarrow U(1)$ is the projection onto the second factor and $\boldsymbol{\Theta}$ is the Cartan 1-form on $U(1)$.

$$\boxed{n > 1}$$

Denote by $U(1) \hookrightarrow P_n \xrightarrow{\mathcal{P}_n} S^2$ the principal $U(1)$-bundle over S^2 with transition function $g_{SN}(\varphi, \theta) = e^{-n\theta\boldsymbol{i}}$. We can identify P_n explicitly as a manifold as follows: Identify the discrete group $\mathbb{Z}_n$ of integers modulo n with the following subgroup of $U(1)$:

$$\mathbb{Z}_n = \{e^{2k\pi\boldsymbol{i}/n} : k = 0, 1, \ldots, n-1\}.$$

Then $\mathbb{Z}_n$ acts on S^3 on the right (because $U(1)$ does). We let $S^3/\mathbb{Z}_n$ be the orbit space, e.g., $S^3/\mathbb{Z}_2 = \ ^3$. One can provide $S^3/\mathbb{Z}_n$ with a manifold structure in the same way as for $\ ^3$. The Hopf map $\mathcal{P}_1 : S^3 \longrightarrow S^2$ carries each orbit of the usual $U(1)$-action on S^3 to a point in S^2 so it does the same for a $\mathbb{Z}_n$-orbit. Moreover, each point of S^2 is the image under $\mathcal{P}_1$ of a $\mathbb{Z}_n$-orbit (indeed, of many $\mathbb{Z}_n$-orbits). Thus, $\mathcal{P}_1$ descends to a surjective map

$$P_n : S^3 / \mathbb{Z}_n \longrightarrow S^2.$$

Also note that the usual $U(1)$-action on S^3 carries any $\mathbb{Z}_n$-orbit onto another $\mathbb{Z}_n$-orbit which is inside the same $U(1)$-orbit (and so has the same image under $\mathcal{P}_n$). Thus, the $U(1)$-action on S^3 descends to a $U(1)$- action on $S^3/\mathbb{Z}_n$ which preserves the fibers of $\mathcal{P}_n$. Local triviality of $\mathcal{P}_n : S^3/\mathbb{Z}_n \longrightarrow S^2$ follows so that

$$U(1) \hookrightarrow S^3 / \mathbb{Z}_n \xrightarrow{\mathcal{P}_n} S^2$$

is a principal $U(1)$-bundle and the transition function g_{SN} is given by $g_{SN}(\varphi, \theta) = e^{-n\,\theta\,\boldsymbol{i}}$. Since a bundle is determined by its transition functions, we have an explicit model for $U(1) \hookrightarrow P_n \xrightarrow{\mathcal{P}_n} S^2$. In particular, $P_n \cong S^3 / \mathbb{Z}_n$.

Remark: $P_n = S^3 / \mathbb{Z}_n$ is an example of a "lens space."

Since the transition function for $U(1) \hookrightarrow S^3 / \mathbb{Z}_n \xrightarrow{\mathcal{P}_n} S^2$ is $g_{SN}(\varphi,\, \theta) = e^{-n\,\theta\,\boldsymbol{i}}, \boldsymbol{\mathcal{A}}_N = -\frac{1}{2} n\boldsymbol{i}\,(1 - \cos\varphi)\, d\theta$ and $\boldsymbol{\mathcal{A}}_S = \frac{1}{2}\, n\boldsymbol{i}(1 + \cos\varphi)\, d\theta$ are gauge potentials on this bundle. The connection $\boldsymbol{\omega}_n$ they determine can be described as follows: The Im -valued 1-form $\tilde{\boldsymbol{\omega}}_n$ on $\ ^2$ given by $\tilde{\boldsymbol{\omega}}_n = \boldsymbol{i}\, n\, \mathrm{Im}(\bar{z}^1 dz^1 + \bar{z}^2 dz^2)$ restricts to an Im -valued 1-form $\iota^* \tilde{\boldsymbol{\omega}}_n$ on S^3 that is invariant under the $\mathbb{Z}_n$-action and so descends to an Im -valued 1-form $\boldsymbol{\omega}_n$ on $S^3 / \mathbb{Z}_n$. $\boldsymbol{\omega}_n$ is the required connection on $U(1) \hookrightarrow S^3 / \mathbb{Z}_n \xrightarrow{\mathcal{P}_n} S^2$.

$$\boxed{n < -1}$$

Here the construction is exactly the same as above for $n > 1$ except that $\mathcal{P}_1$ is replaced by $\mathcal{P}_{-1}$.

2.3 Spin Zero Electrodynamics

We consider next our first example of a gauge theory in which matter fields are present. The gauge field will be an electromagnetic field of the type discussed in Section 2.2. The matter field coupled to this gauge field will represent a charged particle experiencing the effects of the electromagnetic field. According to the General Framework described in Section 2.1, the description of such a particle will require a vector space $\mathcal{V}$ and a representation $\rho : U(1) \longrightarrow GL(\mathcal{V})$ of $U(1)$ on $\mathcal{V}$. Now, in physics, charged particles have wavefunctions with a certain number of complex components, the number of such components being determined by the particle's spin s. Specifically, s is an element of $\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots\}$ and the wavefunction of a particle with spin s

has $2s+1$ components. The simplest case is that of a particle of spin 0 (e.g., a π^- meson) for which the wavefunction takes values in . This is the case we consider here.

Remark: Certain technical complications arise when $s > 0$. For example, an electron has $s = \frac{1}{2}$ and, according to Dirac, the corresponding matter field is defined, not on a $U(1)$-bundle over X, but on a certain $SL(2, \;)$-bundle over X called a spinor bundle. Nevertheless, an electron responds to an electromagnetic, i.e., a $U(1)$-gauge, field. To fit this into our General Framework would require "splicing" the two bundles together into a single bundle on which both objects may be thought to live. We will return to these issues in the next section.

We will adopt the notation of Section 2.1. Thus, X is an open submanifold of ^{1,3} (standard coordinates x^0, x^1, x^2, x^3) and $\boldsymbol{\omega}$ is a connection on a principal $U(1)$-bundle

$$U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$$

over X with curvature $\boldsymbol{\Omega}$. For any cross-section s we write

$$s^*\boldsymbol{\omega} = \boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_\alpha \, dx^\alpha = -\boldsymbol{i}\,\boldsymbol{A} = -\boldsymbol{i} A_\alpha \, dx^\alpha$$

and $$s^*\boldsymbol{\Omega} = \boldsymbol{\mathcal{F}} = \frac{1}{2}\mathcal{F}_{\alpha\beta}\, dx^\alpha \wedge dx^\beta = -\boldsymbol{i}\,\boldsymbol{F} = -\frac{1}{2}\boldsymbol{i} F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta,$$

where $\mathcal{F}_{\alpha\beta} = \partial_\alpha \mathcal{A}_\beta - \partial_\beta \mathcal{A}_\alpha = -\boldsymbol{i}(\partial_\alpha A_\beta - \partial_\beta A_\alpha)$. Now we let $\mathcal{V} =$ (as a 2-dimensional real vector space). The usual inner product on $= \;^2$ is given by

$$\langle z_1, z_2 \rangle = \frac{1}{2}(z_1 \bar{z}_2 + z_2 \bar{z}_1).$$

For each integer n we define a representation $\rho_n : U(1) \longrightarrow GL(\;)$ by

$$\rho_n(g)(z) = g \cdot z = g^n z,$$

where we identify an element of $U(1)$ with a complex number of modulus 1 (these are, in fact, all of the "irreducible" representations of $U(1)$ on). Notice that ρ_n is orthogonal with respect to $\langle \;,\; \rangle$ since

$$\begin{aligned}
\langle \rho_n(g)(z_1), \rho_n(g)(z_2) \rangle &= \langle g^n z_1, g^n z_2 \rangle \\
&= \frac{1}{2}\left((g^n z_1)\left(\overline{g^n z_2}\,\right) + (g^n z_2)\left(\overline{g^n z_1}\,\right) \right) \\
&= \frac{1}{2}\left(g^n (z_1 \bar{z}_2) g^{-n} + g^n (z_2 \bar{z}_1) g^{-n} \right) \\
&= \frac{1}{2}\left(z_1 \bar{z}_2 + z_2 \bar{z}_1 \right) \\
&= \langle z_1, z_2 \rangle .
\end{aligned}$$

The potential function $U : \quad \longrightarrow \quad$ (#7 of the General Framework) is taken to be

$$U(z) = \frac{1}{2}m\langle z, z\rangle = \frac{1}{2}m|\, z\,|^2 = \frac{1}{2}mz\bar{z}$$

where $m \geq 0$ is a constant. Since ρ_n is orthogonal with respect to $\langle\ ,\ \rangle$, U is invariant under the action of $U(1)$ on , as required.

A matter field of type ρ_n is a -valued map ϕ on P that is equivariant with respect to the actions of $U(1)$ on P and , i.e., that satisfies

$$\phi(p \cdot g) = g^{-n}\phi(p)$$

for each $p \in P$ and $g \in U(1)$. The connection $\boldsymbol{\omega}$ determines a covariant exterior derivative $d^{\boldsymbol{\omega}}\phi$ of ϕ and the action $A(\boldsymbol{\omega}, \phi)$ contains an appropriate squared norm $\|\, d^{\boldsymbol{\omega}}\phi\|^{\,2}$ arising from the Minkowski metric on X and the invariant inner product $\langle\ ,\ \rangle$ on . We will defer until later the general procedure for constructing such norms and simply indicate at this point what the result is for the case under consideration and why it is the "natural" choice. For any local cross-section s, $s^*\phi = \phi \circ s$ and, for convenience, we will write $\phi = \phi(x^0, x^1, x^2, x^3)$ for the standard coordinate representation for $s^*\phi$. The corresponding coordinate expression for the pullback of $d^{\boldsymbol{\omega}}\phi$ is

$$(\partial_\alpha \phi - \boldsymbol{i}n\, A_\alpha \phi)\, dx^\alpha$$

(Example #1, page 52). Each $\partial_\alpha\phi - \boldsymbol{i}n\, A_\alpha\phi$ is a function with values in . The invariant inner product $\langle\ ,\ \rangle$ on and the Minkowski inner product $(g(v, w) = v^0w^0 - v^1w^1 - v^2w^2 - v^3w^3 = \eta_{\alpha\beta}v^\alpha w^\beta = \eta^{\alpha\beta}v_\alpha w_\beta$, where $v_\alpha = \eta_{a\alpha}v^a$ and $w_\beta = \eta_{b\beta}w^b)$ combine to give the squared norm

$$\begin{aligned}
\|\, d^{\boldsymbol{\omega}}\phi\,\|^2 &= \eta^{\alpha\beta}(\partial_\alpha\phi - \boldsymbol{i}nA_\alpha\phi)\left(\overline{\partial_\beta\phi - \boldsymbol{i}nA_\beta\phi}\right) \\
&= \eta^{\alpha\beta}(\partial_\alpha\phi - \boldsymbol{i}nA_\alpha\phi)\ (\partial_\beta\bar{\phi} + \boldsymbol{i}nA_\beta\bar{\phi}) \quad (A_\beta \text{ is real}) \\
&= \ (\partial_\alpha\phi - \boldsymbol{i}nA_\alpha\phi)\ (\eta^{\alpha\beta}(\partial_\beta\bar{\phi}) + \boldsymbol{i}n(\eta^{\alpha\beta}A_\beta)\bar{\phi}) \\
&= (\partial_\alpha\phi - \boldsymbol{i}nA_\alpha\phi)\ (\partial^\alpha\bar{\phi} + \boldsymbol{i}nA^\alpha\bar{\phi}),
\end{aligned}$$

where we have written

$$\partial^\alpha = \eta^{\alpha\beta}\partial_\beta, \qquad \alpha = 0, 1, 2, 3$$

(so that $\partial^0 = \partial_0,\ \partial^1 = -\partial_1,\ \partial^2 = -\partial_2$ and $\partial^3 = -\partial_3$) and

$$A^\alpha = \eta^{\alpha\beta}A_\beta, \qquad \alpha = 0, 1, 2, 3.$$

Although the expression we have given for $\|d^{\boldsymbol{\omega}}\phi\|^2$ is local and appears to depend on the choice of the cross-section s, it is, in fact, gauge invariant and therefore determines a globally defined, real-valued function on X. To see this we observe the following: We have already seen (page 59) that a gauge transformation g can be written $g(x) = e^{-\boldsymbol{i}\Lambda(x)}$ and has the following effects on the potential $\boldsymbol{A}$ and the matter field ϕ:

$$\boldsymbol{A} \longrightarrow \boldsymbol{A}^g = \boldsymbol{A} + d\Lambda$$
$$\phi \longrightarrow \phi^g = g^{-1} \cdot \phi = e^{\boldsymbol{i}\,n\Lambda}\phi.$$

Thus,

$$\begin{aligned}\partial_\alpha \phi^g - \boldsymbol{i}n(A^g)_\alpha \phi^g &= \partial_\alpha(e^{\boldsymbol{i}\,n\Lambda}\phi) - \boldsymbol{i}n(A_\alpha + \partial_\alpha \Lambda)(e^{\boldsymbol{i}\,n\Lambda}\phi) \\ &= e^{\boldsymbol{i}\,n\Lambda}\partial_\alpha \phi + \boldsymbol{i}ne^{\boldsymbol{i}\,n\Lambda}(\partial_\alpha \Lambda)\phi - \boldsymbol{i}nA_\alpha e^{\boldsymbol{i}\,n\Lambda}\phi \\ &\qquad - \boldsymbol{i}n(\partial_\alpha \Lambda)e^{\boldsymbol{i}\,n\Lambda}\phi \\ &= e^{\boldsymbol{i}\,n\Lambda}(\partial_\alpha \phi - \boldsymbol{i}nA_\alpha \phi)\end{aligned}$$

and, similarly,

$$\partial^\alpha\, \overline{\phi^g} + \boldsymbol{i}n(A^g)^\alpha \overline{\phi^g} = e^{-\boldsymbol{i}\,n\Lambda}(\partial^\alpha \bar{\phi} + \boldsymbol{i}nA^\alpha \bar{\phi}).$$

Consequently,

$$\begin{aligned}&(\partial_\alpha \phi^g - \boldsymbol{i}n(A^g)_\alpha \phi^g)\left(\partial^\alpha \overline{\phi^g} + \boldsymbol{i}n(A^g)^\alpha \overline{\phi^g}\right) \\ &\quad = (\partial_\alpha \phi - \boldsymbol{i}nA_\alpha \phi)\left(\partial^\alpha \bar{\phi} + \boldsymbol{i}nA^\alpha \bar{\phi}\right)\end{aligned}$$

as required. With this and an appropriate choice of normalizing and coupling constants, we take our action functional to be

$$\begin{aligned}A(\boldsymbol{\omega}, \phi) = \int_X \Big[&-\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} + \frac{1}{2}(\partial_\alpha \phi - \boldsymbol{i}nA_\alpha \phi)\;(\partial^\alpha \bar{\phi} + \boldsymbol{i}nA^\alpha \bar{\phi}) \\ &+ \frac{1}{2}m\phi\bar{\phi}\Big]\, dx^0 dx^1 dx^2 dx^3.\end{aligned}$$

The corresponding Euler-Lagrange equations are

$$\begin{aligned}(\partial_\alpha - \boldsymbol{i}nA_\alpha)\;(\partial^\alpha - \boldsymbol{i}nA^\alpha)\,\phi + m^2\phi = 0& \\ \partial_\alpha F^{\alpha\beta} = 0, \quad \beta = 0,1,2,3&\end{aligned}$$

(see [**Bl**]). The first of these is the **Klein-Gordon equation** (for a spin zero particle of mass m and charge n interacting with a gauge field $\boldsymbol{\mathcal{F}} = -\boldsymbol{i}\boldsymbol{F} = -\boldsymbol{i}d\boldsymbol{A}$ determined by the local gauge potentials $\boldsymbol{\mathcal{A}} = -\boldsymbol{i}A_\alpha dx^\alpha$). The second equation is equivalent to $d^*\boldsymbol{F} = 0$ (see page 62). Since $-\boldsymbol{i}\boldsymbol{F}$ is the pullback of a curvature form, the Bianchi identity gives $d\boldsymbol{F} = 0$ also so $\boldsymbol{F}$ satisfies Maxwell's equations.

Remark: Of course, it was our stated intention to model a spin zero particle in an electromagnetic field, but the point here is that the electromagnetic nature of the field $\boldsymbol{F}$ that appears in the action $A(\boldsymbol{\omega}, \phi)$ is necessitated by the Euler-Lagrange equations. We do not have to impose Maxwell's equations "by hand." Also note that conjugating the Klein-Gordon equation gives

$$(\partial_\alpha + \boldsymbol{i}nA_\alpha)(\partial^\alpha + \boldsymbol{i}nA_\alpha)\bar{\phi} + m^2\bar{\phi} = 0.$$

Thus, if ϕ represents a Klein-Gordon field of mass m and charge n, $\bar{\phi}$ represents a Klein-Gordon field of mass m and charge $-n$. Physicists view ϕ and $\bar{\phi}$ as wavefunctions for a particle/antiparticle pair.

Taking $\boldsymbol{A}$ to be zero above (i.e., "turning off" the electromagnetic field) gives the **free Klein-Gordon equation**

$$\partial_\alpha \partial^\alpha \phi + m^2\phi = 0$$

for a spin zero particle.

Remark: Free objects in physics are also subject to field equations. For example, a free particle of mass m in Newtonian mechanics satisfies $\frac{d}{dt}(m\vec{v}) = \vec{0}$, while in quantum mechanics its wavefunction ψ satisfies the Schroedinger equation

$$-\frac{1}{2m}\vec{\nabla}^2\psi = \boldsymbol{i}\frac{\partial\psi}{\partial t}.$$

Note also that if ϕ satisfies the free Klein-Gordon equation, then so does $\bar{\phi}$.

We will not pursue the business of seeking solutions to these equations and sorting out their physical significance. Indeed, it would seem that no plausible physical interpretation of such solutions exists outside the context of quantum field theory and for this the only service we can perform for our reader is a referral to the physics literature (e.g., [**Gui**], [**Ry**], or [**Wein**]). We will, however, spend a few moments describing an alternative approach to the Klein-Gordon equation which more clearly exposes the philosophical underpinnings of modern gauge theory. Physicists refer to this approach as "minimal coupling" and it begins with the free particle equation

$$\partial_\alpha \partial^\alpha \phi + m^2\phi = 0.$$

This equation itself might be arrived at by "quantizing" the classical relativistic relation $E^2 = \vec{p}^{\,2} + m^2$ between energy and momentum ("quantization" is the mystical process of replacing classical quantities such as these with "corresponding operators" that act on the wavefunction). One then "couples" the particle to the field by replacing the "ordinary derivatives" ∂_α by "covariant derivatives" $\partial_\alpha - \boldsymbol{i}nA_\alpha$ involving the potentials A_α for the field. This, of course, does give the full Klein-Gordon equation, but the motivation is no doubt obscure. Our search for motivation leads to the very early days of quantum mechanics.

In old-fashioned (non-relativistic) quantum mechanics a charged particle (of mass m and charge n) in an electromagnetic field has a wavefunction ψ that satisfies the Schroedinger equation

$$\frac{1}{2m}\left(-\boldsymbol{i}\vec{\nabla} - n\vec{A}\right)^2 \psi = \left(\boldsymbol{i}\frac{\partial}{\partial t} - nV\right)\psi, \tag{2.3.1}$$

where $(V, \vec{A}) = (A^0, A^1, A^2, A^3) = (A_0, -A_1, -A_2, -A_3)$ and $\boldsymbol{A} = A_\alpha dx^\alpha$ satisfies $d\boldsymbol{A} = \boldsymbol{F}$ (see page 63).

Remark: The meaning of $(-\boldsymbol{i}\vec{\nabla} - n\vec{A})^2$ as an operator on ψ is as follows:

$$\begin{aligned}\left(-\boldsymbol{i}\vec{\nabla} - n\vec{A}\right)^2 \psi &= \left(-\boldsymbol{i}\vec{\nabla} - n\vec{A}\right)\cdot\left(-\boldsymbol{i}\vec{\nabla} - n\vec{A}\right)\psi \\ &= \left(-\boldsymbol{i}\vec{\nabla} - n\vec{A}\right)\cdot\left(-\boldsymbol{i}\vec{\nabla}\psi - n\,\psi\vec{A}\right) \\ &= -\vec{\nabla}^2\psi + n\boldsymbol{i}\vec{\nabla}\cdot(\psi\vec{A}) + n\boldsymbol{i}\vec{A}\cdot(\vec{\nabla}\psi) \\ &\qquad + n^2|\,\vec{A}\,|^2\psi.\end{aligned}$$

Now, ψ takes values in and so, at each point, has a modulus and a phase ($\psi = re^{\boldsymbol{i}\theta}$). There is some arbitrariness in the phase, however, since, if a is an element of $U(1)$ (identified with a complex number of modulus 1), then $a\psi$ satisfies (2.3.1) whenever ψ does (being a constant, a just slips outside of all the derivatives in (2.3.1)). Moreover, $a\psi$ differs from ψ only in the phase factor (since $|a| = 1$) and so $|a\psi|^2 = |\psi|^2$. Since all of the physically significant probabilities in quantum mechanics depend only on this squared modulus, ψ and $a\psi$ should represent the same physical object. This freedom to alter the phase of ψ is quite restricted, however. Since a must be constant, any phase shift in the wavefunction must be accomplished at all spatial locations simultaneously. Such a global phase shift, however, violates both the spirit and the letter of relativistic law (you can't do anything "at all spatial locations simultaneously"). Nevertheless, it is difficult to shake the feeling that the physical significance of ψ "should" persist under some sort of phase shift (again, because squared moduli will be unaffected). Notice that the relativistic objection to a phase shift would disappear if one allowed the phase to shift independently at each spacetime point, i.e., if one replaced ψ by $a\psi$, where a is now a function of (x, y, z, t) taking values in $U(1)$. The problem with this, of course, is that, if a is not constant, it will not simply "slip outside" of all the derivatives in (2.3.1). Indeed, product rules will generate all sorts of new terms that do not cancel so there is no reason to suppose that $a\psi$ will even be a solution to (2.3.1). A bit (actually, quite a bit) of vector calculus will, in fact, establish the following: Let ψ be a solution to (2.3.1) and let $\Lambda(x, y, z, t)$ be any smooth, real-valued function. Then

$$\psi' = e^{\boldsymbol{i}\,n\,\Lambda}\psi$$

is a solution to

$$\frac{1}{2m}\left(-\boldsymbol{i}\vec{\nabla} - n\left(\vec{A} + \vec{\nabla}\Lambda\right)\right)^2 \psi' = \left(\boldsymbol{i}\frac{\partial}{\partial t} - n\left(V - \frac{\partial \Lambda}{\partial t}\right)\right)\psi'. \tag{2.3.2}$$

This last result is interesting from a number of different points of view. Physicists in the early part of this century found in it some confirmation of their long held belief that the nonuniqueness of the potential for an electromagnetic field and the nonuniqueness of the phase for a wavefunction were both matters of no physical consequence. The reasoning went something like this: In equation (2.3.1) the electromagnetic field to which ψ is responding is described by the potential $(V, \vec{A})$. The corresponding 1-form is $\boldsymbol{A} = A_\alpha dx^\alpha$, where $A_0 = V$ and $(A_1, A_2, A_3) = -\vec{A}$ (see page 63). In equation (2.3.2), $(V, \vec{A})$ is replaced by $(V - \frac{\partial \Lambda}{\partial t}, \vec{A} + \vec{\nabla}\Lambda)$ and the corresponding 1-form is $\boldsymbol{A} - d\Lambda$ which, of course, describes the same electromagnetic field $\boldsymbol{F} = d\boldsymbol{A} = d(\boldsymbol{A} - d\Lambda)$ since $d^2 = 0$ (see page 63). Classically, the potential was regarded as a convenient computational device, but with no physical significance of its own outside of the fact that it gives rise via differentiation to the electromagnetic field. Thus, (2.3.1) and (2.3.2) should, in some sense, be "equivalent" (describe the same physics). Now, the solutions to (2.3.1) and (2.3.2) are in one-to-one correspondence ($\psi \leftrightarrow e^{\boldsymbol{i}\,n\Lambda}\psi$) and the corresponding functions differ only by the phase factor $e^{\boldsymbol{i}\,n\Lambda}$. Since $|\psi|^2 = |e^{\boldsymbol{i}\,n\Lambda}\psi|^2$ at each point, all of the usual probabilities calculated for ψ and $e^{\boldsymbol{i}\,n\Lambda}\psi$ in quantum mechanics are the same so that these should be two descriptions of the same physical object. Everything seems to fit together quite nicely.

This placid scene was disturbed in the late 1950s when Aharonov and Bohm [**AB**] suggested that, while the phase of a single charge may well be unmeasurable, the *relative* phase of two charged particles that interact should have observable consequences. Their proposed experiment (later confirmed by Chambers in 1960) is described on page 6 of [**N4**]. The potential had now to be taken seriously as a physical field and not merely a mathematical contrivance. This being the case, one is no longer free to regard a phase shift $\psi \longrightarrow e^{\boldsymbol{i}\,n\Lambda}\psi$ as devoid of physical content. Indeed, physicists have gone to quite the other extreme and elevated the invariance of electrodynamics under such local phase transformations to the status of a basic physical principle (this is an instance of the so-called "gauge principle" which lies at the heart of modern gauge theory). To see more clearly the consequences of such a principle we consider the special cases of (2.3.1) and (2.3.2) in which the electromagnetic field is "turned off."

If the electromagnetic field is zero one may choose the potential $\boldsymbol{A}$ for which $(V, \vec{A}) = (0, \vec{0})$ so that (2.3.1) becomes the usual free Schroedinger equation

$$\frac{1}{2m}(-\boldsymbol{i}\vec{\nabla})^2\,\psi = \boldsymbol{i}\frac{\partial}{\partial t}\,\psi. \tag{2.3.3}$$

On the other hand, the same (trivial) electromagnetic field is described by any potential of the form $(0 - \frac{\partial \Lambda}{\partial t}, \vec{0} + \vec{\nabla}\Lambda)$. Begging the indulgence of the reader we would now like to call this $(V, \vec{A})$ and write (2.3.2) as

$$\frac{1}{2m}\left(-\boldsymbol{i}\,\vec{\nabla} - n\,\vec{A}\right)^2 \psi' = \left(\boldsymbol{i}\frac{\partial}{\partial t} - nV\right)\psi'. \tag{2.3.4}$$

Now suppose one adopts as a basic physical principle that electrodynamics should be invariant under local phase shifts. In particular, then equations (2.3.3) and (2.3.4), with $(V, \vec{A}) = (-\frac{\partial \Lambda}{\partial t}, \vec{\nabla}\Lambda)$, are equivalent. Physicists take the following rather remarkable view of this: Even in a vacuum (electromagnetic field zero) the requirement of local phase shift invariance (gauge invariance) necessitates the existence of a physical field $\boldsymbol{A}$ (not the electromagnetic field – that's zero – but the gauge potential field) whose task is to counteract, or balance, the effects of any phase shift and keep the physics invariant. Algebraically, this happens by adding to the operators that appear in the Schroedinger equation terms whose sole purpose is to cancel the extra stuff you get from product rules for $e^{\boldsymbol{i}\,n\,\Lambda}\psi$ when Λ is not constant. The marvelous thing about this way of viewing the situation is that it provides a completely mindless way of ensuring gauge invariance in all sorts of contexts. One need only fudge into the differential operators whatever it takes to cancel the offending extra terms that arise from the product rule. In electrodynamics it works just as well when the field is not turned off. Indeed, "turning the field on," i.e., coupling a particle to an electromagnetic field, can be viewed in exactly the same light. A free particle satisfies (2.3.3). To couple the particle to an electromagnetic field $\boldsymbol{F}$, write $\boldsymbol{F} = d\boldsymbol{A}$, where $\boldsymbol{A} = A_\alpha\, dx^\alpha$, $(A_0, A_1, A_2, A_3) = (A^0, -A^1, -A^2, -A^3) = (V, \vec{A})$ and make the substitutions

$$-\boldsymbol{i}\,\vec{\nabla} \longrightarrow -\boldsymbol{i}\,\vec{\nabla} - n\,\vec{A} \quad \text{and} \quad \boldsymbol{i}\frac{\partial}{\partial t} \longrightarrow \boldsymbol{i}\frac{\partial}{\partial t} - n\,V$$

to obtain

$$\frac{1}{2m}\left(-\boldsymbol{i}\,\vec{\nabla} - n\,\vec{A}\right)^2 \psi = \left(\boldsymbol{i}\frac{\partial}{\partial t} - n\,V\right)\psi$$

which is (2.3.1). The solutions to (2.3.1) will describe the wavefunction of the charged particle coupled to the field $\boldsymbol{F}$. The description is only one among many possible descriptions, each differing from the others by a choice of phase (gauge) determined by the particular $\boldsymbol{A}$ one has chosen for the representation $\boldsymbol{F} = d\boldsymbol{A}$.

Finally, we show that the operator substitutions described above assume a more elegant (and more familiar) form if we recast them in relativistic notation. Here we let

$$(\partial_0, \partial_1, \partial_2, \partial_3) = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right) = \left(\frac{\partial}{\partial t}, \vec{\nabla}\right)$$

and

$$(\partial^0, \partial^1, \partial^2, \partial^3) = (\partial_0, -\partial_1, -\partial_2, -\partial_3) = \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)$$

so

$$\left(\boldsymbol{i}\frac{\partial}{\partial t} - n\,V,\ -\boldsymbol{i}\,\vec{\nabla} - n\,\vec{A}\right) = \boldsymbol{i}\,\left(\frac{\partial}{\partial t} + \boldsymbol{i}\,n\,V, -\vec{\nabla} + \boldsymbol{i}\,n\,\vec{A}\right)$$
$$= \boldsymbol{i}\,\big(\partial^0 + \boldsymbol{i}\,n\,V,\ \partial^1 + \boldsymbol{i}\,n\,A^1,$$
$$\partial^2 + \boldsymbol{i}\,n\,A^2, \partial^3 + \boldsymbol{i}\,n\,A^3\big)\,.$$

Since

$$\left(\boldsymbol{i}\,\frac{\partial}{\partial t}, -\boldsymbol{i}\vec{\nabla}\right) = \boldsymbol{i}\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right) = \boldsymbol{i}\,\left(\partial^0, \partial^1, \partial^2, \partial^3\right),$$

the substitutions above simply amount to

$$\partial^\alpha \longrightarrow \partial^\alpha + \boldsymbol{i}\,n\,A^\alpha, \quad \alpha = 0, 1, 2, 3,$$

or, lowering the indices once again,

$$\partial_\alpha \longrightarrow \partial_\alpha + \boldsymbol{i}\,n\,A_\alpha, \quad \alpha = 0, 1, 2, 3$$

(the sign is + rather than − because of our decision to include the conventional minus sign in $\boldsymbol{\mathcal{A}} = -\boldsymbol{i}\,\boldsymbol{A}$).

We will conclude our discussion of the Klein-Gordon equation by describing its two most important invariance properties: gauge invariance and Lorentz invariance. We have already shown that the action $A(\boldsymbol{\omega}, \phi)$ is invariant under gauge transformations and it follows that the same is true of its Euler-Lagrange equations. Nevertheless, a direct proof is instructive. Thus, we consider the equation

$$(\partial_\alpha - \boldsymbol{i}\,n\,A_\alpha)\,(\partial^\alpha - \boldsymbol{i}\,n\,A^\alpha)\phi + m^2\phi = 0 \tag{2.3.5}$$

and a gauge transformation $g(x) = e^{-\boldsymbol{i}\,\Lambda(x)}$. The corresponding changes in the potential and the matter field are, as usual,

$$\boldsymbol{A} \longrightarrow \boldsymbol{A}^g = \boldsymbol{A} + d\,\Lambda$$

and

$$\phi \longrightarrow \phi^g = g^{-1} \cdot \phi = e^{\boldsymbol{i}\,n\,\Lambda}\phi.$$

Our objective is to show that, if ϕ satisfies (2.3.5), then

$$(\partial_\alpha - \boldsymbol{i}\,n\,(A^g)_\alpha)\,(\partial^\alpha - \boldsymbol{i}\,n\,(A^g)^\alpha)\,\phi^g + m^2\,\phi^g = 0. \tag{2.3.6}$$

First, we compute

$$\begin{aligned}(\partial^\alpha - \boldsymbol{i}\,n\,(A^g)_\alpha)\,\phi^g &= (\partial^\alpha - \boldsymbol{i}\,n\,(A^\alpha + \partial^\alpha \Lambda))\,\left(e^{\boldsymbol{i}\,n\,\Lambda}\phi\right)\\ &= \partial^\alpha\left(e^{\boldsymbol{i}\,n\Lambda}\phi\right) - \boldsymbol{i}\,n\,A^\alpha\,e^{\boldsymbol{i}\,n\,\Lambda}\phi - \boldsymbol{i}\,n\,\partial^\alpha\Lambda\,e^{\boldsymbol{i}\,n\,\Lambda}\phi\\ &= e^{\boldsymbol{i}\,n\,\Lambda}\partial^\alpha\,\phi + \boldsymbol{i}\,n\,e^{\boldsymbol{i}\,n\,\Lambda}\partial^\alpha\Lambda\,\phi\\ &\qquad - \boldsymbol{i}\,n\,A^\alpha\,e^{\boldsymbol{i}\,n\,\Lambda}\,\phi - \boldsymbol{i}\,n\,\partial^\alpha\Lambda\,e^{\boldsymbol{i}\,n\,\Lambda}\phi\\ &= e^{\boldsymbol{i}\,n\,\Lambda}(\partial^\alpha\,\phi - \boldsymbol{i}\,n\,A^\alpha\,\phi)\\ &= e^{\boldsymbol{i}\,n\,\Lambda}(\partial^\alpha - \boldsymbol{i}\,n\,A^\alpha)\,\phi.\end{aligned}$$

In the same way,

$$(\partial_\alpha - \boldsymbol{i}\,n\,(A^g)_\alpha)\left(e^{\boldsymbol{i}\,n\,\Lambda}\phi\right) = e^{\boldsymbol{i}\,n\,\Lambda}(\partial_\alpha - \boldsymbol{i}\,n\,A^\alpha)\,\phi$$

so

$$\begin{aligned}&(\partial_\alpha - \boldsymbol{i}\,n\,(A^g)_\alpha)\,(\partial^\alpha - \boldsymbol{i}\,n\,(A^g)^\alpha)\,\phi^g\\ &\quad = (\partial_\alpha - \boldsymbol{i}\,n\,(A^g)_\alpha)\left(e^{\boldsymbol{i}\,n\,\Lambda}(\partial^\alpha - \boldsymbol{i}\,n\,A^\alpha)\,\phi\right)\\ &\quad = e^{\boldsymbol{i}\,n\,\Lambda}\,(\partial_\alpha - \boldsymbol{i}\,n\,A_\alpha)(\partial^\alpha - \boldsymbol{i}\,n\,A^\alpha)\,\phi.\end{aligned}$$

From this it is clear that (2.3.5) implies (2.3.6). The Klein-Gordon equation is gauge invariant.

We wish to show next that the Klein-Gordon equation is "relativistically invariant." Roughly, this means that the equation has the same mathematical form in all inertial frames of reference, but the precise meaning of such a statement in general will require some discussion. In this section we will be content to spell out explicitly what is being asserted for the Klein-Gordon equation alone. When we turn to the Dirac equation in the next section we will describe precisely what is meant by "relativistic invariance" in general.

Thus far we have written the Klein-Gordon equation (2.3.5) only in standard coordinates x^0, x^1, x^2, x^3 for ^{1,3}, i.e., only in one fixed inertial frame of reference. To emphasize this fact we will now write the derivatives ∂_α and ∂^α explicity as $\partial/\partial x^\alpha$ and $\eta^{\alpha\beta}\,\partial/\partial x^\beta$ so that (2.3.5) becomes

$$\begin{aligned}\eta^{\alpha\beta}\,\left(\frac{\partial}{\partial x^\alpha} - \boldsymbol{i}\,n\,A_\alpha(x^0,\ldots,x^3)\right)\left(\frac{\partial}{\partial x^\beta} - \boldsymbol{i}\,n\,\boldsymbol{A}_\beta(x^0,\right.\\ \left.\ldots,x^3)\right)\phi\,(x^0,\ldots,x^3) + m^2\,\phi\,(x^0,\ldots,x^3) = 0.\end{aligned} \tag{2.3.7}$$

The basic postulate of Special Relativity is that one inertial frame of reference is as good as another. The coordinates y^0, y^1, y^2, y^3 for ^{1,3}, supplied by another such frame of reference are assumed to be related to x^0, x^1, x^2, x^3 by

$$y^\alpha = \Lambda^\alpha{}_\beta x^\beta, \quad \alpha = 0, 1, 2, 3,$$

where $\Lambda = (\Lambda^{\alpha}{}_{\beta})$ is an element of the so-called "proper, orthochronous Lorentz group" $\mathcal{L}_+^{\uparrow} = \{\Lambda : \Lambda^{\top}\, \eta\, \Lambda = \eta,\ \det \Lambda = 1, \Lambda^0{}_0 \geq 1\}$, where η is on page 58.

Remark: The "general Lorentz group" $\mathcal{L}$ consists of those Λ that satisfy $\Lambda^{\top}\, \eta\, \Lambda = \eta$, and these relate arbitrary orthonormal bases in $\quad^{1,3}$. The "orthochronous" condition $\Lambda^0{}_0 \geq 1$ ensures that Λ does not reverse time orientations, i.e., $x^0 \geq 0$ implies $y^0 = \Lambda^0{}_{\alpha}\, x^{\alpha} \geq 0$, for timelike or null vectors, while $\det \Lambda = 1$ ("proper") then guarantees that Λ does not reverse the orientation of the spatial coordinates (x^1, x^2, x^3).

The inverse of Λ is written $\Lambda^{-1} = (\Lambda_{\alpha}{}^{\beta})$ so that

$$x^{\beta} = \Lambda_{\alpha}{}^{\beta}\, y^{\alpha}, \quad \beta = 0, 1, 2, 3.$$

Furthermore, one has

$$\Lambda^{\alpha}{}_{\gamma}\, \Lambda^{\beta}{}_{\delta}\, \eta^{\gamma\delta} = \eta^{\alpha\beta}, \quad \alpha, \beta = 0, 1, 2, 3$$

and

$$\Lambda_{\alpha}{}^{\gamma}\, \Lambda_{\beta}{}^{\delta}\, \eta^{\alpha\beta} = \eta^{\gamma\delta}, \quad \gamma, \delta = 0, 1, 2, 3$$

(physical motivation for and basic properties of the Lorentz group are discussed in some detail in the Introduction and first three sections of Chapter 1 in [**N3**]).

The precise assertion we are making about the Klein-Gordon equation is as follows: If $\phi = \phi(x^0, \ldots, x^3)$ is a solution to (2.3.7) and if we define $\hat{\phi} = \hat{\phi}(y^0, \ldots, y^3)$ by

$$\hat{\phi}\,(y^0, \ldots, y^3) = \phi\,(\Lambda_{\alpha}{}^{0}\, y^{\alpha}, \ldots, \Lambda_{\alpha}{}^{3}\, y^{\alpha}),$$

then $\hat{\phi}$ satisfies

$$\eta^{\alpha\beta} \left(\frac{\partial}{\partial y^{\alpha}} - \boldsymbol{i}\, n \hat{A}_{\alpha}(y^0, \ldots, y^3) \right) \left(\frac{\partial}{\partial y^{\beta}} - \boldsymbol{i}\, n \hat{A}_{\beta}(y^0, \ldots, y^3) \right) \hat{\phi}\,(y^0, \ldots, y^3) + m^2 \hat{\phi}\,(y^0, \ldots, y^3) = 0. \tag{2.3.8}$$

where $\boldsymbol{A} = \hat{A}_{\alpha}\, dy^{\alpha}$ so that $\hat{A}_{\alpha} = \boldsymbol{A}\left(\frac{\partial}{\partial y^{\alpha}}\right) = \boldsymbol{A}\left(\frac{\partial}{\partial x^{\beta}}\, \frac{\partial x^{\beta}}{\partial y^{\alpha}}\right) = \Lambda_{\alpha}{}^{\beta} \boldsymbol{A}\left(\frac{\partial}{\partial x^{\beta}}\right) = \Lambda_{\alpha}{}^{\beta} A_{\beta}$.

Remark: The essential, but rather camouflaged, issue here is that it is up to us to *specify* what the wavefunction is to be in the new coordinates (i.e., how ϕ transforms under $\mathcal{L}_+^{\uparrow}$) and that, in order to satisfy the requirements of special relativity, we must do this in such a way that it satisfies "the same equation in the new coordinate system." In this case, the wavefunction transforms trivially (physicists say, as a scalar), i.e., we simply rewrite ϕ in the new coordinates, but there is no reason to expect such simplicity in general (compare this, for example, with the transformation of a vector field on ^{3} under rotation of the coordinate system).

Having spelled out exactly what we mean by relativistic invariance in this case, the proof is just a simple calculation. First observe that

$$\begin{aligned}
&\left(\frac{\partial}{\partial y^\beta} - \boldsymbol{i}\, n\hat{A}_\beta(y^0,\ldots,y^3)\right)\hat{\phi}\,(y^0,\ldots,y^3)\\
&\quad = \frac{\partial}{\partial y^\beta}\,\hat{\phi}\,(y^0,\ldots,y^3) - \boldsymbol{i}\, n\hat{A}_\beta\,(y^0,\ldots,y^3)\,\hat{\phi}\,(y^0,\ldots,y^3)\\
&\quad = \frac{\partial}{\partial y^\beta}\,\phi\,({\Lambda_a}^0\, y^a,\ldots) - \boldsymbol{i}\, n\hat{A}_\beta\,(y^0,\ldots)\,\phi\,({\Lambda_a}^0\, y^a,\ldots)\\
&\quad = \frac{\partial}{\partial x^\delta}\,\phi\,({\Lambda_a}^0 y^a,\ldots)\frac{\partial x^\delta}{\partial y^\beta} - \boldsymbol{i}\, n\,{\Lambda_\beta}^\delta A_\delta\,({\Lambda_a}^0 y^a,\ldots)\,\phi\,({\Lambda_a}^0 y^a,\ldots)\\
&\quad = {\Lambda_\beta}^\delta\left(\frac{\partial}{\partial x^\delta} - \boldsymbol{i}\, nA_\delta({\Lambda_a}^0\, y^a,\ldots)\right)\phi\,({\Lambda_a}^0\, y^a,\ldots).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\left(\frac{\partial}{\partial y^\alpha} - \boldsymbol{i}\, n\hat{A}_\alpha(y^0,\ldots,y^3)\right)\left(\frac{\partial}{\partial y^\beta} - \boldsymbol{i}n\hat{A}_\beta(y^0,\ldots,y^3)\right)\hat{\phi}\,(y^0,\ldots,y^3)\\
&\quad = {\Lambda_\alpha}^\gamma\,{\Lambda_\beta}^\delta\left(\frac{\partial}{\partial x^\gamma} - \boldsymbol{i}\, nA_\gamma\ ({\Lambda_a}^0\, y^a,\ldots)\right)\left(\frac{\partial}{\partial x^\delta} - \boldsymbol{i}\, nA_\delta({\Lambda_a}^0\, y^a,\ldots)\right)\\
&\qquad\qquad \times\phi\,({\Lambda_a}^0\, y^a,\ldots).
\end{aligned}$$

Since $\eta^{\alpha\beta}\,{\Lambda_\alpha}^\gamma\,{\Lambda_\beta}^\delta = \eta^{\gamma\delta}$, we have

$$\begin{aligned}
&\eta^{\alpha\beta}\left(\frac{\partial}{\partial y^\alpha} - \boldsymbol{i}\, n\hat{A}_\alpha(y^0,\ldots,y^3)\right)\left(\frac{\partial}{\partial y^\beta} - \boldsymbol{i}\, n\hat{A}_\beta(y^0,\ldots,y^3)\right)\hat{\phi}(y^0,\ldots,y^3)\\
&\quad = \eta^{\gamma\delta}\left(\frac{\partial}{\partial x^\gamma} - \boldsymbol{i}\, nA_\gamma(x^0,\ldots,x^3)\right)\left(\frac{\partial}{\partial x^\delta} - \boldsymbol{i}\, nA_\delta(x^0,\ldots,x^3)\right)\phi(x^0,\ldots,x^3)
\end{aligned}$$

and from this it is clear that (2.3.7) implies (2.3.8).

2.4 Spin One-Half Electrodynamics

Electrons, protons and neutrons (unlike the π mesons of Section 2.3) have spin $s = \frac{1}{2}$ and so, according to the generally accepted scheme of (nonrelativistic) quantum mechanics, should have a wavefunction with $2\left(\frac{1}{2}\right) + 1 = 2$ components. We intend to say just a few words on the rationale behind this and then refer those who are interested to the elegant, and quite accessible, account of these phenomena in Volume III of *The Feynman Lectures on Physics* [**Fey**].

The classical Bohr picture of an atom (negatively charged electrons revolving around a positively charged nucleus) suggests that an orbiting electron actually constitutes a tiny current loop. Such a current loop produces a magnetic field which, at large distances, is the same as that of a magnetic dipole (located at the center of the loop and perpendicular to the plane of the loop). Such a dipole has a magnetic moment (a vector describing its orientation and strength). Consequently, an electron in an atom has associated with it an "orbital magnetic moment." Now, magnetic moments behave in interesting and well-understood ways when subjected to external magnetic fields. In 1922 (just before the advent of quantum mechanics), Stern and Gerlach carried out an experiment designed to detect these effects for the orbital magnetic moment of an electron. From the point of view of classical physics (the only point of view available at the time), the results were quite shocking. Somewhat later, the quantum mechanics of Schroedinger and Heisenberg, when applied to the orbital magnetic moment of the electron, provided a qualitative, but not quantitative explanation of the outcome. Finally, it was suggested by Uhlenbeck and Goudsmit that this discrepancy (and various others associated with the anomalous Zeeman effect and the splitting of certain spectral lines) could be accounted for if one assumed that the electron had associated with it an additional magnetic moment, not arising from its orbital motion, but rather from a sort of "intrinsic" angular momentum or "spinning" of the electron. The suggestion was not that an electron actually spins on some axis in the same way that the earth does on its, but rather that it possesses some intrinsic property (called "spin") that manifests itself in an external magnetic field by mimicing the behavior of the magnetic moment of a spinning charged ball. This intrinsic magnetic moment vector, however, must be of a rather peculiar sort that one could only encounter in quantum mechanics. The Stern-Gerlach experiment suggested that its component in *any* spatial direction could take on only one of two possible values (± 1 with the proper choice of units). This has the following consequence. Let us select (arbitrarily) some direction in space (say, the z-direction of some coordinate system). The intrinsic magnetic moment of an electron has, at each point, a z-component σ_z that can take on one of the two values ± 1. Which value it has will determine how the electron responds to certain magnetic fields and so a complete description of the electron's wavefunction must contain this information. More precisely, the wavefunction must be regarded as a function of not only x, y, z and t, but of σ_z as well.

$$\psi = \psi(x, y, z, t, \sigma_z)$$

However, since σ_z can assume only the two values ± 1, such a wavefunction is equivalent to a *pair* of functions $\psi_1(x, y, z, t) = \psi(x, y, z, t, 1)$ and $\psi_2(x, y, z, t) = \psi(x, y, z, t, -1)$. It is convenient to put these two together into a column vector and adopt the point of view that an electron (or any spin one-half particle) has a two-component wavefunction

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

The probabilistic interpretations of the wavefunctions in quantum mechanics now run as follows: For each fixed t and each region $R \subseteq {}^3$, $\int_R \psi_1 \bar{\psi}_1$ is the probability at time t that the particle will be detected in R with its spin vector directed in the *positive* z-direction; $\int_R \psi_2 \bar{\psi}_2$ is the probability that it will be found in R with its spin vector in the *negative* z-direction; $\int_R \psi_1 \bar{\psi}_1 + \psi_2 \bar{\psi}_2$ is the probability that it will be found in R at all.

Pauli formulated a theory of the electron along the lines suggested above and, although this work was later superceded by that of Dirac, it provides an instructive warm-up and we will spend a few moments outlining some of its essential features. We will consider only stationary states $\Psi(x, y, z, t) = \psi(x, y, z)\, e^{-\boldsymbol{i}\omega t}$ and will focus our attention on the spatial part $\psi\,(x, y, z)$ (a proper treatment of time dependence should be relativistic, which Pauli's theory was not). Thus, we are interested in the two-component object

$$\psi\,(x, y, z) = \begin{pmatrix} \psi_1\,(x, y, z) \\ \psi_2\,(x, y, z) \end{pmatrix}.$$

Here x, y and z presumably represent "standard" coordinates in ^{3} and we (along with Pauli) will require that our theory be independent of which particular oriented, orthonormal basis for ^{3} gives rise to these coordinates, i.e., that it be invariant under the rotation group $SO(3)$. We are therefore led to ask the following question: How is the two-component wavefunction $\left(\begin{smallmatrix} \psi_1 \\ \psi_2 \end{smallmatrix}\right)$ transformed if the coordinate system is subjected to the rotation corresponding to some element of $SO(3)$?

To answer this question let us first review the manner in which such issues are addressed in classical physics and then decide what, if any, modifications are required by quantum mechanics. If the coordinate system is subjected to a rotation $R \in SO(3)$, then the state of the electron will be described by a pair $\left(\begin{smallmatrix} \psi'_1 \\ \psi'_2 \end{smallmatrix}\right)$, where each ψ'_i is a function of the new coordinates x', y' and z'. Now, it is conceivable that each ψ'_i is simply ψ_i expressed in terms of these new coordinates. This is, indeed, what we found to be the case for the Klein-Gordon equation (see page 83). However, it is also conceivable that the dependence of $\left(\begin{smallmatrix} \psi'_1 \\ \psi'_2 \end{smallmatrix}\right)$ on $\left(\begin{smallmatrix} \psi_1 \\ \psi_2 \end{smallmatrix}\right)$ is more analogous to the transformation law for an ordinary vector, or tensor field on ^{3}. We will not prejudge this issue here, but will simply make a few tentative assumptions about this dependence. We assume, for example, that ψ'_1 and ψ'_2 are linear functions of ψ_1 and ψ_2. The reason is that, presumably, any differential equations one might arrive at for the wavefunction will be generalizations of the (linear) Schroedinger equation and should (at least as a first guess) themselves be linear. Thus, for some 2×2 complex matrix $T = T(R)$ we have

$$\begin{pmatrix} \psi_1' \\ \psi_2' \end{pmatrix} = T(R) \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

(here it is understood that both sides have been written in terms of one of the two coordinate systems, $x,\ y,\ z$ or $x',\ y'$ and z'). Assuming also (for the moment) that the wavefunction is uniquely determined in each coordinate system we find that a rotation by $R_2 \in SO(3)$ followed by a rotation by $R_1 \in SO(3)$ must have the same effect as the rotation $R_1 R_2 \in SO(3)$. Thus, we must have

$$T(R_1 R_2) = T(R_1) T(R_2).$$

Similarly, each $T(R)$ must be invertible and satisfy

$$T(R^{-1}) = (T(R))^{-1}$$

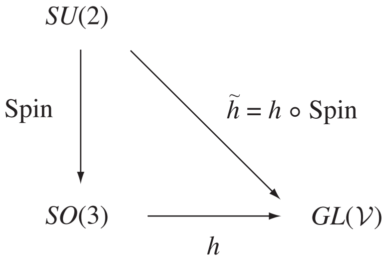
What we find then is that the rule T which associates with every $R \in SO(3)$ the corresponding transformation matrix $T(R)$ for $\left(\begin{smallmatrix} \psi_1 \\ \psi_2 \end{smallmatrix} \right)$ is a homomorphism into the group of invertible, 2×2, complex matrices. Identifying this latter group with $GL(\ ^2)$ we find that T is a representation of $SO(3)$ on $\ ^2$. The reason this information is useful is that all of the representations of $SO(3)$ are known. These are usually described somewhat indirectly as follows: In Appendix A of [**N4**] it is shown that $SU(2)$ is the (double) covering group of $SO(3)$. More precisely, there exists a smooth, surjective group homomorphism

$$\text{Spin} : SU(2) \longrightarrow SO(3)$$

with kernel $\pm \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ and with the property that each point of $SO(3)$ has an open neighborhood V whose inverse image under Spin is a disjoint union of (two) open sets in $SU(2)$, each of which is mapped diffeomorphically onto V by Spin. Now, consider a representation

$$h : SO(3) \longrightarrow GL(\mathcal{V})$$

of $SO(3)$. Composing with Spin then gives a representation of $SU(2)$. Every representation of $SO(3)$ "comes from" a representation of $SU(2)$. The converse is not true, however. That is, a given representation $\tilde{h} : SU(2) \longrightarrow GL(\mathcal{V})$ of $SU(2)$ will not induce a representation of $SO(3)$ unless $\tilde{h}(-g) = \tilde{h}(g)$ for every $g \in SU(2)$. The representations of $SU(2)$ that do *not* satisfy this condition are sometimes referred to in the physics literature as "2-valued representations of $SO(3)$," although they are not representations of $SO(3)$ at all, of course.



Now, it is quite easy to write out some rather obvious representations of $SU(2)$. Let $[z_1, z_2]$ be the vector space of all polynomials with complex coefficients in the two unknowns z_1 and z_2. For each $k = 0, 1, \ldots$, let $\mathcal{V}_k$ be the subspace consisting of those polynomials that are homogeneous of degree k:

$$c_0 {z_1}^k + c_1 {z_1}^{k-1} z_2 + \ \cdots \ + c_k {z_2}^k.$$

A basis for $\mathcal{V}_k$ consists of all polynomials ${z_1}^{k-r} {z_2}^r$, $r = 0, 1, \ldots, k$. Each $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $SU(2)$ gives rise to a linear transformation on $\mathcal{V}_k$ which carries ${z_1}^{k-r} {z_2}^r$ onto $(z_1')^{k-r} (z_2')^r$, where

$$\begin{pmatrix} z_1' \\ z_2' \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

These linear transformations are clearly invertible and so the assignment, to each $g \in SU(2)$, of the corresponding element of $GL(\mathcal{V}_k)$ is a representation of $SU(2)$, usually denoted

$$D^{\frac{k}{2}} : SU(2) \longrightarrow GL(\mathcal{V}_k),$$

and called the **spin-j representation**, where $j = \frac{k}{2}$. One can show (see [**vdW**]) that each of these representations is *irreducible* (i.e., that there is no proper subspace of $\mathcal{V}_k$ that is invariant under every $D^{\frac{k}{2}}(g)$, $g \in SU(2)$) and that every irreducible representation of $SU(2)$ with complex representation space is equivalent to one of these (two representations $D_1 : G \longrightarrow GL(\mathcal{V}_1)$ and $D_2 : G \longrightarrow GL(\mathcal{V}_2)$ of a group G are *equivalent* if it is possible to choose bases for $\mathcal{V}_1$ and $\mathcal{V}_2$ so that, for each $g \in G$, the matrices of $D_1(g)$ and $D_2(g)$ are the same). Furthermore, any representation of $SU(2)$ can be constructed from these irreducible representations by forming finite direct sums (the *direct sum* of $D_1 : G \longrightarrow GL(\mathcal{V}_1)$ and $D_2 : G \longrightarrow GL(\mathcal{V}_2)$ is the representation $D_1 \oplus D_2 : G \longrightarrow GL(\mathcal{V}_1 \oplus \mathcal{V}_2)$ defined by $(D_1 \oplus D_2)(g)(v_1, v_2) = (D_1(g)(v_1), D_2(g)(v_2))$. In effect, we now have all of the representations of $SU(2)$.

The polynomials have now served their purpose and it will be convenient to note that $\mathcal{V}_k$ has complex dimension $k+1$ and so can be identified with ${}^{k+1}$ by identifying $z_1{}^{k-r} z_2{}^r$, $r = 0, 1, \ldots, k$, with the standard basis for ${}^{k+1}$. The linear transformations $D^{\frac{k}{2}}(g)$ can therefore be identified with $(k+1) \times (k+1)$ complex matrices. For example, $k = 0$ gives the trivial representation of $SU(2)$ on $(D^0(g) = (1)$ for each $g \in SU(2))$, while $k = 1$ gives the identity representation of $SU(2)$ on ^{2} $(D^{\frac{1}{2}}(g) = g$ for every $g \in SU(2))$. Note that $D^{\frac{1}{2}}$ is the only irreducible representation of $SU(2)$ on ^{2}. The only other way to get a representation of $SU(2)$ on ^{2} is to form the direct sum of two copies of D^0:

$$(D^0 \oplus D^0)(g) = (1) \oplus (1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This, of course, leaves everything in ^{2} alone. Finally notice that $D^{\frac{1}{2}}(-g) = -D^{\frac{1}{2}}(g)$ and $(D^0 \oplus D^0)(-g) = (D^0 \oplus D^0)(g)$ so only this second example descends to a representation of $SO(3)$ (the trivial representation of $SO(3)$ on ^{2}).

The situation we have just described would seem to present us with something of a dilemma. To ensure the rotational invariance of Pauli's theory of the electron we were led to seek a representation of $SO(3)$ on ^{2} that would transform the two-component wavefunctions $\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ when the coordinate system is rotated. We find now that there is only one such $(D^0 \oplus D^0)$ and this is the trivial representation. Under this representation the transformed wavefunction $\begin{pmatrix} \psi_1' \\ \psi_2' \end{pmatrix}$ would simply be $\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ written in terms of the new coordinates. However, this is clearly not consistent with the phenomenon (spin one-half) which led us to two-component wavefunctions in the first place. Recall that $\psi_1(x,y,z) = \psi(x,y,z,1)$ and $\psi_2(x,y,z) = \psi(x,y,z,-1)$, where ± 1 are the possible z-components of the intrinsic magnetic moment of the electron. A rotation which reverses the direction of the z-axis must interchange ψ_1 and ψ_2 and so cannot leave $\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ unchanged. Thus, $D^0 \oplus D^0$ is not consistent with the structure we are attempting to model. Must we conclude then that Pauli's proposal is doomed to failure?

To extricate ourselves from this dilemma we must understand that there is an essential feature of quantum mechanics that requires an adjustment in the classical picture we painted earlier (pages 86–87). Our conclusion that the transformation matrices $T(R)$ satisfy $T(R_1 R_2) = T(R_1)T(R_2)$ and $(T(R))^{-1} = T(R^{-1})$ and therefore give rise to a representation of $SO(3)$ followed from the assumption that the wavefunction is uniquely determined in each coordinate system. This, however, is not (quite) the case. For example, both $\pm \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ represent the same state of our electron since an overall sign change has no effect on the probabilities described earlier (page 86) and all of the physical content of the wavefunction is contained in such probabilities. It follows, in particular, that for a given $R \in SO(3)$, the transformation matrix $T(R)$ is determined only up to sign. Physicists would call T a "2-valued representation" of $SO(3)$. This makes no sense, of course, but we have just seen exactly how one can make sense of it. The appropriate tactic is to "go to the covering space," i.e., to represent a rotation, not by an element of $SO(3)$, but rather by two elements of $SU(2)$ and seek the transformation law for $\binom{\psi_1}{\psi_2}$ among the representations of $SU(2)$. No longer being constrained to select a representation of $SU(2)$ that descends to $SO(3)$, we have available one more option, i.e., $D^{\frac{1}{2}}$. With this choice each $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $SU(2)$ would transform the wavefunction $\binom{\psi_1}{\psi_2}$ as follows:

$$\begin{pmatrix} \psi_1' \\ \psi_2' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}.$$

Notice, in particular, that if $g = \begin{pmatrix} 0 & -\boldsymbol{i} \\ -\boldsymbol{i} & 0 \end{pmatrix}$, then Spin $(\pm g)$ is the rotation about the x-axis through π (Appendix A, [**N4**]) and therefore reverses the z-axis. Since

$$\begin{pmatrix} 0 & -\boldsymbol{i} \\ -\boldsymbol{i} & 0 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = -\boldsymbol{i} \begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix}$$

and $-\boldsymbol{i}\binom{\psi_2}{\psi_1}$ represents the same state as $\binom{\psi_2}{\psi_1}$, the representation $D^{\frac{1}{2}}$, unlike $D^0 \oplus D^0$, is at least consistent with our proposed model of spin one-half.

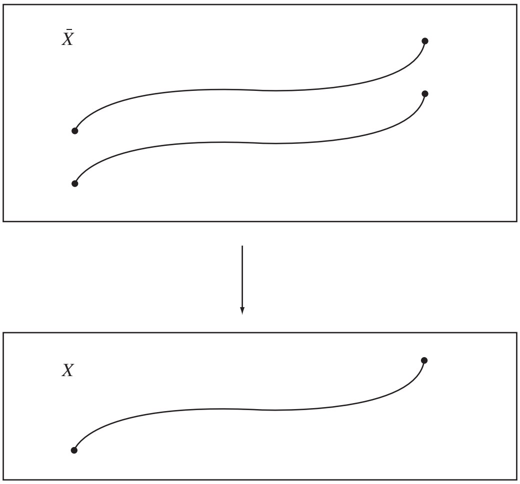
The program we have just described can leave one with the queasy feeling of ambiguity. Suppose that one is given a frame in ^{3} and wants to rotate by $R \in SO(3)$ to a new frame and thereby a new representation of the wavefunction. If Spin $(\pm g) = R$, then $D^{\frac{1}{2}}(g)\binom{\psi_1}{\psi_2}$ and $D^{\frac{1}{2}}(-g)\binom{\psi_1}{\psi_2} = -D^{\frac{1}{2}}(g)\binom{\psi_1}{\psi_2}$ are physically equivalent and that's fine, but it is difficult not to ask oneself, "But, really, which is it?" The answer, interestingly enough, resides in the topologies of $SU(2)$, $SO(3)$ and the spinor map. One is forced to regard a rotation of frames in ^{3} not as an instantaneous jump from one to the other, but as a physical process that begins with one frame and continuously rotates the axes to the new position. The transformation law for the wavefunction depends not only on the end result of the rotation, but also on "how you got there." For example, the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}$$

represents a rotation through t radians about the x-axis. As t varies from 0 to 2π one has a continuous sequence of rotations (i.e., a curve $R_1(t)$ in

$SO(3)$) that represents the physical process of rotating a frame through one complete turn (360°) about its x-axis (identify each element of $SO(3)$ with the configuration of the axes that would result from applying that rotation to the initial configuration). The curve $R_2(t)$ in $SO(3)$ defined by the same formula, but with $0 \le t \le 4\pi$ represents a rotation about the x-axis through 720°. Both R_1 and R_2 begin and end with the same configuration, but there is a real difference, both physically and mathematically.

Spin : $SU(2) \longrightarrow SO(3)$ is a covering space (Exercise A.13, [**N4**]) and covering spaces have the property that curves in the covered space lift uniquely to curves in the covering space once an initial point is selected (Corollary 1.5.13, [**N4**]).



In particular, given a curve in $SO(3)$ (representing a continuous rotation of one frame into another) and a choice of either g or $-g$ above the initial point (frame), there is a uniquely determined element of $SU(2)$ above the terminal point that represents the transformation law that gives the wavefunction in the new frame. No ambiguity at all!

For the curves R_1 and R_2 in $SO(3)$, both of which begin and end at the identity, one can write out the lifts g_1 and g_2 starting at $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ explicitly. Both are given by

$$\begin{pmatrix} \cos\frac{t}{2} & -\,\boldsymbol{i}\sin\frac{t}{2} \\ -\boldsymbol{i}\sin\frac{t}{2} & \cos\frac{t}{2} \end{pmatrix}$$

but with $0 \le t \le 2\pi$ for g_1 and $0 \le t \le 4\pi$ for g_2 (see Appendix A, [**N4**]). But notice that g_1 is a path in $SU(2)$ from $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ to $-\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, whereas g_2 begins and ends at $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$. Thus, a rotation of the frame through 360° changes the sign of the wavefunction, but a rotation through 720° leaves the sign unchanged, even though both rotations begin and end with the same configuration of the axes. Mathematically, the difference between R_1 and R_2 is that they represent two different homotopy classes in $\pi_1(SO(3)) \cong \mathbb{Z}_2$. R_2 is nullhomotopic since it is Spin $\circ g_2$ and g_2 is a loop at $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ in $SU(2) \cong S^3$, but R_1 is not because it lifts to a path g_1 from $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ to $-\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, in $SU(2)$.

The next step in this program would be to look at the differential equations proposed by Pauli for $\begin{pmatrix}\psi_1\\\psi_2\end{pmatrix}$ and decide whether or not they assume the same form when the coordinate system is rotated and the wavefunction is transformed by $D^{\frac{1}{2}}$ (they do!). Since Pauli's theory was eventually abandoned (because it is not relativistically invariant) we shall not pursue this here, but will instead turn to the profoundly successful alternative proposed by Dirac.

Remark: Before moving forward with our discussion of spin one-half we observe that the arguments we presented (on pages 85–86) for modeling such a particle with a wavefunction having two components really had nothing specific to do with spin one-half. The essential feature that led to the doubling of the number of components was the existence of an "internal structure" that could be represented by a parameter (σ_z for the electron) that could assume precisely two values. There are other examples of this sort of thing, e.g., the "isotopic spin" parameter of a nucleon which determines (in the absence of electromagnetic fields) whether the particle is a proton or a neutron. The two components of the nucleon wavefunction then represent the "proton part" and the "neutron part" of the doublet. Indeed, it was this example that provided the initial motivation for Yang-Mills theory ([**YM**]).

Dirac set out to construct a relativistically invariant equation that would be satisfied by the wavefunction of a spin one-half particle. He reasoned that his equation should, in some sense, "imply" the Klein-Gordon equation since, as we have noted, this is just the quantized version of the relativistic energy-momentum relation $E^2 = \vec{p}^{\,2} + m^2$. However, he also sought to remedy certain physical problems associated with the appearance of the second time derivative in the Klein-Gordon equation (for a discussion of these see Chapter 6 of [**Hol**]). Thus, Dirac sought a first order linear equation which, upon iteration, yielded the Klein-Gordon equation. Somewhat more precisely, Dirac was in the market for a first order differential operator

$$\not{D} = \gamma^0\partial_0 + \gamma^1\partial_1 + \gamma^2\partial_2 + \gamma^3\partial_3 = \gamma^\alpha\partial_\alpha \tag{2.4.1}$$

($\partial_\alpha = \frac{\partial}{\partial x^\alpha}$, $\alpha = 0, 1, 2, 3$) such that, when applied to the equation

$$\not{D}\phi = -\boldsymbol{i}m\phi, \tag{2.4.2}$$

the result is the free Klein-Gordon equation

$$\eta^{\alpha\beta}\partial_\alpha\partial_\beta\phi = -m^2\phi. \tag{2.4.3}$$

The problem is to determine the γ^α, $\alpha = 0, 1, 2, 3$, so that this will be the case. Apply the operator $\not{D}$ in (2.4.1) to both sides of (2.4.2).

$$\begin{aligned} \not{D}\,(\not{D}\,\phi) &= \not{D}\,(-\boldsymbol{i}m\phi) \\ (\gamma^\alpha\partial_\alpha)\,(\gamma^\beta\partial_\beta\phi) &= -\boldsymbol{i}m\,\not{D}\phi \\ \gamma^\alpha\gamma^\beta\partial_\alpha\,(\partial_\beta\phi) &= -\boldsymbol{i}m\,(-\boldsymbol{i}m\phi) \\ (\gamma^\alpha\gamma^\beta\partial_\alpha\partial_\beta)\,\phi &= -m^2\phi \end{aligned} \tag{2.4.4}$$

This will agree with (2.4.3) if

$$\gamma^\alpha\gamma^\beta\partial_\alpha\partial_\beta = \eta^{\alpha\beta}\partial_\alpha\partial_\beta,$$

i.e., if

$$\gamma^\alpha\gamma^\beta = \eta^{\alpha\beta}, \quad \alpha, \beta = 0, 1, 2, 3. \tag{2.4.5}$$

Now, (2.4.5) clearly cannot be satisfied if the γ^α are taken to be numbers (none can be 0 since $\eta^{\alpha\alpha} = \pm 1$, but $\gamma^\alpha\gamma^\beta = 0$ if $\alpha \neq \beta$). Dirac's idea was to allow ϕ to have more than one complex component and interpret (2.4.5) as matrix equations (one for each $\alpha, \beta = 0, 1, 2, 3$). Specifically, if

$$\phi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_n \end{pmatrix}$$

and $\partial_\beta\phi$ is computed entrywise, then $\not{D}\phi = \gamma^\alpha\partial_\alpha\phi$ would require the γ^α to have n columns (the reason we do not immediately follow Pauli's lead and take $n = 2$ will become clear shortly). To iterate the operator and define $\not{D}(\not{D}\phi)$ requires that the γ^α have n rows as well. Now, (2.4.5) must be interpreted as matrix equations

$$\gamma^\alpha\gamma^\beta = \eta^{\alpha\beta}\,\mathrm{id}, \quad \alpha, \beta = 0, 1, 2, 3,$$

where id is the $n \times n$ identity matrix. Notice, however, that each $\eta^{\alpha\beta}$ id is a symmetric matrix. To accommodate this fact we observe that, because $\partial_\alpha\partial_\beta\phi = \partial_\beta\partial_\alpha\phi$, (2.4.4) can be written

$$\frac{1}{2}\left(\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha\right)\partial_\alpha\partial_\beta\phi = -m^2\phi$$

so that we might just as well have written (2.4.5) as

$$\frac{1}{2}\left(\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha\right) = \eta^{\alpha\beta}$$

and so the matrix conditions we are currently trying to satisfy may be taken to be

$$\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\eta^{\alpha\beta}\mathrm{id}, \quad \alpha, \beta = 0, 1, 2, 3. \tag{2.4.6}$$

Finding square matrices that satisfy (2.4.6) is actually a problem familiar to algebraists. What we are looking for here is a matrix representation of the Clifford algebra of $(\ ^4, \langle\ ,\ \rangle)$, where $\langle\ ,\ \rangle$ is the Minkowski inner product. There are many solutions, but the smallest n for which such matrices can be found is $n = 4$. Thus, Pauli's choice of $n = 2$ cannot succeed in this context (we will eventually have to sort out how to reconcile this with the arguments we presented earlier to the effect that a spin one-half wavefunction should have two components). It is, in fact, easy to write down a set of 4×4 matrices satisfying conditions (2.4.6). Letting

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\boldsymbol{i} \\ \boldsymbol{i} & 0 \end{pmatrix}, \text{ and } \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

be the Pauli spin matrices and taking σ_0 to be the 2×2 identity matrix, one easily verifies the usual commutation relations

$$\begin{aligned} \sigma_i^2 &= \sigma_0, & i &= 1, 2, 3 \\ \sigma_i \sigma_j &= -\sigma_j \sigma_i, & i, j &= 1, 2, 3, \quad i \neq j. \end{aligned} \tag{2.4.7}$$

Now, we define

$$\gamma^0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\gamma^1 = \begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\gamma^2 = \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \boldsymbol{i} \\ 0 & 0 & -\boldsymbol{i} & 0 \\ 0 & -\boldsymbol{i} & 0 & 0 \\ \boldsymbol{i} & 0 & 0 & 0 \end{pmatrix}$$

$$\gamma^3 = \begin{pmatrix} 0 & -\sigma_3 \\ \sigma_3 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

It is now a simple matter to verify that the conditions in (2.4.6) are satisfied by these matrices, e.g.,

$$
\begin{aligned}
&\gamma^1\gamma^2 + \gamma^2\gamma^1 \\
&\quad = \begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix}\begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix} \\
&\quad = \begin{pmatrix} -\sigma_1\sigma_2 & 0 \\ 0 & -\sigma_1\sigma_2 \end{pmatrix} + \begin{pmatrix} -\sigma_2\sigma_1 & 0 \\ 0 & -\sigma_2\sigma_1 \end{pmatrix} \\
&\quad = \begin{pmatrix} -(\sigma_1\sigma_2 + \sigma_2\sigma_1) & 0 \\ 0 & -(\sigma_1\sigma_2 + \sigma_2\sigma_1) \end{pmatrix} \\
&\quad = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} [4pt] = 2\eta^{12}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
\gamma^1\gamma^1 + \gamma^1\gamma^1 = 2\gamma^1\gamma^1 &= 2\begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix} \\
&= 2\begin{pmatrix} -\sigma_1^2 & 0 \\ 0 & -\sigma_1^2 \end{pmatrix} = -2\begin{pmatrix} \sigma_0 & 0 \\ 0 & \sigma_0 \end{pmatrix} \\
&= 2\eta^{11}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$

etc.

Remark: There are many other possible choices for $\gamma^0,\ \gamma^1,\ \gamma^2$ and γ^3, many of which are used in the physics literature. Indeed, for any nonsingular matrix B, one can replace each γ^α by $B\gamma^\alpha B^{-1}$ and obtain another set of "Dirac matrices" satisfying (2.4.6). Conversely, one can show (see, e.g., pages 104–106 of [**Gre**]) that any set of 4×4 matrices satisfying (2.4.6) differs from our choice by such a similarity transformation. Algebraically, this means that, up to equivalence, there is only one representation of the Clifford algebra of $\;^{1,3}$ by 4×4 matrices. The choice we have made is called the **Weyl**, or **chiral representation**.

With $\gamma^0,\ \gamma^1,\ \gamma^2$ and γ^3 the 4×4 matrices described above, the wavefunction for our spin one-half particle has four components

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{pmatrix}$$

and the **Dirac equation** (expressed in standard coordinates on ^{1,3}) is

$$\not{D}\phi = -\boldsymbol{i}m\phi$$
$$\gamma^\alpha \partial_\alpha \phi = -\boldsymbol{i}m\phi$$

$$\begin{pmatrix} 0 & 0 & \partial_0 - \partial_3 & -\partial_1 + \boldsymbol{i}\partial_2 \\ 0 & 0 & -\partial_1 - \boldsymbol{i}\partial_2 & \partial_0 + \partial_3 \\ \partial_0 + \partial_3 & \partial_1 - \boldsymbol{i}\partial_2 & 0 & 0 \\ \partial_1 + \boldsymbol{i}\partial_2 & \partial_0 - \partial_3 & 0 & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{pmatrix}$$
$$= -\boldsymbol{i}m \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{pmatrix}.$$

We will discuss the relativistic invariance of the Dirac equation shortly, but first we would like to show that, in the massless ($m = 0$) case, there is a simpler solution to the problem of finding a first order operator $D = \gamma^\alpha \partial_\alpha$ which, when applied to the equation

$$D\phi = 0$$

yields the $m = 0$ Klein-Gordon equation

$$\eta^{\alpha\beta} \partial_\alpha \partial_\beta \phi = 0.$$

For this purpose we write this last equation as

$$\partial_0 \partial_0 \phi = \delta^{ij} \partial_i \partial_j \phi. \tag{2.4.8}$$

Now, $D\phi = 0$ can be written

$$\gamma^0 \partial_0 \phi = -\gamma^i \partial_i \phi,$$

or, assuming γ^0 is invertible,

$$\partial_0 \phi = -\mu^i \partial_i \phi \quad (\mu^i = (\gamma^0)^{-1} \gamma^i, \ i = 1, 2, 3).$$

Then

$$\begin{aligned} \partial_0 \partial_0 \phi &= \partial_0(-\mu^i \partial_i \phi) = -\mu^i \partial_0 \partial_i \phi \\ &= -\mu^i \partial_i \partial_0 \phi = -\mu^i \partial_i (-\mu^j \partial_j \phi) \end{aligned}$$

$$\partial_0 \partial_0 \phi = \mu^i \mu^j \partial_i \partial_j \phi. \tag{2.4.9}$$

Now compare (2.4.8) and (2.4.9). Again, $\mu^i\mu^j = \delta^{ij}$ cannot be satisfied by numbers so we rewrite (2.4.9) as

$$\partial_0\partial_0\phi = \frac{1}{2}(\mu^i\mu^j + \mu^j\mu^i)\partial_i\partial_j\phi, \tag{2.4.10}$$

and seek matrices satisfying

$$\mu^i\mu^j + \mu^j\mu^i = 2\delta^{ij}\ \mathrm{id}, \quad i, j = 1, 2, 3. \tag{2.4.11}$$

In the terminology of algebra, matrices satisfying the conditions (2.4.11) constitute a matrix representation of the Clifford algebra of $(\ \ ^3, \langle\ ,\ \rangle)$, where $\langle\ ,\ \rangle$ is the usual positive definite inner product on $\ \ ^3$. Notice that the conditions (2.4.11) coincide with the commutation relations (2.4.7) for the Pauli spin matrices σ_1, σ_2 and σ_3. In particular, there is now a 2×2 solution to the problem so, in the massless case, our wavefunction can have two components

$$\phi = \begin{pmatrix} \phi_3 \\ \phi_4 \end{pmatrix}$$

(the reason for the peculiar numbering will become clear soon). One obtains an operator $D = \gamma^\alpha\partial_\alpha$ of the required type by taking, for example, $\gamma^0 = \sigma_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\gamma^i = \mu^i = -\sigma_i$, $i = 1, 2, 3$. Thus,

$$\begin{aligned} D &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\partial_0 + \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}\partial_1 + \begin{pmatrix} 0 & \boldsymbol{i} \\ -\boldsymbol{i} & 0 \end{pmatrix}\partial_2 + \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\partial_3 \\ &= \begin{pmatrix} \partial_0 - \partial_3 & -\partial_1 + \boldsymbol{i}\partial_2 \\ -\partial_1 - \boldsymbol{i}\partial_2 & \partial_0 + \partial_3 \end{pmatrix}. \end{aligned}$$

The corresponding equation $D\phi = 0$ then becomes

$$\begin{pmatrix} \partial_0 - \partial_3 & -\partial_1 + \boldsymbol{i}\partial_2 \\ -\partial_1 - \boldsymbol{i}\partial_2 & \partial_0 + \partial_3 \end{pmatrix} \begin{pmatrix} \phi_3 \\ \phi_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.4.12}$$

and is known as the **Weyl neutrino equation**. Notice that this is just what one would obtain from the Dirac equation with $m = 0$ and a wave- function of the form

$$\begin{pmatrix} 0 \\ 0 \\ \phi_3 \\ \phi_4 \end{pmatrix}.$$

For future reference we note also that the $m = 0$ Dirac equation for a wavefunction of the form

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ 0 \\ 0 \end{pmatrix}$$

reduces to

$$\begin{pmatrix} \partial_0 + \partial_3 & \partial_1 - \boldsymbol{i}\partial_2 \\ \partial_1 + \boldsymbol{i}\partial_2 & \partial_0 - \partial_3 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.4.13}$$

and that the coefficient matrix in (2.4.13) is the formal conjugate, transposed inverse of the coefficient matrix in (2.4.12). The significance of these observations will emerge in our discussion of the transformation properties of the wavefunctions and the corresponding invariance properties of the equations.

Now we return to the issue of the Lorentz invariance of the Dirac equation. The problem, as it was for the Klein-Gordon equation (page 82), is to show that the Dirac equation has the same form in any other coordinate system $y^0,\ y^1,\ y^2,\ y^3$ for $\ ^{1,3}$, related to the standard coordinates $x^0,\ x^1,\ x^2,\ x^3$ by

$$y^\alpha = \Lambda^\alpha{}_\beta x^\beta, \qquad \alpha = 0, 1, 2, 3,$$

where $\Lambda = (\Lambda^\alpha{}_\beta) \in \mathcal{L}_+^\uparrow$. However, since the Dirac wavefunction has four complex components, this will require finding a representation $T : \mathcal{L}_+^\uparrow \longrightarrow GL(\ ^4)$ of $\mathcal{L}_+^\uparrow$ on $\ ^4$ which, if taken to be the transformation law for the wavefunction, preserves the form of the Dirac equation (cf., the discussion of the two-component Pauli theory on pages 86–93). As was the case for $SO(3)$ in the Pauli theory, it so happens that all of the representations $\mathcal{L}_+^\uparrow$ are known, that they are most conveniently described in terms of a two- fold covering group of $\mathcal{L}_+^\uparrow$ and that, because of the nature of a quantum mechanical wavefunction, it is actually the representations of the covering group that do *not* descend to $\mathcal{L}_+^\uparrow$ that turn out to be of most interest. We begin with a brief summary of the relevant results (see Chapter 3 for more details).

Identifying $\mathcal{L}_+^\uparrow$ with a subset of $\ ^{16}$ one finds that it is a submanifold diffeomorphic to $SO(3) \times\ ^3$ and so is a 6-dimensional Lie group containing $SO(3)$ as a closed subgroup. We denote by $SL(2,\ )$ the group of 2×2 complex matrices with determinant one. Identifying $SL(2,\ )$ with a subset of $\ ^4 =\ ^8$ one finds that it is a submanifold diffeomorphic to $S^3 \times\ ^3$ and so it is also a 6-dimensional (simply connected) Lie group. Note that $SU(2)$ is a closed subgroup of $SL(2,\ )$. Now, we have already described a two-fold covering map

$$\text{Spin}: SU(2) \longrightarrow SO(3)$$

and we wish now to show that this is, in fact, the restriction to $SU(2)$ of a two-fold covering map of $SL(2,\ )$ onto $\mathcal{L}_+^\uparrow$, also denoted

$$\text{Spin}: SL(2,\ ) \longrightarrow \mathcal{L}_+^\uparrow.$$

The construction of this map is carried out in detail in Section 1.7 of [**N3**] so we will be brief. $\ ^{1,3}$ can be identified with the linear space $\mathcal{H}$ of 2×2

complex Hermitian matrices

$$x = \begin{pmatrix} x^0 + x^3 & x^1 - \boldsymbol{i}x^2 \\ x^1 + \boldsymbol{i}x^2 & x^0 - x^3 \end{pmatrix}$$
$$= x^0\sigma_0 + x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3 = x^\alpha\sigma_\alpha,$$

where the squared norm is taken to be the determinant. Observe that each of the coordinates x^α can be expressed as

$$x^\alpha = \frac{1}{2}\text{trace}(\sigma_\alpha x)$$

so that

$$x = \sum_{\alpha=0}^{3} \frac{1}{2}\text{trace}(\sigma_\alpha x)\sigma_\alpha.$$

Now, for each $g \in SL(2, \ )$ we define a map $\Lambda_g : \mathcal{H} \longrightarrow \mathcal{H}$ by

$$\Lambda_g(x) = gx\bar{g}^\top.$$

Note that $\Lambda_g(x)$ is, indeed, in $\mathcal{H}$ because $\overline{\Lambda_g(x)}^\top = (\overline{gx\bar{g}^\top})^\top = (\bar{g}\bar{x}g^\top)^\top = gx\bar{g}^\top = \Lambda_g(x)$. Also note that, for $g \in SU(2) \subseteq SL(2, \ )$, $\bar{g}^\top = g^{-1}$. Now, Λ_g is surely linear and satisfies $\det(\Lambda_g(x)) = \det(gx\bar{g}^\top) = \det(x)$ so it preserves the Minkowski inner product on $\mathcal{H} = \ ^{1,3}$. Thus, Λ_g is an orthogonal transformation and, with a bit more work (page 77, [**N3**]), one can show that it is proper and orthochronous. Now let

$$\Lambda_g(x) = y^\alpha\sigma_\alpha.$$

Then

$$\begin{aligned} y^\alpha &= \frac{1}{2}\text{trace}(\sigma_\alpha\Lambda_g(x)) = \frac{1}{2}\text{trace}(\sigma_\alpha g x\bar{g}^\top) \\ &= \frac{1}{2}\text{trace}(\sigma_\alpha g(x^\beta\sigma_\beta)\bar{g}^\top) \\ &= \frac{1}{2}\text{trace}(\sigma_\alpha g\sigma_\beta\bar{g}^\top)x^\beta \\ &= \Lambda^\alpha{}_\beta x^\beta \end{aligned}$$

where

$$\Lambda^\alpha{}_\beta = \frac{1}{2}\text{trace}(\sigma_\alpha g\sigma_\beta\bar{g}^\top), \qquad \alpha, \beta = 0, 1, 2, 3.$$

Thus, $(\Lambda^\alpha{}_\beta)_{\alpha,\beta=0,1,2,3}$ is in $\mathcal{L}_+^\uparrow$ and we define Spin: $SL(2, \ ) \longrightarrow \mathcal{L}_+^\uparrow$ by

$$\text{Spin}(g) = (\Lambda^\alpha{}_\beta)_{\alpha,\,\beta\,=\,0,1,2.3}$$

for each $g \in SL(2, \ )$. One then shows that Spin is a two-fold covering group for $\mathcal{L}_+^\uparrow$ and that its restriction to $SU(2)$ agrees with the map of the same name discussed on page 88. Before proceeding we will record one additional fact that we will need shortly. Notice that

$$\Lambda^{\alpha}{}_{\beta} = \frac{1}{2}\text{trace}(\sigma_\alpha g \sigma_\beta \bar{g}^\top) = \frac{1}{2}\text{trace}(\sigma_\beta(\bar{g}^\top \sigma_\alpha g))$$

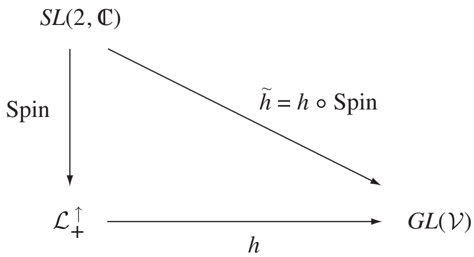
so that

$$\bar{g}^\top \sigma_\alpha g = \sum_{\beta=0}^{3} \Lambda^{\alpha}{}_{\beta}\sigma_\beta. \tag{2.4.14}$$

Now we proceed just as we did for $SO(3)$ in our discussion of the Pauli theory. Consider a representation

$$h : \mathcal{L}_+^\uparrow \longrightarrow GL(\mathcal{V})$$

of $\mathcal{L}_+^\uparrow$. Composing with Spin then gives a representation of $SL(2, \ )$.



Thus, every representation of $\mathcal{L}_+^\uparrow$ comes from a representation of $SL(2, \ )$, but conversely, a representation of $SL(2, \ )$ will descend to a representation of $\mathcal{L}_+^\uparrow$ if and only if it takes the same value at $\pm g$ for each $g \in SL(2, \ )$ (all others are of the "2-valued" variety).

The representations of $SL(2, \ )$ are all known and are described in some detail in Section 3.1 of [**N3**]. We will limit ourselves to a brief discussion of just those items that are relevant to our present context. There are a few obvious representations of $SL(2, \ )$ on $\ ^n$. By sending every element of $SL(2, \ )$ to the $n \times n$ identity matrix one obtains the trivial representation which leaves everything in $\ ^n$ fixed. When $n = 2$ one also has the identity representation which sends each $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ in $SL(2, \ )$ to the linear transformation on $\ ^2$ defined by

$$\begin{pmatrix} z^1 \\ z^2 \end{pmatrix} \longrightarrow \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} z^1 \\ z^2 \end{pmatrix} = \begin{pmatrix} \alpha z^1 + \beta z^2 \\ \gamma z^1 + \delta z^2 \end{pmatrix}.$$

This representation is generally denoted

$$D^{(\frac{1}{2},0)} : SL(2, \ ) \longrightarrow GL(\ ^2)$$

and called the **left-handed spinor representation** of $SL(2, \ )$. Identifying the linear transformation $D^{(\frac{1}{2},0)}(g)$ on $\ ^2$ with its matrix relative to the standard basis for $\ ^2$ one can write

$$D^{(\frac{1}{2},0)}(g) = g$$

for each $g \in SL(2,\ )$. Another, somewhat less obvious, representation

$$D^{(0,\frac{1}{2})} : SL(2,\ ) \longrightarrow GL(\ ^2)$$

of $SL(2,\ )$ on $\ ^2$ sends each g to the linear transformation whose matrix relative to the standard basis is the inverse of the conjugate transpose $\bar{g}^\top$ of g (called the **right-handed spinor representation** of $SL\,(2,\ )$). For simplicity we again write

$$D^{(0,\frac{1}{2})}(g) = (\bar{g}^\top)^{-1}.$$

Remarks: The map that sends g to (the linear transformation on $\ ^2$ whose matrix is) $(g^\top)^{-1}$ is also a representation of $SL(2,\ )$, but it is equivalent to $D^{(\frac{1}{2},0)}$. The reason is that, if $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, then, for any $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $SL(2,\ )$,

$$BgB^{-1} = (g^\top)^{-1}.$$

It follows that $D^{(0,\frac{1}{2})}$ is equivalent to the **conjugation representation** $g \longrightarrow \bar{g}$. $D^{(\frac{1}{2},0)}$ and $D^{(0,\frac{1}{2})}$ are, however, *not* equivalent as representations of $SL(2,\ )$. To see this, suppose there were a matrix B such that $BgB^{-1} = (\bar{g}^\top)^{-1}$ for each $g \in SL(2,\ )$. Then, in particular, we would have $\text{trace}(g) = \text{trace}(\bar{g}^\top)^{-1}$ for each $g \in SL(2,\ )$. This, however, is not true for $g = \begin{pmatrix} -2\boldsymbol{i} & 0 \\ 0 & \frac{1}{2}\boldsymbol{i} \end{pmatrix} \in SL(2,\ )$ since $(\bar{g}^\top)^{-1} = \begin{pmatrix} -\frac{1}{2}\boldsymbol{i} & 0 \\ 0 & 2\boldsymbol{i} \end{pmatrix}$. It is interesting to note, however, that $g \longrightarrow g$ and $g \longrightarrow (\bar{g}^\top)^{-1}$ *are* equivalent as representations of $SU(2)$ since there $\bar{g}^\top = g^{-1}$ so $(\bar{g}^\top)^{-1} = g$.

There is a precise sense in which all of the representations of $SL(2,\ )$ can be constructed from $D^{(\frac{1}{2},0)}$ and $D^{(0,\frac{1}{2})}$. Rather than describing this procedure in general we will be content to illustrate it in the only case of real interest to us, i.e., the representations of $SL(2,\ )$ on $\ ^4$. The most obvious way to construct a representation on $\ ^4$ from two representations on $\ ^2$ is by forming their direct sum. For example, we can define the representation

$$D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})} : SL(2,\ ) \longrightarrow GL(\ ^4)$$

by

$$D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}(g) = \begin{pmatrix} g & 0 \\ 0 & (\bar{g}^\top)^{-1} \end{pmatrix},$$

where all of the entries are 2×2 matrices and we are again identifying a linear transformation on ^{4} with its matrix relative to the standard basis. Similarly, one can define the direct sum of any such pair. A somewhat less obvious procedure for building a representation on ^{4} is the tensor product of two representations on ^{2}. For instance, the representation

$$D^{(\frac{1}{2},\frac{1}{2})} = D^{(\frac{1}{2},0)} \otimes D^{(0,\frac{1}{2})} : SL(2, \quad) \longrightarrow GL(\quad{}^4)$$

can be described as follows: For each $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in SL(2, \quad)$, we define

$$\begin{aligned} D^{(\frac{1}{2},\frac{1}{2})}(g) &= \begin{pmatrix} \alpha(\bar{g}^\top)^{-1} & \beta(\bar{g}^\top)^{-1} \\ \gamma(\bar{g}^\top)^{-1} & \delta(\bar{g}^\top)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \alpha\bar{\delta} & -\alpha\bar{\gamma} & \beta\bar{\delta} & -\beta\bar{\gamma} \\ -\alpha\bar{\beta} & \alpha\bar{\alpha} & -\beta\bar{\beta} & \beta\bar{\alpha} \\ \gamma\bar{\delta} & -\gamma\bar{\gamma} & \delta\bar{\delta} & -\delta\bar{\gamma} \\ -\gamma\bar{\beta} & \gamma\bar{\alpha} & -\delta\bar{\beta} & \delta\bar{\alpha} \end{pmatrix}. \end{aligned}$$

Note that $D^{(\frac{1}{2},\frac{1}{2})}(-g) = D^{(\frac{1}{2},\frac{1}{2})}(g)$ so $D^{(\frac{1}{2},\frac{1}{2})}$ descends to a representation of $\mathcal{L}_+^\uparrow$ on ^{4} (also denoted $D^{(\frac{1}{2},\frac{1}{2})}$). One can show that $D^{(\frac{1}{2},\frac{1}{2})}$ is equivalent to the natural (vector) representation of $\mathcal{L}_+^\uparrow$ on ^{1,3}. Similarly, one can define such representations as $D^{(1,0)} = D^{(\frac{1}{2},0)} \otimes D^{(\frac{1}{2},0)}$. One can show that these are irreducible, whereas such things as $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$, of course, are not. From our point of view the important fact is that we have just described *all* of the representations of $SL(2, \quad)$ on ^{4} (up to equivalence). Proving the relativistic invariance of the Dirac equation therefore amounts to searching among these few representations of $SL(2, \quad)$ on ^{4} for one which, if adopted as the transformation law for the 4-component Dirac wavefunction, will lead to a transformed wavefunction that satisfies the same (Dirac) equation in the transformed coordinate system.

Begin with the Dirac equation in standard coordinates $x = (x^0, x^1, x^2, x^3)$ on ^{1,3}.

$$\left(\gamma^\beta \partial_\beta + \boldsymbol{i}m\right)\phi(x) = 0. \tag{2.4.15}$$

Introduce new coordinates $y = (y^0, y^1, y^2, y^3)$ on ^{1,3} by

$$y^\alpha = \Lambda^\alpha{}_\beta x^\beta, \quad \alpha = 0, 1, 2, 3,$$

where $\Lambda = (\Lambda^\alpha{}_\beta) \in \mathcal{L}_+^\uparrow$ and define $\hat{\partial}_\alpha = \frac{\partial}{\partial y^\alpha}$, $\alpha = 0, 1, 2, 3$. Then

$$\partial_\beta = \frac{\partial}{\partial x^\beta} = \frac{\partial y^\alpha}{\partial x^\beta}\frac{\partial}{\partial y^\alpha} = \Lambda^\alpha{}_\beta \hat{\partial}_\alpha.$$

Let $g \in SL(2, \;)$ be such that Spin $(\pm g) = (\Lambda^\alpha{}_\beta)$. Thus,

$$\Lambda^\alpha{}_\beta = \frac{1}{2}\text{trace}\left(\sigma_\alpha g \sigma_\beta \bar{g}^\top\right), \qquad \alpha, \beta = 0, 1, 2, 3.$$

Our objective is to find a representation

$$\rho : SL(2, \;) \longrightarrow GL(\;^4)$$

such that, if

$$\hat{\phi}(y) = \rho(g)(\phi(\Lambda^{-1}y)),$$

then (2.4.15) implies

$$\left(\gamma^\alpha \hat{\partial}_\alpha + \boldsymbol{i}m\right)\hat{\phi}(y) = 0. \tag{2.4.16}$$

We begin by simply rewriting (2.4.15) in the new coordinates

$$\gamma^\beta \partial_\beta \phi + \boldsymbol{i}m\phi = 0$$

$$\gamma^\beta\left(\Lambda^\alpha{}_\beta\, \hat{\partial}_\alpha\right)\left((\rho(g))^{-1}\hat{\phi}(y)\right) + \boldsymbol{i}m(\rho(g))^{-1}\hat{\phi}(y) = 0$$

$$\gamma^\beta(\rho(g))^{-1}\Lambda^\alpha{}_\beta \hat{\partial}_\alpha \hat{\phi}(y) + (\rho(g))^{-1}\left(\boldsymbol{i}m\hat{\phi}(y)\right) = 0.$$

Multiply through by $\rho(g)$ to obtain

$$\rho(g)\gamma^\beta(\rho(g))^{-1}\Lambda^\alpha{}_\beta \hat{\partial}_\alpha \hat{\phi}(y) + \boldsymbol{i}m\hat{\phi}(y) = 0$$

$$\left(\left(\rho(g)\gamma^\beta(\rho(g))^{-1}\Lambda^\alpha{}_\beta\right)\hat{\partial}_\alpha + \boldsymbol{i}m\right)\hat{\phi}(y) = 0$$

which will agree with (2.4.16) if and only if

$$\rho(g)\gamma^\beta(\rho(g))^{-1}\Lambda^\alpha{}_\beta = \gamma^\alpha. \tag{2.4.17}$$

This then is the condition that our representation ρ must satisfy in order to preserve the form of the Dirac equation. We obtain a more convenient form of this condition as follows:

$$(\rho(g))\gamma^\beta(\rho(g))^{-1}\Lambda^\alpha{}_\beta = \gamma^\alpha$$

$$(\rho(g))\gamma^\beta(\rho(g))^{-1} = \Lambda_\alpha{}^\beta \gamma^\alpha$$

$$\gamma^\beta(\rho(g))^{-1} = (\rho(g))^{-1}\Lambda_\alpha{}^\beta \gamma^\alpha$$

$$\gamma^\beta = (\rho(g))^{-1}\Lambda_\alpha{}^\beta \gamma^\alpha (\rho(g))$$

$$\gamma^\beta = \Lambda_\alpha{}^\beta \Big((\rho(g))^{-1}\gamma^\alpha (\rho(g)) \Big)$$

$$\Lambda^\alpha{}_\beta \gamma^\beta = (\rho(g))^{-1}\gamma^\alpha \rho(g)$$

$$(\rho(g))^{-1}\gamma^\alpha \rho(g) = \Lambda^\alpha{}_\beta \gamma^\beta, \quad \alpha = 0,1,2,3. \tag{2.4.18}$$

At this point one need only check each of the representations ρ of $SL(2,\quad)$ on $\;^4$ described earlier in the hope of finding one that satisfies (2.4.18). One's hopes are not dashed. The winner is $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$, as we now show. With $\gamma^0, \gamma^1, \gamma^2$ and γ^3 as on page 96 and

$$\rho(g) = \begin{pmatrix} g & 0 \\ 0 & (\bar{g}^\top)^{-1} \end{pmatrix}$$

we compute

$$(\rho(g))^{-1}\gamma^0 \rho(g) = \begin{pmatrix} 0 & g^{-1}\sigma_0 (\bar{g}^\top)^{-1} \\ \bar{g}^\top \sigma_0 g & 0 \end{pmatrix} \tag{2.4.19}$$

and, for $i = 1,2,3$,

$$(\rho(g))^{-1}\gamma^i \rho(g) = \begin{pmatrix} 0 & -g^{-1}\sigma_i (\bar{g}^\top)^{-1} \\ \bar{g}^\top \sigma_i g & 0 \end{pmatrix}. \tag{2.4.20}$$

On the other hand,

$$\begin{aligned} \Lambda^\alpha{}_\beta \gamma^\beta &= \Lambda^\alpha{}_0 \gamma^0 + \Lambda^\alpha{}_1 \gamma^1 + \Lambda^\alpha{}_2 \gamma^2 + \Lambda^\alpha{}_3 \gamma^3 \\ &= \Lambda^\alpha{}_0 \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} + \Lambda^\alpha{}_1 \begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix} \\ &\quad + \Lambda^\alpha{}_2 \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix} + \Lambda^\alpha{}_3 \begin{pmatrix} 0 & -\sigma_3 \\ \sigma_3 & 0 \end{pmatrix} \end{aligned}$$

so

$$\Lambda^\alpha{}_\beta \gamma^\beta = \begin{pmatrix} 0 & \Lambda^\alpha{}_0 \sigma_0 - \sum_{i=1}^3 \Lambda^\alpha{}_i \sigma_i \\ \sum_{\beta=0}^3 \Lambda^\alpha{}_\beta \sigma_\beta & 0 \end{pmatrix}. \tag{2.4.21}$$

Now, for any $\alpha = 0,1,2,3$, the 21-block in $(\rho(g))^{-1}\gamma^\alpha \rho(g)$ is $\bar{g}^\top \sigma_\alpha g$ and the 21-block in $\Lambda^\alpha{}_\beta \gamma^\beta$ is $\sum_{\beta=0}^3 \Lambda^\alpha{}_\beta \sigma_\beta$ and these are equal by (2.4.14). The equality of the 12-blocks in (2.4.18) follows by taking inverses on both sides of (2.4.14).

More precisely, one observes that

$$(\bar{g}^\top \sigma_\alpha g)^{-1} = g^{-1}\sigma_\alpha^{-1}(\bar{g}^\top)^{-1} = g^{-1}\sigma_\alpha(\bar{g}^\top)^{-1}$$

and, from a brief calculation that we will leave for the reader,

$$\left(\sum_{\beta=0}^{3} \Lambda^\alpha{}_\beta \sigma_\beta\right)^{-1} = \begin{cases} \Lambda^0{}_0\sigma_0 - \sum_{i=1}^{3} \Lambda^0{}_i\sigma_i, & \alpha = 0 \\ -\Lambda^\alpha{}_0\sigma_0 + \sum_{i=1}^{3} \Lambda^\alpha{}_i\sigma_i, & \alpha = 1,2,3 \end{cases} .$$

With this we have established the Lorentz invariance of the Dirac equation.

The emergence of the representation $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$ as the appropriate transformation law for a Dirac wavefunction has interesting and important consequences that we will briefly explore. Let us write the Dirac wavefunction ϕ as

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{pmatrix} = \begin{pmatrix} \phi_L \\ \phi_R \end{pmatrix},$$

where $\phi_L = \binom{\phi_1}{\phi_2}$ and $\phi_R = \binom{\phi_3}{\phi_4}$. Since ϕ transforms according to $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$, ϕ_L and ϕ_R transform according to $D^{(\frac{1}{2},0)}$ and $D^{(0,\frac{1}{2})}$, respectively. Furthermore, the Dirac equation becomes a pair of coupled equations for ϕ_L and ϕ_R:

$$\gamma^\alpha \partial_\alpha \phi = -\boldsymbol{i} m\phi$$

$$\begin{pmatrix} 0 & \sigma_0\partial_0 - \sum_{i=1}^{3} \sigma_i\partial_i \\ \sigma_0\partial_0 + \sum_{i=1}^{3} \sigma_i\partial_i & 0 \end{pmatrix} \begin{pmatrix} \phi_L \\ \phi_R \end{pmatrix} = -\boldsymbol{i} m \begin{pmatrix} \phi_L \\ \phi_R \end{pmatrix}$$

$$\begin{cases} \left(\sigma_0\partial_0 - \sum_{i=1}^{3} \sigma_i\partial_i\right) \phi_R = -\boldsymbol{i} m\phi_L \\ \left(\sigma_0\partial_0 + \sum_{i=1}^{3} \sigma_i\partial_i\right) \phi_L = -\boldsymbol{i} m\phi_R \end{cases} . \tag{2.4.22}$$

In particular, in the massless ($m = 0$) case one obtains two uncoupled equations

$$\left(\sigma_0\partial_0 - \sum_{i=1}^{3}\sigma_i\partial_i\right)\phi_R = 0 \tag{2.4.23}$$

$$\left(\sigma_0\partial_0 + \sum_{i=1}^{3}\sigma_i\partial_i\right)\phi_L = 0 \tag{2.4.24}$$

which are, of course, just equations (2.4.12) and (2.4.13). To understand the significance of ϕ_R and ϕ_L in general, the relationship between our current model of spin one-half particles as 4-component objects and our earlier (2-component) view of spin one-half (pages 85–86) and just what (2.4.23) and (2.4.24) have to do with neutrinos (see page 100) we must discuss yet another symmetry (invariance property) of the Dirac equation.

The Dirac equation is invariant under the proper, orthochronous Lorentz group $\mathcal{L}_+^\uparrow$ because we were able to find a ("2-valued") representation of $\mathcal{L}_+^\uparrow$ which, if taken to be the transformation law for the wavefunction, led to a transformed wavefunction that satisfied the Dirac equation in the transformed coordinate system. Now we wish to consider a coordinate transformation, called **spatial inversion**, that does not correspond to an element of $\mathcal{L}_+^\uparrow$. Its matrix is

$$\pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and its effect is simply to switch the orientation of the spatial coordinate system. Note that, since π^2 is the 4×4 identity matrix, π generates a group of coordinate transformations which we may denote $\mathbb{Z}_2$, since that's what it is isomorphic to. In order to prove the invariance of the Dirac equation under spatial inversions we will find a representation $\rho\colon \mathbb{Z}_2 \longrightarrow GL(\ ^4)$ which, if taken to be the transformation law for ϕ, leads to a transformed wavefunction that satisfies the Dirac equation in the transformed coordinate system. Note that, since $(\gamma^0)^2$ is the 4×4 identity matrix, the assignments

$$\begin{aligned} \text{id} &\longrightarrow \text{id} \\ \pi &\longrightarrow \gamma^0 \end{aligned}$$

define a representation ρ of $\mathbb{Z}_2$ on $\ ^4$. We will show that this does the job.

Begin with the Dirac equation (2.4.15) in standard coordinates $x = (x^0, x^1, x^2, x^3)$ on $\ ^{1,3}$. Introduce new coordinates $y = (y^0, y^1, y^2, y^3)$ on $\ ^{1,3}$ by $y^\alpha = \Lambda^\alpha{}_\beta x^\beta,\ \alpha = 0,1,2,3$, where $\Lambda = (\Lambda^\alpha{}_\beta)$ is in $\mathbb{Z}_2$. Define $\hat{\partial}_\alpha = \partial/\partial y^\alpha,\ \alpha = 0,1,2,3$. Then $\partial_\beta = \Lambda^\alpha{}_\beta \hat{\partial}_\alpha$. Now define $\hat{\phi}(y)$ by

$$\hat{\phi}\,(y) = \rho(\Lambda)(\phi\,(\Lambda^{-1}y)).$$

Substituting into (2.4.15) as on pages 106–107 gives

$$\Big((\rho(\Lambda)\gamma^\beta(\rho(\Lambda))^{-1}\,\Lambda^\alpha{}_\beta)\hat{\partial}_\alpha + \boldsymbol{i}m\Big)\hat{\phi}(y) = 0$$

which will be the Dirac equation if

$$(\rho(\Lambda))\gamma^\beta(\rho(\Lambda))^{-1}\Lambda^\alpha{}_\beta = \gamma^\alpha, \quad \alpha = 0,1,2,3,$$

i.e., if

$$(\rho(\Lambda))^{-1}\gamma^\alpha\rho(\Lambda) = \Lambda^\alpha{}_\beta\gamma^\beta, \quad \alpha = 0,1,2,3 \tag{2.4.25}$$

(see page 107). Now, (2.4.25) is obviously satisfied if $\Lambda = \text{id}$ so we need only verify that it is also satisfied if $\Lambda = \pi$. In this case, (2.4.25) becomes

$$\gamma^0\gamma^\alpha\gamma^0 = \pi^\alpha{}_\beta\gamma^\beta, \qquad \alpha = 0,1,2,3,$$

i. e.,

$$\gamma^0\gamma^0\gamma^0 = \gamma^0$$

and

$$\gamma^0\gamma^i\gamma^0 = -\gamma^i \quad i = 1,2,3.$$

Since these are all easy to verify directly we have established the invariance of the Dirac equation under spatial inversion.

Notice, in particular, that if we write our Dirac wavefunction $\phi = \left(\begin{smallmatrix}\phi_L\\ \phi_R\end{smallmatrix}\right)$ as on page 109, then, under a spatial inversion, it transforms as follows:

$$\phi = \begin{pmatrix}\phi_L\\ \phi_R\end{pmatrix} \longrightarrow \gamma^0\phi = \begin{pmatrix}0 & \sigma_0\\ \sigma_0 & 0\end{pmatrix}\begin{pmatrix}\phi_L\\ \phi_R\end{pmatrix} = \begin{pmatrix}\phi_R\\ \phi_L\end{pmatrix}.$$

Switching the orientation of the spatial coordinate axes interchanges ϕ_L and ϕ_R and one therefore thinks of these two components as having opposite "handedness" or "chirality." In particular, in the $m = 0$ case one can regard (2.4.23) (respectively, (2.4.24)) as equations for a particle that is massless, spin one-half and "left-handed" (respectively, "right-handed"). Since the discovery that, in β-decay processes (in which a neutrino is emitted), parity is not conserved, these equations have been regarded as potential models for the neutrino. Lee and Yang ([**LY**]) suggested (2.4.23), but Feynman and Gell-Mann ([**FG-M**]) showed that the experimental evidence suggests (2.4.24) ("neutrinos spin to the left"). In the general (massive) case one thinks of a spin one-half particle as having two chiral components ϕ_L and ϕ_R, each of which has two additional components representing the possible spin states.

One final invariance property of the Dirac equation is worthy of note. If θ is some real *constant* so that $e^{\boldsymbol{i}\theta}$ is in $U(1)$, then

$$(\gamma^\alpha\partial_\alpha + \boldsymbol{i}m)\phi = 0$$

obviously implies

$$(\gamma^\alpha \partial_\alpha + \boldsymbol{i}m)(e^{\boldsymbol{i}\theta}\phi) = 0.$$

The Dirac equation is therefore invariant under the global $U(1)$-action $\phi \longrightarrow e^{\boldsymbol{i}\theta}\phi$. Just as was the case for the Schroedinger and Klein-Gordon equations, elevating this global symmetry to a local gauge symmetry in which θ is a function of x^0, x^1, x^2 and x^3 requires the presence of an electromagnetic gauge potential. In the physics literature such potentials are included by "minimal coupling," i.e., by replacing the ordinary derivatives ∂_α in the Dirac equation by "covariant derivatives" $\partial_\alpha + \boldsymbol{i}n\boldsymbol{A}_\alpha$ (see pages 76–80).

The problem of molding all of the information we have assembled thus far into a gauge theory model of the type described in Section 2.1 is complicated by a number of issues. Recall that in the spin zero case (Section 2.2) we began with an electromagnetic field (i.e., a connection on a $U(1)$-bundle over spacetime) and the various representations of $U(1)$ on   and, from them, constructed complex scalar fields, an action and the corresponding Euler-Lagrange equations. The resulting Klein-Gordon equations happened to be Lorentz invariant. Our point of departure in this section has been to insist at the outset on the relativistic invariance of the free particle equations. This led us to a wavefunction taking its values in $\ ^4$ whose *external* symmetry (Lorentz invariance) was expressed in the form of a transformation law corresponding to a specific representation of the double cover $SL(2,\ )$ of $\mathcal{L}_+^\uparrow$. This suggests that, in a corresponding gauge theory model, the wavefunction is a matter field on some $SL(2,\ )$-bundle over spacetime. However, electrons are coupled to electromagnetic fields and these are connections on $U(1)$-bundles, not $SL(2,\ )$-bundles, over spacetime. Furthermore, gauge invariance refers specifically to the *internal* symmetry of a particle reflected in the behavior of its wavefunction under changes in the local gauge potentials for the electromagnetic field so this notion also "lives" in a $U(1)$-bundle. To build a proper gauge theory model for Dirac electrons coupled to electromagnetic fields will require the "splicing together" of the external $SL(2,\ )$-bundle and the internal $U(1)$-bundle into a single $SL(2,\ ) \times U(1)$-bundle on which both the electron and the electromagnetic field may be thought to live. This turns out to be a relatively simple thing to do and we will outline the construction shortly.

A more delicate, and much more interesting, obstacle is one that we could evade altogether by simply continuing to restrict our attention to the spacetime $\ ^{1,3}$ and its open submanifolds. From the perspective of the workaday world of particle physics this would be an entirely reasonable choice since it amounts to ignoring gravitational effects and these are generally negligible in elementary particle interactions in the laboratory. From the perspective of topology (and nonperturbative quantum field theory), however, such a choice would "evade" the best part. We will conclude this section with a brief synopsis of the issues involved in describing spin one-half particles that live in more general spacetimes where gravitational effects are not neglected. We will deal with these issues in detail in the remaining chapters of the book.

A spacetime is a 4-dimensional (second countable, Hausdorff) manifold X with a "Lorentz metric" $\boldsymbol{g}$ (this is a semi-Riemannian metric with the property that each tangent space $T_x(X)$ has a basis $\{e_0,\ e_1,\ e_2,\ e_3\}$ for which $\boldsymbol{g}(x)(e_\alpha, e_\beta) = \eta_{\alpha\beta}$). Thus, each $T_x(X)$ with its inner product $\boldsymbol{g}(x)$ can be identified with $\ ^{1,3}$. The general Lorentz group $\mathcal{L}$ therefore acts on the orthonormal bases of each $T_x(X)$. We are, however, only interested in bases related by elements of $\mathcal{L}_+^\uparrow$. Although one can isolate such a collection of bases at each $T_x(X)$ individually just by selecting an isomorphism onto $\ ^{1,3}$, an unambiguous choice over the entire manifold X is possible only if X is assumed orientable and "time orientable." We will discuss this latter condition in more detail in Chapter 3; essentially, one assumes the existence of a vector field $\boldsymbol{V}$ on X that is timelike ($\boldsymbol{g}(x)(\boldsymbol{V}(x), \boldsymbol{V}(x))) > 0$ for each $x \in X$) and so makes a smooth selection over X of a timelike direction at each point that we may (arbitrarily) decree "future-directed." We will adopt both of these assumptions and thereby obtain, at each point, a family of oriented, time oriented, orthonormal bases for the tangent space related by elements of $\mathcal{L}_+^\uparrow$.

Remark: We will find that compact spacetimes necessarily violate certain rather basic notions of causality (they contain closed timelike curves). For this reason we will henceforth restrict our attention to the noncompact variety.

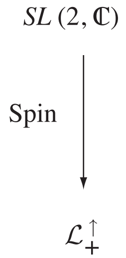
Now, Lorentz invariance means invariance under $\mathcal{L}_+^\uparrow$. When $X = \ ^{1,3}$, bundles over X are trivial so gauge fields, matter fields, etc., can all be identified with objects defined on X. Furthermore, each tangent space can be canonically identified with X itself so $\mathcal{L}_+^\uparrow$ acts on X. In the general case, none of this is true. In particular, choosing Lorentz frames and acting by $\mathcal{L}_+^\uparrow$ on such frames cannot take place globally on all of X, but only point by point. To describe all of this precisely we will build (in Chapter 3) the "oriented, time oriented, orthonormal frame bundle" of X. This is a principal $\mathcal{L}_+^\uparrow$-bundle

$$\mathcal{L}_+^\uparrow \hookrightarrow \mathcal{L}(X) \xrightarrow{\mathcal{P}_\mathcal{L}} X$$

over X whose fibers consist of the oriented, time oriented, orthonormal bases for the tangent spaces to X. The action of $\mathcal{L}_+^\uparrow$ on $\mathcal{L}(X)$ will simply carry one such basis onto another (above the same point in X) and one can make sense of "Lorentz invariance" for matter fields associated with this bundle by some representation of $\mathcal{L}_+^\uparrow$.

The frame bundle $\mathcal{L}_+^\uparrow \hookrightarrow \mathcal{L}(X) \longrightarrow X$ exists for every oriented, time oriented spacetime and so presents no real obstacle to our program. However, there is an obstacle (or, rather, "obstruction"). The fibers of $\mathcal{L}(X)$ are all isomorphic to $\mathcal{L}_+^\uparrow$, but the Dirac wavefunction did not arise from a representation of $\mathcal{L}_+^\uparrow$ on $\ ^4$. Rather, it was determined by the representation $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$

of $SL(2, \ )$ on $\ ^4$. Thus, to regard a Dirac electron as a matter field in the sense of Section 2.1 it will be necessary to "globalize" over all of X the double cover



The frame bundle provides a copy of $\mathcal{L}_+^\uparrow$ above every $x \in X$ so what we need is an $SL(2, \ )$-bundle

$$SL(2, \ ) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X \tag{2.4.26}$$

over X and a map of $S(X)$ onto $\mathcal{L}(X)$ that is, in effect, the spinor map of $\mathcal{P}_S^{-1}(x)$ onto $\mathcal{P}_{\mathcal{L}}^{-1}(x)$ for each $x \in X$. More precisely, a **spinor structure** for X consists of a principal $SL(2, \ )$-bundle (2.4.26) over X and a map

$$\lambda : S(X) \longrightarrow \mathcal{L}(X)$$

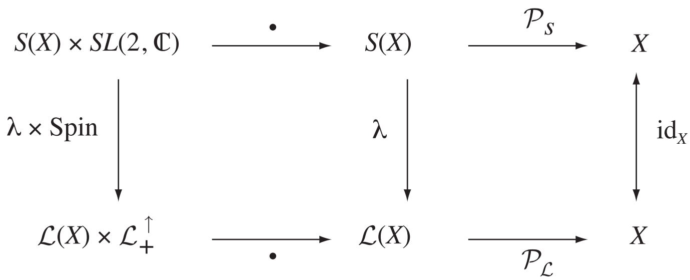
such that

$$\mathcal{P}_{\mathcal{L}}(\lambda(p)) = \mathcal{P}_S(p)$$

and

$$\lambda(p \cdot g) = \lambda(p) \ \cdot \ \text{Spin}(g)$$

for all $p \in S(X)$ and all $g \in SL(2, \ )$. The following diagram therefore commutes.



It is at this point that we encounter our obstruction. Not every oriented, time oriented spacetime X admits a spinor structure and, for those which do not, it is simply not possible to define the Dirac wavefunction for a massive, spin one-half particle. Confidence in the Dirac equation is such that this is generally regarded as adequate justification for dismissing as physically unacceptable any spacetime without a spinor structure. From our point of view the interesting part of all of this is that the existence or nonexistence of a spinor structure is a purely topological question about X. We will show that there is a certain Čech cohomology class $w_2(X) \in \check{\mathrm{H}}^2(X;\mathbb{Z}_2)$, called the "second Stiefel-Whitney class" of X, the vanishing of which is a necessary and sufficient condition for the existence of a spinor structure on $X : w_2(X)$ is the obstruction to the existence of a spinor structure (see Section 6.5).

We will show that all of the spacetimes of interest to us do, in fact, admit spinor structures. With this the construction of a gauge theory model for a free Dirac electron proceeds as follows: X is a (noncompact) spacetime manifold with $w_2(X) = 0$. The vector space $\mathcal{V}$ in which the wavefunction will take its values is $\;^4$. The (external) symmetry group G is $SL(2,\;)$ and the representation ρ of $SL(2,\;)$ on $\;^4$ is

$$\rho = D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})} : SL(2,\;) \longrightarrow GL(\;^4).$$

One can define an inner product $\langle\;,\;\rangle$ on $\;^4$ relative to which this representation is orthogonal as follows: First define the "twisted" Hermitian form H: $\;^4 \times \;^4 \longrightarrow \;$ by

$$\mathrm{H}((z_1,\ldots,z_4),(w_1,\ldots,w_4)) = z_1\bar{w}_3 + z_2\bar{w}_4 + z_3\bar{w}_1 + z_4\bar{w}_2.$$

Regarding $z,\ w \in \;^4$ as column matrices, this is equivalent to

$$\mathrm{H}(z,\ w) = z^\top \gamma^0 w,$$

where

$$\gamma^0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}$$

as on page 96. A simple computation shows that

$$\mathrm{H}(\rho(g)(z), \rho(g)(w)) = \mathrm{H}(z,w)$$

for all $g \in SL(2,\;)$. The required inner product on $\;^4$ is then given by

$$\langle z,w\rangle = \frac{1}{2}(\mathrm{H}(z,w) + \mathrm{H}(w,z)).$$

The principal $SL(2, \quad)$-bundle over X is a spinor bundle

$$SL(2, \quad) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X.$$

A **free Dirac electron** is a corresponding matter field, i.e., a smooth map

$$\phi : S(X) \longrightarrow \quad ^4$$

that is equivariant, i.e., satisfies

$$\phi(p \cdot g) = g^{-1} \cdot \phi(p) = \begin{pmatrix} g^{-1} & 0 \\ 0 & \bar{g}^{\top} \end{pmatrix} \phi(p).$$

Remarks: These electrons are "free" in the sense that they are not coupled to an electromagnetic field. Such an electromagnetic field does not live on the spinor bundle, but rather (as a connection) on a $U(1)$-bundle over X. Shortly we will describe how to splice the spinor bundle and the $U(1)$-bundle together into a single $SL(2, \quad) \times U(1)$-bundle on which an interaction can be described. Notice, however, that our "free" electron is not entirely free if the underlying spacetime X represents a nontrivial gravitational field. Such influences enter these considerations in the form of a connection on the spinor bundle that is essentially the lift of the canonical (Levi-Civita) connection on the frame bundle (see Section 3.3). With this and the potential function U : $\quad ^4 \longrightarrow \quad$ given by $U(z) = \frac{1}{2}m\|z\|^2 = \frac{1}{2}m\langle z, z\rangle$ one can write down an action whose Euler-Lagrange equations constitute the general spacetime version of the Dirac equation. The details are available in Section 6.4 of [**Bl**]. Needless to say, when $X = \quad ^{1,3}$ and the matter field is pulled back by the standard global cross-section of the (necessarily) trivial spinor bundle, this reduces to $(\gamma^{\alpha}\partial_{\alpha} + \boldsymbol{i}m)\phi = 0$.

A free Dirac electron can, of course, also be regarded as a cross-section of the vector bundle associated to $SL(2, \quad) \hookrightarrow S(X) \longrightarrow X$ by the representation $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$ (see pages 49–50). More generally, if ρ is any representation of $SL(2, \quad)$ on some $\quad ^k$, then an equivariant $\quad ^k$-valued map on $S(X)$ (or, equivalently, a cross-section of the associated vector bundle $S(X) \times_{\rho} \quad ^k$) is called a k-component **spinor field** of type ρ on X. 4-component spinor fields of type $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$ are generally called **Dirac spinor** fields. 2-component spinor fields of type $D^{(\frac{1}{2},0)}$, or $D^{(0,\frac{1}{2})}$ are called **Weyl spinor** fields.

Finally, we outline the "splicing" procedure for building from the spinor bundle (where electrons live) and a $U(1)$-bundle (where electromagnetic fields live) a single $SL(2, \quad) \times U(1)$-bundle (where both live and therefore can interact). Consider, in general, two principal bundles over the base X:

$$G_1 \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} X$$

$$G_2 \hookrightarrow P_2 \xrightarrow{\mathcal{P}_2} X$$

Let

$$P_1 \circ P_2 = \{(p_1, p_2) \in P_1 \times P_2 : \mathcal{P}_1(p_1) = \mathcal{P}_2(p_2)\}.$$

Then $P_1 \circ P_2$ is a submanifold of $P_1 \times P_2$ and will be the total space of the spliced bundle. Define

$$\mathcal{P}_{12} : P_1 \circ P_2 \longrightarrow X$$

by

$$\mathcal{P}_{12}(p_1, p_2) = \mathcal{P}_1(p_1) = \mathcal{P}_2(p_2).$$

Then $\mathcal{P}_{12}$ is a smooth map of $P_1 \circ P_2$ onto X. Define a smooth right action of $G_1 \times G_2$ on $P_1 \circ P_2$ by

$$(p_1, p_2) \cdot (g_1, g_2) = (p_1 \cdot g_1, p_2 \cdot g_2).$$

Then

$$G_1 \times G_2 \hookrightarrow P_1 \circ P_2 \xrightarrow{\mathcal{P}_{12}} X$$

is a smooth principal $G_1 \times G_2$-bundle over X.

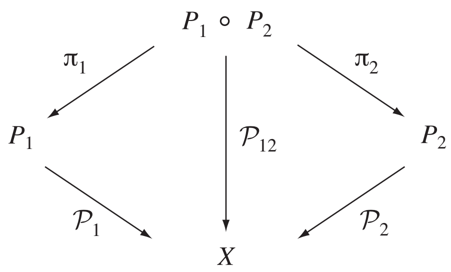
Next we define maps $\pi_1 : P_1 \circ P_2 \longrightarrow P_1$ and $\pi_2 : P_1 \circ P_2 \longrightarrow P_2$ by $\pi_i(p_1, p_2) = p_i,\ i = 1, 2$. Letting e_1 and e_2 denote the identities in G_1 and G_2 we identify $\{e_1\} \times G_2$ with G_2 and $G_1 \times \{e_2\}$ with G_1. We then have principal bundles

$$G_2 \hookrightarrow P_1 \circ P_2 \xrightarrow{\pi_1} P_1$$

and

$$G_1 \hookrightarrow P_1 \circ P_2 \xrightarrow{\pi_2} P_2$$

for which the following diagram commutes:



Now, suppose we have connections $\boldsymbol{\omega}_1$ on $\mathcal{P}_1 : P_1 \longrightarrow X$ and $\boldsymbol{\omega}_2$ on $\mathcal{P}_2 : P_2 \longrightarrow X$. Identify $\mathcal{G}_1$ with $\mathcal{G}_1 \times \{0\} \subseteq \mathcal{G}_1 \oplus \mathcal{G}_2$ and $\mathcal{G}_2$ with $\{0\} \times \mathcal{G}_2 \subseteq \mathcal{G}_1 \oplus \mathcal{G}_2$. Then $\pi_1^* \boldsymbol{\omega}_1$ is a connection on $\pi_2 : P_1 \circ P_2 \longrightarrow P_2$, $\pi_2^* \boldsymbol{\omega}_2$ is a connection on $\pi_1 : P_1 \circ P_2 \longrightarrow P_1$ and $\pi_1^* \boldsymbol{\omega}_1 \oplus \pi_2^* \boldsymbol{\omega}_2$ is a connection on $\mathcal{P}_{12} : P_1 \circ P_2 \longrightarrow X$.

Finally, let $\mathcal{V}$ be a vector space and let $\rho_1 : G_1 \longrightarrow GL(\mathcal{V})$ and $\rho_2 : G_2 \longrightarrow GL(\mathcal{V})$ be two representations that satisfy

$$\rho_1(g_1) \circ \rho_2(g_2) = \rho_2(g_2) \circ \rho_1(g_1) \qquad (2.4.27)$$

for all $g_1 \in G_1$ and $g_2 \in G_2$. Then we can define

$$\rho_1 \times \rho_2 : G_1 \times G_2 \longrightarrow GL(\mathcal{V})$$

by

$$(\rho_1 \times \rho_2)(g_1, g_2) = \rho_1(g_1) \circ \rho_2(g_2) = \rho_2(g_2) \circ \rho_1(g_1)$$

and obtain a representation of $G_1 \times G_2$ on $\mathcal{V}$ with associated left action on $\mathcal{V}$ given by

$$(g_1, g_2) \cdot \xi = (\rho_1(g_1) \circ \rho_2(g_2))(\xi).$$

Remark: ρ_1 and ρ_2 must commute, i.e., satisfy (2.4.27), to ensure that $\rho_1 \times \rho_2$ is a representation:

$$\begin{aligned}(\rho_1 \times \rho_2)\ ((g_1, g_2)\ (g_1', g_2')) = (\rho_1 \times \rho_2)\ ((g_1 g_1', g_2 g_2')) = \\ \rho_1(g_1 g_1') \circ \rho_2(g_2 g_2') &= \rho_1(g_1) \circ \rho_1(g_1') \circ \rho_2(g_2) \circ \rho_2(g_2') \\ &= \rho_1(g_1) \circ \rho_2(g_2) \circ \rho_1(g_1') \circ \rho_2(g_2') \\ &= (\rho_1 \times \rho_2)(g_1, g_2) \circ (\rho_1 \times \rho_2)\ (g_1',\ g_2').\end{aligned}$$

Now, a matter field on $G_1 \times G_2 \hookrightarrow P_1 \circ P_2 \xrightarrow{\mathcal{P}_{12}} X$ associated with $\rho_1 \times \rho_2$: $G_1 \times G_2 \longrightarrow GL(\mathcal{V})$ is a map $\phi : P_1 \circ P_2 \longrightarrow \mathcal{V}$ satisfying

$$\phi((p_1, p_2) \cdot (g_1, g_2)) = (g_1^{-1},\ g_2^{-1}) \cdot \phi\,(p_1, p_2),$$

i.e.,

$$\phi\,((p_1 \cdot g_1, p_2 \cdot g_2)) = \left(\rho_1\left(g_1^{-1}\right)\right)\left(\left(\rho_2(g_2^{-1}\right)\right)\ (\phi\,(p_1, p_2))).$$

Now we apply this construction to the following special case. Begin with an oriented, time oriented spacetime X and a spinor bundle

$$SL(2,\quad) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X$$

on X (this will be $G_1 \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} X$). Let $\boldsymbol{\omega}_1$ be the spinor connection on $S(X)$ referred to in the Remark on page 117. Take $\mathcal{V} = \ {}^4$ and let ρ_1 be the representation $D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$ of $SL(2,\ \ )$ on ^{4}. Thus,

$$(\rho_1(g_1)) \begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix} = \begin{pmatrix} g_1 & 0 \\ 0 & \left(\bar{g}_1^{\top}\right)^{-1} \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix}$$

for each g_1 in $SL(2,\ \ )$. Next let $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}_2} X$ be some principal $U(1)$-bundle over X and let $\boldsymbol{\omega}_2$ be a connection on it (representing some electromagnetic field to which the Dirac electron will respond). Take $\rho_2 \colon U(1) \longrightarrow GL(\;\;^4)$ to be the representation given by

$$(\rho_2\,(g_2))\begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix} = g_2 \begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix} = \begin{pmatrix} g_2 z_1 \\ \vdots \\ g_2 z_4 \end{pmatrix}$$

for each $g_2 \in U(1)$. Note that $\rho_1(g_1) \circ \rho_2(g_2) = \rho_2(g_2) \circ \rho_1(g_1)$ as required in (2.4.27). Thus, we have a representation

$$\rho_1 \times \rho_2 : SL(2, \;\;) \times U(1) \longrightarrow GL(\;\;^4)$$

given by

$$\begin{aligned}(\rho_1 \times \rho_2)(g_1, g_2)\begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix} &= \rho_1(g_1) \circ \rho_2\,(g_2)\begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix} \\ &= \begin{pmatrix} g_1 & 0 \\ 0 & \left(\bar{g}_1^{\top}\right)^{-1} \end{pmatrix}\begin{pmatrix} g_2 z_1 \\ \vdots \\ g_2 z_4 \end{pmatrix} \\ &= g_2 \begin{pmatrix} g_1 & 0 \\ 0 & \left(\bar{g}_1^{\top}\right)^{-1} \end{pmatrix}\begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix}.\end{aligned}$$

Now we splice the two bundles together to obtain

$$SL\,(2, \;\;) \times U(1) \hookrightarrow S(X) \circ P \longrightarrow X.$$

A **Dirac electron** (coupled to the $U(1)$-potential $\boldsymbol{\omega}_2$) is then a smooth map $\phi : S(X) \circ P \longrightarrow \;\;^4$ satisfying

$$\phi\,((p_1, p_2) \cdot (g_1, g_2)) = g_2^{-1}\begin{pmatrix} g_1^{-1} & 0 \\ 0 & \bar{g}_1^{\top} \end{pmatrix}\phi\,(p_1, p_2),$$

i.e.,

$$\phi\,(\,p_1 \cdot g_1,\, p_2 \cdot g_2\,) = \begin{pmatrix} (g_1\,g_2)^{-1} & 0 \\ 0 & \left(\overline{g_1\,g_2}\right)^{\top} \end{pmatrix}\phi\,(\,p_1,\, p_2),$$

where $g_1\,g_2$ is the entrywise product of $g_1 = e^{\boldsymbol{i}\theta_1}$ with $g_2 \in SL(2, \;\;)$. With the equipment now available one can write down an action functional whose Euler-Lagrange equations describe the interaction of a massive, spin one-half particle with an electromagnetic field (the details are available in Section 7.2 of [**Bl**]).

As one final illustration of this technique we will sketch an analogous construction for the interaction of a nucleon with a classical Yang-Mills field (for the details, see Section 7.3 of [**Bl**]). Here we face the same problem as in the case of a Dirac electron coupled to an electromagnetic field. A Yang-Mills field is given by a connection on a principal $SU(2)$-bundle over spacetime (Section 6.3 of [**N4**]), whereas a nucleon (proton/neutron) is a massive, spin one-half particle and therefore lives in a spinor bundle. There is an additional complication, however. A nucleon is a proton/neutron doublet, i.e., its wave-function has a proton component and a neutron component and so must take its values in $\mathcal{V} = \ {}^4 \oplus \ {}^4 = \ {}^8$.

We begin then with an oriented, time oriented spacetime X and a spinor bundle

$$SL(2,\ \ ) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X.$$

$\boldsymbol{\omega}_1$ is again the spinor connection referred to in the Remark on page 117. Now take $\mathcal{V} = \ {}^4 \oplus \ {}^4$, which we identify with the set of $\left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right)$ with $v_1, v_2 \in \ {}^4$. Letting $\rho = D^{(\frac{1}{2},0)} \oplus D^{(0,\frac{1}{2})}$ we define $\rho_1 : SL(2,\ \ ) \longrightarrow \ {}^4 \oplus \ {}^4$ by

$$(\rho_1(g_1))(v) = (\rho_1(g_1)) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} (\rho(g_1))(v_1) \\ (\rho(g_1))(v_2) \end{pmatrix}.$$

Now let $SU(2) \hookrightarrow P \xrightarrow{\mathcal{P}_2} X$ be some principal $SU(2)$-bundle over X and $\boldsymbol{\omega}_2$ some connection on it (representing the Yang-Mills potential to which the nucleon is coupled). Define $\rho_2 : SU(2) \longrightarrow GL(\ {}^4 \oplus \ {}^4)$ as follows: For each

$$g_2 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

in $SU(2)$,

$$(\rho_2(g_2))(v) = (\rho_2(g_2)) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$= \begin{pmatrix} \alpha v_1 + \beta v_2 \\ \gamma v_1 + \delta v_2 \end{pmatrix}.$$

A simple calculation shows that $\rho_1(g_1) \circ \rho_2(g_2) = \rho_2(g_2) \circ \rho_1(g_1)$ so we have a representation

$$\rho_1 \times \rho_2 : SL(2,\ \ ) \times SU(2) \longrightarrow GL(\ {}^4 \oplus \ {}^4).$$

Letting

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_4 \\ w_1 \\ \vdots \\ w_4 \end{pmatrix}$$

we have

$$\begin{aligned}
(\rho_1 \times \rho_2(g_1, g_2))(v) &= \begin{pmatrix} (\rho(g_1))(\alpha v_1 + \beta v_2) \\ (\rho(g_1))(\gamma v_1 + \delta v_2) \end{pmatrix} \\
&= \begin{pmatrix} \begin{pmatrix} g_1 & 0 \\ 0 & \left(\bar{g}_1^\top\right)^{-1} \end{pmatrix} & \begin{pmatrix} \alpha z_1 + \beta w_1 \\ \vdots \\ \alpha z_4 + \beta w_4 \end{pmatrix} \\ \\ \begin{pmatrix} g_1 & 0 \\ 0 & \left(\bar{g}_1^\top\right)^{-1} \end{pmatrix} & \begin{pmatrix} \gamma z_1 + \delta w_1 \\ \vdots \\ \gamma z_4 + \delta w_4 \end{pmatrix} \end{pmatrix}.
\end{aligned}$$

A **nucleon field** (coupled to an $SU(2)$-Yang-Mills potential $\boldsymbol{\omega}_2$) is then a smooth map $\phi : S(X) \circ P \longrightarrow \ ^4 \oplus \ ^4$ such that

$$\phi\left((p_1, p_2) \cdot (g_1, g_2)\right) = (g_1^{-1}, g_2^{-1}) \cdot \phi(p_1, p_2).$$

Writing $\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ we have

$$\begin{pmatrix} \phi_1(p_1 \cdot g_1, p_2 \cdot g_2) \\ \phi_2(p_1 \cdot g_1, p_2 \cdot g_2) \end{pmatrix} = \begin{pmatrix} (\rho(g_1^{-1}))(\alpha\phi_1(p_1 \cdot g_1, p_2 \cdot g_2) \\ \qquad\qquad + \beta\phi_2(p_1 \cdot g_1, p_2 \cdot g_2)) \\ (\rho(g_1^{-1}))(\gamma\phi_1(p_1 \cdot g_1, p_2 \cdot g_2) \\ \qquad\qquad + \delta\phi_2(p_1 \cdot g_1, p_2 \cdot g_2)) \end{pmatrix}.$$

ϕ_1 is called the **proton component** of ϕ, while ϕ_2 is its **neutron component**. Note the "mixing" of the proton and neutron components due to the $SU(2)$-action.

2.5 $SU(2)$-Yang-Mills-Higgs Theory on ^{n}

The motivation behind our final example of a classical gauge theory is a process known as "dimensional reduction." The example itself is of profound significance to physics, arises naturally from the pure Yang-Mills theory on ^{4} that was the subject of **[N4]**, provides an extraordinary insight into the true nature of the Dirac magnetic monopole and gives rise to even deeper connections between physics and topology than those we have encountered thus far. We will begin by simply enumerating, without motivation, the eight items required (in Section 2.1) for the construction of a classical gauge theory. Then we will review the structure of pure $SU(2)$- Yang-Mills theory on $\ ^4$ and show how our example arises from it. Finally, we will describe a number of the remarkable properties of the model, both physical and mathematical.

The base manifold is $X = \ ^n$ with its usual orientation and Riemannian metric. For the vector space $\mathcal{V}$ we take the Lie algebra $su(2)$ of 2×2 complex matrices that are skew-Hermitian and tracefree. The (positive definite) inner product on $su(2)$ is given by $\langle A, B \rangle = -2$ trace (AB) ($-$trace(AB) is the Killing form of $su(2)$ and the 2 is a matter of convenience). The Lie group G is taken to be $SU(2)$ and $\rho : SU(2) \longrightarrow GL(su(2))$ is the adjoint representation

$$ad_g(A) = gAg^{-1}$$

for all $g \in SU(2)$ and $A \in su(2)$. Note that $\langle ad_g(A), ad_g(B) \rangle = \langle gAg^{-1},\, gBg^{-1} \rangle = -2$ trace$((gAg^{-1})(gBg^{-1})) = -2$ trace$(g(AB)g^{-1}) = -2$ trace$(AB) = \langle A, B \rangle$, as required. Since every bundle over $\ ^n$ is trivial, we will trivialize at the outset and take

$$SU(2) \hookrightarrow \ ^n \times SU(2) \stackrel{\mathcal{P}}{\longrightarrow} \ ^n$$

as our bundle, where the right action of $SU(2)$ on $\ ^n \times SU(2)$ is given by

$$p \cdot g = (x, h) \cdot g = (x, hg)$$

for all $p = (x, h) \in \ ^n \times SU(2)$ and all $g \in SU(2)$. There is a natural global cross-section $s : \ ^n \longrightarrow \ ^n \times SU(2)$ given by

$$s(x) = (x, e)$$

where, for convenience, we write e for the identity element in $SU(2)$. Any other global cross-section then has the form

$$\begin{aligned} s^g : \ ^n &\longrightarrow \ ^n \times SU(2) \\ s^g(x) &= s(x) \cdot g(x) \\ &= (x, e) \cdot g(x) \\ &= (x, g(x)) \end{aligned}$$

for some smooth map $g : \ ^n \longrightarrow SU(2)$ (Exercise 4.3.5 of [**N4**]). The cross-section s^g gives rise to an automorphism of the bundle (i.e., a global gauge transformation) in the usual way (see page 343 of [**N4**]):

$$\begin{aligned} s(x) \cdot h &\longrightarrow s^g(x) \cdot h \\ (x, e) \cdot h &\longrightarrow (x, g(x)) \cdot h \\ (x, h) &\longrightarrow (x, g(x)h). \end{aligned}$$

Thus, one can identify a gauge transformation with a smooth map $g :\ ^n \longrightarrow SU(2)$ which multiplies in the fibers on the left.

Because of the triviality of the bundle, any connection $\boldsymbol{\omega}$ on $SU(2) \hookrightarrow\ ^n \times SU(2) \longrightarrow\ ^n$ is uniquely determined by its gauge potential

$$\boldsymbol{\mathcal{A}} = s^* \boldsymbol{\omega},$$

which is an $su(2)$-valued 1-form on ^{n}. Furthermore, any $su(2)$-valued 1-form on ^{n} is the pullback by s of a unique connection on $SU(2) \hookrightarrow\ ^n \times SU(2) \longrightarrow\ ^n$ (page 333, [**N4**]). Thus, we may restrict our attention entirely to globally defined gauge potentials $\boldsymbol{\mathcal{A}}$ on ^{n}. Relative to standard coordinates on ^{n} we write

$$\boldsymbol{\mathcal{A}} = s^* \boldsymbol{\omega} = \mathcal{A}_\alpha dx^\alpha,$$

where each $\mathcal{A}_\alpha,\ \alpha = 1, \ldots, n$, takes values in $su(2)$ (see (2.5.2) below). A gauge transformation $g :\ ^n \longrightarrow SU(2)$ gives a new gauge potential

$$\boldsymbol{\mathcal{A}}^g = (s^g)^* \boldsymbol{\omega}$$

related to $\boldsymbol{\mathcal{A}}$ by

$$\boldsymbol{\mathcal{A}}^g = g^{-1} \boldsymbol{\mathcal{A}} g + g^{-1} dg,$$

where dg is the entrywise exterior derivative of $g :\ ^n \longrightarrow SU(2)$ and the products are matrix products (we will do an explicit calculation of this sort for the "t' Hooft-Polyakov monopole" somewhat later). The curvature $\boldsymbol{\Omega}$ of $\boldsymbol{\omega}$ is likewise uniquely determined by the field strength

$$\boldsymbol{\mathcal{F}} = s^* \boldsymbol{\Omega} = d\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}} = \frac{1}{2} \mathcal{F}_{\alpha\beta} dx^\alpha \wedge dx^\beta,$$

where

$$\mathcal{F}_{\alpha\beta} = \partial_\alpha \mathcal{A}_\beta - \partial_\beta \mathcal{A}_\alpha + [\mathcal{A}_\alpha, \mathcal{A}_\beta], \quad \alpha, \beta = 1, \ldots, n.$$

A gauge transformation $g :\ ^n \longrightarrow SU(2)$ gives a new field strength

$$\boldsymbol{\mathcal{F}}^g = (s^g)^* \boldsymbol{\Omega}$$

related to $\boldsymbol{\mathcal{F}}$ by

$$\boldsymbol{\mathcal{F}}^g = g^{-1} \boldsymbol{\mathcal{F}} g.$$

In this context a matter field is a smooth $su(2)$-valued map Φ on $^n \times SU(2)$ that satisfies

$$\Phi(p \cdot g) = g^{-1} \cdot \Phi(p)$$
$$\Phi((x,h) \cdot g) = ad_{g^{-1}} (\Phi(x,h))$$
$$\Phi\ (x, hg\ ) = g^{-1} \Phi(x,h) g$$

for all $(x,h) \in\ ^n \times SU(2)$ and all $g \in SU(2)$. When, as in this case, $\mathcal{V}$ is the Lie algebra $\mathcal{G}$ of the structure group G and ρ is the adjoint representation of G on $\mathcal{G}$, a matter field is referred to as a **Higgs field**. The triviality of the bundle in our present circumstances allows us to identify the Higgs field with its pullback by the global cross-section s:

$$\phi = s^*\mathbf{\Phi} = \mathbf{\Phi} \circ s$$
$$\phi(x) = \mathbf{\Phi}(x, e).$$

Under a gauge transformation $g : \quad ^n \longrightarrow SU(2)$,

$$\phi^g = (s^g)^*\mathbf{\Phi} = g^{-1}\phi g$$

because

$$\begin{aligned} ((s^g)^*\mathbf{\Phi})(x) &= \mathbf{\Phi}(s^g(x)) \\ &= \mathbf{\Phi}(x, g) \\ &= \mathbf{\Phi}(x, eg) \\ &= g^{-1}\mathbf{\Phi}(x, e)g \\ &= g^{-1}\phi(x)g. \end{aligned}$$

The next item on the agenda (#7 of Section 2.1) is the potential function $U : su(2) \longrightarrow \quad$. This plays a rather peculiar role in the story we wish to tell. Initially we adopt what is called the **Georgi-Glashow potential** $U : su(2) \longrightarrow \quad$ given by

$$U(A) = \frac{\lambda}{8}(||A||^2 - 1)^2,$$

where $\lambda \geq 0$ is a constant and $\| A \|^2 = \langle A, A \rangle = -2\ \mathrm{trace}(A^2)$, noting that $U(g \cdot A) = U(gAg^{-1}) = U(A)$ as required. Shortly, however, we will take λ to be zero and retain only a vestige of the potential in the form of an asymptotic boundary condition that it imposes on the Higgs field ϕ (see pages 132–133).

In order to describe the appropriate action (#8 of Section 2.1) for our example we must anticipate a few results on differential forms that will be proved later (Chapter 4). We will content ourselves with just a brief description of those particular items required for the example. We have already introduced real- and vector-valued 0-forms (pages 10 and 12), 1-forms (pages 9 and 12), and 2-forms (pages 11 and 12). k-forms, for integers $k \geq 3$, are defined analogously and all of the familiar algebraic and analytic operations on 0-, 1-, and 2-forms extend to this more general context. For example, a real-valued 3-form on a manifold X is a map $\boldsymbol{\alpha}$ that assigns to each $p \in X$ a real-valued trilinear function $\boldsymbol{\alpha}_p$ on $T_p(X) \times T_p(X) \times T_p(X)$ that is skew-symmetric (changes sign when-ever two of its arguments are interchanged) and smooth in the sense that, for any $\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3 \in \mathcal{X}(X)$, the function $\boldsymbol{\alpha}(\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3)$ on X defined by $(\boldsymbol{\alpha}(\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3))(p) = \boldsymbol{\alpha}_p(\boldsymbol{V}_1(p), \boldsymbol{V}_2(p), \boldsymbol{V}_3(p))$ is in $C^\infty(X)$. These arise, for example, as exterior derivatives of 2-forms and wedge products of 1-forms and 2-forms, or of three 1-forms (all of which will be defined carefully in Chapter 4).

In general, the set of k-forms on an n-dimensional manifold X is denoted $\Lambda^k(X)$ and admits a natural $C^\infty(X)$-module structure (and so in particular, is a real vector space). If (U, φ) is a chart for X with coordinate functions

$x^1, \ldots, x^n$, then any $\boldsymbol{\alpha} \in \Lambda^k(X)$ has a local coordinate expression

$$\boldsymbol{\alpha} = \frac{1}{k!}\, \alpha_{i_1 \cdots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

(summation over $i_1, \ldots, i_k = 1, \ldots, n$), where each $\alpha_{i_1 \cdots i_k}$ is in $C^\infty(U)$. The exterior derivative $d\boldsymbol{\alpha}$ of $\boldsymbol{\alpha} \in \Lambda^k(X)$ is an element of $\Lambda^{k+1}(X)$ which, locally, is given by

$$d\boldsymbol{\alpha} = \frac{1}{k!}(d\alpha_{i_1 \cdots i_k}) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

There are no nonzero k-forms on X if $k > n$ and, for $0 \leq k \leq n$, we will show that the dimension of $\Lambda^k(X)$ as a $C^\infty(X)$-module is $\binom{n}{k}$. Since $\binom{n}{n-k} = \binom{n}{k}$, the modules $\Lambda^k(X)$ and $\Lambda^{n-k}(X)$ are isomorphic. We will find that, when X is oriented and has a metric (Riemannian or semi-Riemannian), then there is a natural isomorphism

$$^* : \Lambda^k(X) \longrightarrow \Lambda^{n-k}(X),$$

called the Hodge star operator. Moreover, $\dim \Lambda^n(X) = \binom{n}{n} = 1$ and, when X is oriented and has a metric, there is a distinguished generator for $\Lambda^n(X)$ called the metric volume form and denoted **vol** (in standard coordinates on ^{n} this is just $dx^1 \wedge \cdots \wedge dx^n$). In particular, for any $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \Lambda^k(X), \boldsymbol{\alpha} \wedge {}^*\boldsymbol{\beta}$ is in $\Lambda^n(X)$ and so is a multiple, by some element of $C^\infty(X)$, of **vol**. We denote this element of $C^\infty(X)$ by $\langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle$:

$$\boldsymbol{\alpha} \wedge^* \boldsymbol{\beta} = \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle \mathbf{vol}.$$

This defines an inner product on $\Lambda^k(X)$ and, when $\boldsymbol{\beta} = \boldsymbol{\alpha}$, we will write $\langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \rangle = \| \boldsymbol{\alpha} \|^2$. In the Riemannian case we will find that $^{**}\boldsymbol{\beta} = (-1)^{k(n-k)}\boldsymbol{\beta}$ and that it follows from this that the Hodge star operator is actually an isometry. k-forms that vanish outside of a compact set can be integrated over k-dimensional manifolds. Indeed, we will find that they can even be integrated over k-dimensional regions with a sufficiently smooth $(k-1)$-dimensional "boundary" and it is in this context that we will prove a version of Stokes' Theorem relating the integral of a $(k-1)$-form $\boldsymbol{\alpha}$ over this boundary to the integral of $d\boldsymbol{\alpha}$ over the region it bounds.

Much of what we have just said about real-valued forms extends easily to vector-valued forms by simply doing everything (evaluation at tangent vectors, exterior derivative, Hodge star, etc.) componentwise relative to some basis for the vector space (we will show that it all turns out to be independent of the choice of basis). There are a few troublesome items (e.g., wedge products) that we will treat carefully in Chapter 4 and simply illustrate here for the particular vector space of interest in our example i.e., $su(2)$. We take as a basis for $su(2)$ the set $\{T_1, T_2, T_3\}$, where

$$T_1 = -\frac{1}{2}\boldsymbol{i}\sigma_1, \quad T_2 = -\frac{1}{2}\boldsymbol{i}\sigma_2, \quad T_3 = -\frac{1}{2}\boldsymbol{i}\sigma_3,$$

and

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\boldsymbol{i} \\ \boldsymbol{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli spin matrices. It is easy to see that $\{T_1, T_2, T_3\}$ is orthonormal with respect to the inner product $\langle A, B \rangle = -2 \text{ trace}(AB)$ on $su(2)$. Then any $su(2)$-valued k-form $\boldsymbol{\varphi}$ on $\ ^n$ (e.g, $\boldsymbol{\mathcal{A}}$ or $\boldsymbol{\mathcal{F}}$) can be regarded as a matrix of complex k-forms

$$\boldsymbol{\varphi} = \boldsymbol{\varphi}^a T_a = -\frac{1}{2}\boldsymbol{\varphi}^a(\boldsymbol{i}\sigma_a) = -\frac{1}{2}\begin{pmatrix} \boldsymbol{\varphi}^3\boldsymbol{i} & \boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\boldsymbol{i} \\ -\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\boldsymbol{i} & -\boldsymbol{\varphi}^3\boldsymbol{i} \end{pmatrix} \tag{2.5.1}$$

where $\boldsymbol{\varphi}^1$, $\boldsymbol{\varphi}^2$, and $\boldsymbol{\varphi}^3$ are in $\Lambda^k(\ ^n)$. Alternatively one can write

$$\boldsymbol{\varphi}^a = \frac{1}{k!}\varphi^a_{i_1\cdots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \quad a = 1, \ldots, n$$

and then

$$\begin{aligned} \boldsymbol{\varphi} = \boldsymbol{\varphi}^a T_a &= \left(\frac{1}{k!}\varphi^a_{i_1\cdots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) T_a \\ &= \frac{1}{k!}\left(\varphi^a_{i_1\cdots i_k} T_a\right) dx^{i_1} \wedge \cdots \wedge dx^{i_k} \end{aligned}$$

$$\boldsymbol{\varphi} = \frac{1}{k!}\,\varphi_{i_1\cdots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \tag{2.5.2}$$

where

$$\begin{aligned} \varphi_{i_1\cdots i_k} &= \varphi^a_{i_1\cdots i_k} T_a \\ &= -\frac{1}{2}\begin{pmatrix} \varphi^3_{i_1\cdots i_k}\boldsymbol{i} & \varphi^2_{i_1\cdots i_k} + \varphi^1_{i_1\cdots i_k}\boldsymbol{i} \\ -\varphi^2_{i_1\cdots i_k} + \varphi^1_{i_1\cdots i_k}\boldsymbol{i} & -\varphi^3_{i_1\cdots i_k}\,\boldsymbol{i} \end{pmatrix} \end{aligned} \tag{2.5.3}$$

is a smooth map into $su(2)$ for each $i_1 \ldots,\ i_k = 1, \ldots,\ n$.

We will, in Chapter 4, discuss various natural ways of defining a wedge product for matrix-valued forms such as these. From our point of view at the moment, the most useful such notion can be described as follows: The wedge product of two complex-valued forms is obtained by multiplying the forms as if they were complex numbers, but with real and imaginary parts multiplied by the ordinary wedge product of real-valued forms, i.e.,

$$(\boldsymbol{\varphi}^1 + \boldsymbol{\varphi}^2\boldsymbol{i}) \wedge (\boldsymbol{\psi}^1 + \boldsymbol{\psi}^2\boldsymbol{i}) = (\boldsymbol{\varphi}^1 \wedge \boldsymbol{\psi}^1 - \boldsymbol{\varphi}^2 \wedge \boldsymbol{\psi}^2) + (\boldsymbol{\varphi}^1 \wedge \boldsymbol{\psi}^2 + \boldsymbol{\varphi}^2 \wedge \boldsymbol{\psi}^1)\boldsymbol{i}$$

Now, if $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ are both $su(2)$-valued and written as in (2.5.1), then $\boldsymbol{\varphi} \wedge \boldsymbol{\psi}$ is obtained by simply forming their matrix product, with entries multipled by the complex wedge product described above. We will illustrate the procedure with an example that will also allow us to write out our action functional. We consider a k-form $\boldsymbol{\varphi}$ with values in $su(2)$ and written in the form (2.5.1) and will compute $\boldsymbol{\varphi} \wedge {}^*\boldsymbol{\varphi}$, where ${}^*\boldsymbol{\varphi}$ is the Hodge dual of $\boldsymbol{\varphi}$, computed componentwise (i.e., entrywise). Thus,

$$\boldsymbol{\varphi} \wedge {}^*\boldsymbol{\varphi} = \frac{1}{4} \begin{pmatrix} \boldsymbol{\varphi}^3\,\boldsymbol{i} & \boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\boldsymbol{i} \\ -\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\boldsymbol{i} & -\boldsymbol{\varphi}^3\boldsymbol{i} \end{pmatrix} \begin{pmatrix} {}^*\boldsymbol{\varphi}^3\,\boldsymbol{i} & {}^*\boldsymbol{\varphi}^2 + {}^*\boldsymbol{\varphi}^1\boldsymbol{i} \\ -{}^*\boldsymbol{\varphi}^2 + {}^*\boldsymbol{\varphi}^1\boldsymbol{i} & -{}^*\boldsymbol{\varphi}^3\boldsymbol{i} \end{pmatrix}$$

$$= \frac{1}{4} \begin{pmatrix} \begin{array}{l} -\boldsymbol{\varphi}^3 \wedge {}^*\boldsymbol{\varphi}^3 \\ \quad +(\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\,\boldsymbol{i}) \\ \qquad \wedge(-{}^*\boldsymbol{\varphi}^2 + {}^*\boldsymbol{\varphi}^1\,\boldsymbol{i}) \end{array} & \begin{array}{l} (\boldsymbol{\varphi}^3\,\boldsymbol{i}) \wedge ({}^*\boldsymbol{\varphi}^2 + {}^*\boldsymbol{\varphi}^1\,\boldsymbol{i}) \\ \quad -(\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\,\boldsymbol{i}) \\ \qquad \wedge({}^*\boldsymbol{\varphi}^3\,\boldsymbol{i}) \end{array} \\ \\ \begin{array}{l} (-\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\,\boldsymbol{i}) \\ \quad \wedge({}^*\boldsymbol{\varphi}^3\,\boldsymbol{i}) - (\boldsymbol{\varphi}^3\,\boldsymbol{i}) \\ \qquad \wedge(-{}^*\boldsymbol{\varphi}^2 + {}^*\boldsymbol{\varphi}^1\,\boldsymbol{i}) \end{array} & \begin{array}{l} (-\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\,\boldsymbol{i}) \\ \quad \wedge({}^*\boldsymbol{\varphi}^2 + {}^*\boldsymbol{\varphi}^1\,\boldsymbol{i}) \\ \qquad -\boldsymbol{\varphi}^2 \wedge {}^*\boldsymbol{\varphi}^3 \end{array} \end{pmatrix}.$$

Each entry is computed in the same way. For example,

$$\begin{aligned} -\boldsymbol{\varphi}^3 \wedge {}^*\boldsymbol{\varphi}^3 + (\boldsymbol{\varphi}^2 + \boldsymbol{\varphi}^1\boldsymbol{i}) \wedge (-{}^*\boldsymbol{\varphi}^2 \wedge {}^*\boldsymbol{\varphi}^1\boldsymbol{i}) \\ = -\|\boldsymbol{\varphi}^3\|^2\mathbf{vol} - \|\boldsymbol{\varphi}^2\|^2\mathbf{vol} - \|\boldsymbol{\varphi}^1\|^2\mathbf{vol} \\ + \ (\langle\boldsymbol{\varphi}^2, \boldsymbol{\varphi}^1\rangle\mathbf{vol} - \langle\boldsymbol{\varphi}^1,\ \boldsymbol{\varphi}^2\rangle\mathbf{vol})\boldsymbol{i} \\ = -(\|\boldsymbol{\varphi}^1\|^2 + \|\boldsymbol{\varphi}^2\|^2 + \|\boldsymbol{\varphi}^3\|^2)\mathbf{vol} \end{aligned}$$

and similarly for the rest. In particular, the $(2,2)$-entry is the same so

$$-2\mathrm{trace}(\boldsymbol{\varphi} \wedge {}^*\boldsymbol{\varphi}) = (\|\boldsymbol{\varphi}^1\|^2 + \|\boldsymbol{\varphi}^2\|^2 + \|\boldsymbol{\varphi}^3\|^2)\mathbf{vol}.$$

We define

$$\|\boldsymbol{\varphi}\|^2 = \|\boldsymbol{\varphi}^1\|^2 + \|\boldsymbol{\varphi}^2\|^2 + \|\boldsymbol{\varphi}^3\|^2$$

so that

$$-2\mathrm{trace}(\boldsymbol{\varphi} \wedge {}^*\boldsymbol{\varphi}) = \|\boldsymbol{\varphi}\|^2\mathbf{vol}. \tag{2.5.4}$$

Remark: More generally, if $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ are any two $su(2)$-valued k-forms on ^{n} and $\{T_1, T_2, T_3\}$ is *any* orthonormal basis for $su(2)$ and if we write $\boldsymbol{\varphi} = \boldsymbol{\varphi}^a\,T_a$ and $\boldsymbol{\psi} = \boldsymbol{\psi}^a\,T_a$, then we can define $\langle\boldsymbol{\varphi}, \boldsymbol{\psi}\rangle = \langle\boldsymbol{\varphi}^1, \boldsymbol{\psi}^1\rangle + \langle\boldsymbol{\varphi}^2, \boldsymbol{\psi}^2\rangle + \langle\boldsymbol{\varphi}^3, \boldsymbol{\psi}^3\rangle$ and show, as above, that

$$-2\mathrm{trace}(\boldsymbol{\varphi} \wedge {}^*\boldsymbol{\psi}) = \langle\boldsymbol{\varphi}, \boldsymbol{\psi}\rangle\ \mathbf{vol}.$$

With the machinery we have assembled thus far we can write out the **Yang-Mills-Higgs action functional** $A(\boldsymbol{\mathcal{A}}, \phi)$ for our example:

$$
\begin{aligned}
A(\boldsymbol{\mathcal{A}}, \phi) \quad &= \int_{{}^n} \Big(-\operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge {}^*\boldsymbol{\mathcal{F}}) - \operatorname{trace}(d^{\mathcal{A}}\,\phi \wedge d^{\mathcal{A}}\phi) \\
&\qquad\qquad + \frac{\lambda}{8}\,{}^*(\|\phi\|^2 - 1)^2 \Big) \\
&= \frac{1}{2}\int_{{}^n} \Big(\|\boldsymbol{\mathcal{F}}\|^2 + \|d^{\mathcal{A}}\phi\|^2 \\
&\qquad\qquad + \frac{\lambda}{4}\,(\|\phi\|^2 - 1)^2 \Big)\; dx^1 \wedge \cdots \wedge dx^n,
\end{aligned}
\tag{2.5.5}
$$

where $d^{\mathcal{A}}\phi = d\phi + [\boldsymbol{\mathcal{A}}, \phi]$ is the covariant exterior derivative of the Higgs field ϕ. We observe first that $A(\boldsymbol{\mathcal{A}}, \phi)$ is gauge invariant, i.e., that the effect of a gauge transformation $g : {}^n \longrightarrow SU(2)$ ($\boldsymbol{\mathcal{A}} \longrightarrow \boldsymbol{\mathcal{A}}^g$, $\boldsymbol{\mathcal{F}} \longrightarrow \boldsymbol{\mathcal{F}}^g$ and $\phi \longrightarrow \phi^g$) is to leave the integral unchanged. We have already seen that $\boldsymbol{\mathcal{F}}^g = g^{-1}\,\boldsymbol{\mathcal{F}} g$ and $\phi^g = g^{-1}\phi g$ and we show now that

$$
d^{\boldsymbol{\mathcal{A}}^g}\,\phi^g = g^{-1}\Big(d^{\boldsymbol{\mathcal{A}}}\,\phi\Big)\,g \tag{2.5.6}
$$

as well (we will need to use a few simple algebraic properties of d, e.g., the product rule for matrix products, that will be proved in Chapter 4). Indeed, $d^{\boldsymbol{\mathcal{A}}}\,\phi = d\phi + [\boldsymbol{\mathcal{A}}, \phi]$ implies

$$
g^{-1}\Big(d^{\boldsymbol{\mathcal{A}}}\,\phi\Big)g = g^{-1}\,d\phi g + g^{-1}[\boldsymbol{\mathcal{A}}, \phi]g
$$

and

$$
\begin{aligned}
d^{\boldsymbol{\mathcal{A}}^g}\,\phi^g &= d\phi^g + [\boldsymbol{\mathcal{A}}^g, \phi^g] = d\Big(g^{-1}\phi g\Big) + \Big[g^{-1}\boldsymbol{\mathcal{A}}g + g^{-1}dg,\, g^{-1}\,\phi g\Big] \\
&= d\Big(g^{-1}\,\phi g\Big) + \Big[g^{-1}\boldsymbol{\mathcal{A}}g,\, g^{-1}\,\phi g\Big] + \Big[g^{-1}dg,\, g^{-1}\,\phi g\Big] \\
&= g^{-1}[\boldsymbol{\mathcal{A}}, \phi]g + d\Big(g^{-1}\,\phi g\Big) + \Big[g^{-1}dg,\, g^{-1}\,\phi g\Big] \\
&= g^{-1}[\boldsymbol{\mathcal{A}}, \phi]g + g^{-1}d(\phi g) + dg^{-1}\,\phi g \\
&\qquad\quad + g^{-1}\,dg\,g^{-1}\phi g - g^{-1}\phi\,g\,g^{-1}dg \\
&= g^{-1}[\boldsymbol{\mathcal{A}}, \phi]g + g^{-1}\,\phi dg + g^{-1}\,d\phi g \\
&\qquad\quad + dg^{-1}\,\phi g + g^{-1}\,dg\,g^{-1}\,\phi g - g^{-1}\,\phi dg \\
&= g^{-1}[\boldsymbol{\mathcal{A}}, \phi]g + g^{-1}\,d\phi g + dg^{-1}\,\phi g + g^{-1}\,dg\,g^{-1}\,\phi g \\
&= g^{-1}[\boldsymbol{\mathcal{A}}, \phi]g + g^{-1}d\phi g = g^{-1}\Big(d^{\boldsymbol{\mathcal{A}}}\,\phi\Big)g
\end{aligned}
$$

because

$$
\begin{aligned}
g^{-1}g \; = id &\Longrightarrow g^{-1}\, dg + dg^{-1}\, g = 0 \\
&\Longrightarrow g^{-1}\, dg = -dg^{-1}\, g \\
&\Longrightarrow g^{-1}\, dg\, g^{-1}\, \phi g = -dg^{-1}\, \phi g.
\end{aligned}
$$

Gauge invariance of the action will therefore follow if we can show that $\|g^{-1}\boldsymbol{\varphi} g\|^2 = \|\boldsymbol{\varphi}\|^2$ for any $su(2)$-valued form $\boldsymbol{\varphi}$. But if we write $\boldsymbol{\varphi} = \boldsymbol{\varphi}^a\, T_a$, then $g^{-1}\boldsymbol{\varphi} g = \boldsymbol{\varphi}^a\,(g^{-1}\, T_a\, g)$ and, since $\langle g^{-1}Ag, g^{-1}Bg\rangle = \langle A, B\rangle, \{g^{-1}T_1 g, g^{-1}T_2 g, g^{-1}T_3 g\}$ is also an orthonormal basis for $su(2)$ so this is clear (see the Remark on page 130).

We are interested in finite action, stationary configurations $(\boldsymbol{\mathcal{A}}, \phi)$, i.e., solutions to the Euler-Lagrange equations for the action (2.5.5) for which $A(\boldsymbol{\mathcal{A}}, \phi) < \infty$. As it happens, no such solutions exist when $n > 4$ (see **[JT]**). When $n = 2$ such solutions do exist and they are called **vortices**. These are studied exhaustively in **[JT]**, but we will have no more to say about them. When $n = 4$ any such solution is gauge equivalent to a pure Yang-Mills field $(\lambda = 0, \phi = 0)$ of the type discussed in **[N4]** (we will briefly review this material shortly). Our primary concern is with the case $n = 3$ where finite action, stationary configurations are, for reasons we hope to make clear, called **monopoles**. In fact, we intend to discuss only a special case in which just a vestige of the Georgi-Glashow potential survives. This special case arises in the following way: The requirement that $A(\boldsymbol{\mathcal{A}}, \phi) < \infty$ implies that, as $|x| \longrightarrow \infty$ in ^{3},

$$
\left\| \boldsymbol{\mathcal{F}} \right\| \longrightarrow 0 \tag{2.5.7}
$$

$$
\left\| d^{\mathcal{A}}\, \phi \right\| \longrightarrow 0 \tag{2.5.8}
$$

and, at least if $\lambda \neq 0$,

$$
\left\| \phi \right\| \longrightarrow 1. \tag{2.5.9}
$$

Indeed, it is shown in Chapter 4, Sections 10–15, of **[JT]** that each of these limits is achieved uniformly. Now, when $\lambda = 0$ there is no reason to suppose that finite action implies $\|\phi\| \longrightarrow 1$ as $|x| \longrightarrow \infty$. However, **[JT]** also shows that, even in this case, there is some constant $c \geq 0$ such that $\|\phi\| \longrightarrow c$ uniformly as $|x| \longrightarrow \infty$ and that if $c \neq 0$, one can rescale in ^{3} to obtain a new configuration $(\boldsymbol{\mathcal{A}}'(x), \phi'(x)) = (c^{-1}\, \boldsymbol{\mathcal{A}}(c^{-1}\, x),\; c^{-1}\, \phi(c^{-1}x))$ which is a finite action stationary point for the action A with $\lambda = 0$ and satisfies $\|\phi'\| \longrightarrow 1$ uniformly as $|x| \longrightarrow \infty$. In effect, one loses nothing, even in the $\lambda = 0$ case, by restricting attention to those configurations for which (2.5.7), (2.5.8) *and* (2.5.9) are satisfied. This, then, is precisely what we intend to.

Remark: Before abandoning the self-interaction term, however, we point out some general features of the action (2.5.5). In particular, we note that there are some obvious absolute minima. Indeed, $A(\boldsymbol{\mathcal{A}}, \phi)$ is obviously zero whenever $\boldsymbol{\mathcal{A}} = 0$ and $\phi = \phi_0$ is a constant in $su(2)$ with $\|\phi_0\| = 1$. Such an absolute minimum is called a *ground state* for the system. The corresponding quantum state of lowest energy is called a *vacuum state* and physicists perform perturbation calculations about such vacuum states. The point here is that such vacuum states are not unique (as long as $\boldsymbol{\mathcal{A}} = 0$, any $\phi_0 \in S^2 \subseteq su(2)$ will give rise to such a state). A specific choice of such a ϕ_0 is said to *break the symmetry* from $SU(2)$ to $U(1)$ and we wish to very briefly explain the terminology (see Section 10.3 of [**Bl**] for more details). A gauge transformation $g : \ ^3 \longrightarrow SU(2)$ acts on ϕ by $\phi \longrightarrow \phi^g = g^{-1}\,\phi g$. If the ground state is to be gauge invariant, then we must have $g^{-1}\,\phi_0\,g = \phi_0$ and this occurs only for g in the isotropy subgroup of ϕ_0 in $SU(2)$ under the adjoint action. Now, we claim that this isotropy subgroup is a copy $U(1)$. To see this, identify $su(2)$ with $\ ^3$ and $SU(2)\,/\pm$ id with $SO(3)$ (page 88). Then the adjoint action of $SU(2)$ on $su(2)$ corresponds to the natural action of $SO(3)$ on $\ ^3$ (this is proved, although not stated in these terms in Appendix A to [**N4**]). But this natural action of $SO(3)$ on $\ ^3$ (rotation) is transitive on $S^2 \subseteq \ ^3$ (page 27) Now, if $H = \{g \in SU(2) : g^{-1}\,\phi_0\,g = \phi_0\}$, then (since $SU(2)$ is compact), $SU(2)\,/\,H \cong S^2$ (Remark on page 27). Thus, $\dim H = 1$. But H is closed in $SU(2)$ so it too is compact and $H \cong U(1)\,(S^1$ is the only compact, connected 1-manifold; see Section 5–11 of [**N1**]). The ground states of our $SU(2)$ Yang-Mills-Higgs theory are therefore invariant only under a $U(1)$ subgroup of $SU(2)$. This is an instance of the phenomenon of *spontaneous symmetry breaking* in which a field theory with an exact symmetry group G (e.g., $SU(2)$) gives rise to ground states that are invariant only under some proper subgroup H (e.g., $U(1)$) of G.

With these few remarks behind us we now turn to the case ($\lambda = 0$) of most interest to us. More precisely, we consider the action

$$\begin{aligned} A(\boldsymbol{\mathcal{A}}, \phi) &= \int_{\ ^3} \Big(-\operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge {}^*\boldsymbol{\mathcal{F}}) - \operatorname{trace}(d^{\mathcal{A}}\phi \wedge {}^*d^{\mathcal{A}}\phi) \Big) \\ &= \frac{1}{2} \int_{\ ^3} \Big(\|\boldsymbol{\mathcal{F}}\|^2 + \|d^{\mathcal{A}}\,\phi\|^2 \Big)\, dx^1 \wedge dx^2 \wedge dx^3 \end{aligned} \tag{2.5.10}$$

and take as our configuration space

$$C = \left\{ (\boldsymbol{\mathcal{A}}, \phi) : A(\boldsymbol{\mathcal{A}}, \phi) < \infty,\ \lim_{R \longrightarrow \infty}\ \sup_{|x| \geq R} \big|1 - \|\phi\|\big| = 0 \right\} \tag{2.5.11}$$

(the second condition reflecting our decision to restrict attention to finite action configurations for which $\|\phi\| \longrightarrow 1$ as $|x| \longrightarrow \infty$ in $\ ^3$). Writing out the Euler-Lagrange equations for the action (2.5.11) yields what are called the **Yang-Mills-Higgs (YMH) equations**

$$\begin{cases} {}^*d^{\mathcal{A}}\,{}^*\boldsymbol{\mathcal{F}} = [d^{\mathcal{A}}\phi, \phi] \\ {}^*d^{\mathcal{A}}\;{}^*d^{\mathcal{A}}\phi = 0 \end{cases}. \tag{2.5.12}$$

The configuration $(\boldsymbol{\mathcal{A}}, \phi)$ must also satisfy the following *Bianchi identities*:

$$\begin{cases} d^{\mathcal{A}}\,\boldsymbol{\mathcal{F}} = 0 \\ d^{\mathcal{A}}\,d^{\mathcal{A}}\phi = [\boldsymbol{\mathcal{F}}, \phi] \end{cases}. \tag{2.5.13}$$

Thus we are looking for solutions to (2.5.12) that live in C. We shall find some interesting ones, but not by studying (2.5.12) directly. As it happens, there is a simpler set of first order equations whose solutions necessarily also satisfy (2.5.12) and, in fact, give the absolute minima of the action (2.5.10).

Remark: This is entirely analogous to the situation encountered in [**N4**], where the (anti-) self-dual equations on ^{4} gave the absolute minima of the Yang-Mills action. Shortly we will review that situation and find that there is a closer connection than simple analogy.

The best way to see where these equations come from is as follows: We denote by $\langle\ ,\ \rangle$ the inner product we have defined on $su(2)$-valued forms on ^{3} (Remark, page 130). Notice that, on ^{3}, $\boldsymbol{\mathcal{F}}$ and ${}^*d^{\mathcal{A}}\phi$ are both 2- forms and (since the metric on ^{3} is Riemannian), $\|d^{\mathcal{A}}\phi\|^2 = \|{}^*d^{\mathcal{A}}\phi\|^2$. Now notice that

$$\begin{aligned} \|\boldsymbol{\mathcal{F}}\|^2 + \left\|d^{\mathcal{A}}\phi\right\|^2 &= \|\boldsymbol{\mathcal{F}}\|^2 + \left\|{}^*d^{\mathcal{A}}\phi\right\|^2 \\ &= \langle \boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}\rangle + \left\langle {}^*d^{\mathcal{A}}\phi,\, {}^*d^{\mathcal{A}}\phi \right\rangle \\ &= \left\langle \boldsymbol{\mathcal{F}} - {}^*d^{\mathcal{A}}\phi, \boldsymbol{\mathcal{F}} - {}^*d^{\mathcal{A}}\phi \right\rangle + 2\left\langle \boldsymbol{\mathcal{F}}, {}^*d^{\mathcal{A}}\phi \right\rangle \\ &= \left\| \boldsymbol{\mathcal{F}} - {}^*d^{\mathcal{A}}\phi \right\|^2 + 2\left\langle \boldsymbol{\mathcal{F}}, {}^*d^{\mathcal{A}}\phi \right\rangle \end{aligned}$$

and, similarly,

$$\|\boldsymbol{\mathcal{F}}\|^2 + \left\|d^{\mathcal{A}}\phi\right\|^2 = \left\|\boldsymbol{\mathcal{F}} + {}^*d^{\mathcal{A}}\phi\right\|^2 - 2\left\langle \boldsymbol{\mathcal{F}}, {}^*d^{\mathcal{A}}\phi \right\rangle.$$

It follows that $(\boldsymbol{\mathcal{A}}, \phi)$ achieves an absolute minimum when

$$\boldsymbol{\mathcal{F}} = \pm{}^*d^{\mathcal{A}}\phi. \tag{2.5.14}$$

These are the (**Bogomolny**) **monopole equations** and any configuration $(\boldsymbol{\mathcal{A}}, \phi)$ which satisfies them also satisfies the YMH equations (2.5.12) (either observe that an absolute minimum for $A(\boldsymbol{\mathcal{A}}, \phi)$ is necessarily a stationary value, or simply substitute $\boldsymbol{\mathcal{F}} = \pm{}^*d^{\mathcal{A}}\phi$ into (2.5.12) and use the Bianchi identities (2.5.13)). For the record we write out these equations in standard coordinates on ^{3}, using a formula for the Hodge dual of a 1-form on ^{3}

that we will prove in Chapter 4. With $\boldsymbol{\mathcal{A}} = \mathcal{A}_k\,dx^k$ and $\boldsymbol{\mathcal{F}} = \frac{1}{2}\,\mathcal{F}_{ij}\,dx^i \wedge dx^j$, (2.5.14) becomes

$$\mathcal{F}_{ij} = \pm \sum_{k=1}^{3} \epsilon_{ijk}(\partial_k\,\phi + [\mathcal{A}_k, \phi]), \quad i, j = 1, 2, 3, \tag{2.5.15}$$

where ϵ_{ijk} is the Levi-Civita symbol (anti-symmetric in $i\,j\,k$ and $\epsilon_{123} = 1$).

We will have something to say shortly about why these are called "monopole" equations. First we wish to give some indication of how the example before us now actually arises quite naturally out of the pure Yang-Mills theory on ^{4} discussed in [**N4**]. This is the $n = 4$ case of the model we have constructed when, in the action (2.5.5), $\lambda = 0$ and $\phi = 0$. In this special case the action (which now depends only on $\boldsymbol{\mathcal{A}}$) is called the **Yang-Mills-action** and written

$$\begin{aligned} \mathcal{YM}(\boldsymbol{\mathcal{A}}) &= -\int_{\ {}^4} \text{trace}(\boldsymbol{\mathcal{F}} \wedge {}^*\boldsymbol{\mathcal{F}}) \\ &= \frac{1}{2}\int_{\ {}^4} \|\boldsymbol{\mathcal{F}}\|^2\,dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4. \end{aligned} \tag{2.5.16}$$

The Euler-Lagrange equations now become the **Yang-Mills equations**

$$d^{\mathcal{A}}\,{}^*\boldsymbol{\mathcal{F}} = 0, \tag{2.5.17}$$

while the Bianchi identities become

$$d^{\mathcal{A}}\,\boldsymbol{\mathcal{F}} = 0. \tag{2.5.18}$$

Now notice that if we have a potential $\boldsymbol{\mathcal{A}}$ for which the field strength $\boldsymbol{\mathcal{F}}$ is **self-dual** (**SD**)

$${}^*\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{F}}, \tag{2.5.19}$$

or **anti-self-dual** (**ASD**)

$${}^*\boldsymbol{\mathcal{F}} = -\boldsymbol{\mathcal{F}}, \tag{2.5.20}$$

then the Bianchi identity (2.5.18) implies that $\boldsymbol{\mathcal{F}}$ necessarily satisfies the Yang-Mills equations (2.5.17). Such potentials $\boldsymbol{\mathcal{A}}$ are called $SU(2)$ **instantons** on ^{4} and, in [**N4**], a number of examples (called the **BPST instantons**) were described. Written in quaternionic notation (i.e., identifying ^{4} with and $su(2)$ with the Lie algebra Im of pure imaginary quaternions) these can be written

$$\boldsymbol{\mathcal{A}}_{\lambda,n}(q) = \text{Im}\left(\frac{\bar{q} - \bar{n}}{\lambda^2 + |q - n|^2}\,dq\right), \tag{2.5.21}$$

where $\lambda > 0$ and $n \in$ are parameters called the **scale** and **center** of the instanton, respectively. The corresponding field strengths are

$$\boldsymbol{\mathcal{F}}_{\lambda,n}(q) = \frac{\lambda^2}{(\lambda^2 + |q - n|^2)^2}\,d\bar{q} \wedge dq. \tag{2.5.22}$$

These are ASD and satisfy

$$\frac{1}{2}\,\|\boldsymbol{\mathcal{F}}_{\lambda,n}(q)\|^2 = \frac{48\lambda^2}{(\lambda^2+|q-n|^2)^4} \tag{2.5.23}$$

and

$$\mathcal{YM}(\boldsymbol{\mathcal{A}}_{\lambda,n}) = \int_{\ ^4} \frac{48\,\lambda^2}{(\lambda^2+|q-n|^2)^4}\, dq^0\, dq^1\, dq^2\, dq^3 = 8\pi^2. \tag{2.5.24}$$

Notice that all of these potentials have the same Yang-Mills action (total field strength). For a fixed center n, $\|\boldsymbol{\mathcal{F}}_{\lambda,n}(q)\|^2$ has a maximum value of $96/\lambda^2$ at $q = n$ and, as $\lambda \longrightarrow 0$, this maximum value approaches infinity in such a way that the integrals over $\ ^4$ remain constant at $8\pi^2$.

The fact that all of these BPST potentials $\boldsymbol{\mathcal{A}}_{\lambda,n}$ have the same Yang-Mills action has a much more profound significance than may be apparent at first glance. It was shown in [**N4**] (see also Examples 3 and 4, pages 40–42, of Chapter 1) that each $\boldsymbol{\mathcal{A}}_{\lambda,n}$ is the pullback to $\ ^4$ via stereographic projection of a connection $\boldsymbol{\omega}_{\lambda,n}$ on the Hopf bundle $SU(2) \hookrightarrow S^7 \longrightarrow \ ^1$. Thinking of $\ ^1 \cong S^4$ as the one-point compactification of $\ ^4$, one can reverse one's point of view here and say that, by virtue of the asymptotic behavior implicit in the fact that $\mathcal{YM}(\boldsymbol{\mathcal{A}}_{\lambda,n}) < \infty$, each $\boldsymbol{\mathcal{A}}_{\lambda,n}$ "extends to the point at infinity." Indeed, Uhlenbeck's Removable Singularities Theorem asserts that, if $\boldsymbol{\mathcal{A}}$ is an $SU(2)$ gauge potential on $\ ^4$ with $\mathcal{YM}(\boldsymbol{\mathcal{A}}) < \infty$, then $\boldsymbol{\mathcal{A}}$ always "extends to the point at infinity" in the sense that there exists a unique $SU(2)$-bundle over S^4 and a connection $\boldsymbol{\omega}$ on it such that $\boldsymbol{\mathcal{A}} = (s \circ \varphi^{-1})^*\boldsymbol{\omega}$, where s is a cross-section of the bundle defined on the complement of some point in S^4 and φ is a stereographic projection to $\ ^4$. The specific bundle to which the potential $\boldsymbol{\mathcal{A}}$ "extends" is, moreover, determined by the value of $\mathcal{YM}(\boldsymbol{\mathcal{A}})$. It is essential to understand precisely what is being asserted here so, before returning to the Bogomolny monopole equations, we will elaborate.

According to the Classification Theorem (page 34), the set of equivalence classes of principal $SU(2)$-bundles over S^4 is in one-to-one correspondence with the elements of the homotopy group $\pi_3(SU(2)) \cong \pi_3(S^3) \cong \mathbb{Z}$. There are various ways of associating with each $SU(2)$-bundle over S^4 an integer that characterizes it up to equivalence, but from our point of view the most useful of these arises in the theory of characteristic classes. We have already had one brief encounter with characteristic classes (the first Chern class, in Section 2.2) and will return to the general theory in Chapter 6. For the present we will content ourselves with a brief, informal description of those aspects of the subject relevant to characterizing a principal $SU(2)$-bundle

$$SU(2) \hookrightarrow P \xrightarrow{\mathcal{P}} S^4$$

over S^4 up to equivalence.

Choose any connection $\boldsymbol{\omega}$ on the bundle and let $\boldsymbol{\Omega}$ denote its curvature (we will prove in Chapter 3 that connections exist on any smooth principal bundle). The local field strengths $\tilde{\mathcal{F}} = s^* \boldsymbol{\omega}$ for various cross-sections generally do not agree on the intersections of their domains and so do not piece together into a globally defined 2-form on all of S^4. Indeed, we know that if s^g is another cross-section, then $\tilde{\mathcal{F}}^g = g^{-1}\tilde{\mathcal{F}}\, g$ on the intersection. However, certain algebraic combinations of the local field strengths can be found which *do* agree on the intersections and so do piece together into globally defined forms on S^4. Essentially, all that is required is a symmetric, multilinear function on $su(2)$ that is ad-invariant, i.e., takes the same values at A and $g^{-1}Ag$ for all $A \in su(2)$ and $g \in SU(2)$. The trace is an obvious choice and this gave rise, in Section 2.2, to the first Chern class which, for $U(1)$-bundles over spacetime, is the cohomology class of the electromagnetic field strength. For $SU(2)$-bundles, however, the local field strengths take values in $su(2)$ where everything has trace zero so this will not get us very far. The trace of the product would be the next likely candidate and this gets us very far indeed. We define a 4-form on all of S^4 by decreeing that, relative to any local cross-section, it is given by

$$\frac{1}{8\pi^2}\text{trace}(\tilde{\mathcal{F}} \wedge \tilde{\mathcal{F}})$$

(the $\frac{1}{8\pi^2}$ is a normalizing constant whose purpose we will describe shortly). Now, *a priori* this is a complex 4-form on S^4, but it can be shown to be real-valued (because $\tilde{\mathcal{F}}$ is skew-Hermitian). Being a 4-form on the 4-dimensional manifold S^4, it is also closed (i.e., has exterior derivative zero) and therefore determines a de Rham cohomology class

$$c_2(P) = \frac{1}{8\pi^2}\left[\text{trace}(\tilde{\mathcal{F}} \wedge \tilde{\mathcal{F}})\right] \in H^4_{\text{de R}}(S^4). \tag{2.5.25}$$

Remarkably, this cohomology class does not depend on the initial choice of the connection $\boldsymbol{\omega}$ from which it arose; it is a characteristic class for the bundle. $c_2(P)$ is called the **second Chern class** of the bundle. Any representative of the class $c_2(P)$ is a 4-form on the compact 4-manifold S^4 and so can be integrated over S^4. It will follow from Stokes' Theorem that two forms in the same cohomology class have the same integral so, in effect, we may integrate $c_2(P)$ over S^4. The result is written

$$c_2(P)[S^4] = \int_{S^4} c_2(P) = \frac{1}{8\pi^2}\int_{S^4}\text{trace}(\tilde{\mathcal{F}} \wedge \tilde{\mathcal{F}}) \tag{2.5.26}$$

and called the **second Chern number** of the bundle (physicists call $-c_2(P)[S^4]$ the **topological charge** or **instanton number** of the bundle). The $\frac{1}{8\pi^2}$ ensures that $c_2(P)[S^4]$ is an *integer* (not obvious, but true) and it is this integer that labels the equivalence classes of $SU(2)$-bundles over S^4. More precisely, two principal $SU(2)$-bundles over S^4 are equivalent if and only if their second Chern numbers are equal.

To relate this to our gauge potentials on ^{4} recall that there is a stereographic projection φ, defined at all but one point of S^4, that is an orientation preserving (conformal) diffeomorphism onto ^{4}. The integrals we define will be invariant under such maps and unaffected by the omission of one point so $c_2(P)[S^4]$ can be computed by integrating over ^{4} the pullback of $\frac{1}{8\pi^2}$ trace$(\tilde{\boldsymbol{\mathcal{F}}} \wedge \tilde{\boldsymbol{\mathcal{F}}})$ by φ. Pullback commutes with trace and the wedge product and $(\varphi^{-1})^* \tilde{\boldsymbol{\mathcal{F}}} = (\varphi^{-1})^* (s^* \varOmega) = (s \circ \varphi^{-1})^* \boldsymbol{\varOmega} = \boldsymbol{\mathcal{F}}$ is a gauge potential on ^{4}. Thus,

$$c_2(P)[S^4] = \frac{1}{8\pi^2} \int_{\ 4} \operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}). \tag{2.5.27}$$

Now here's the good part: If $\boldsymbol{\mathcal{A}}$ is an ASD potential on ^{4}, then ${}^*\boldsymbol{\mathcal{F}} = -\boldsymbol{\mathcal{F}}$ so

$$\begin{aligned} \mathcal{YM}(\boldsymbol{\mathcal{A}}) &= \int_{\ 4} -\operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge {}^*\boldsymbol{\mathcal{F}}) \\ &= \int_{\ 4} \operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}) \qquad (ASD). \end{aligned} \tag{2.5.28}$$

If this is finite, the Removable Singularities Theorem assures us that $\boldsymbol{\mathcal{A}}$ extends to some principal $SU(2)$-bundle over S^4. The Chern number of this bundle can be computed from any connection on it so we might as well use the extension of $\boldsymbol{\mathcal{A}}$. Then, comparing (2.5.27) and (2.5.28) gives

$$c_2(P)[S^4] = \frac{1}{8\pi^2} \mathcal{YM}(\boldsymbol{\mathcal{A}}) \qquad (ASD). \tag{2.5.29}$$

The Yang-Mills action of $\boldsymbol{\mathcal{A}}$ (i.e., its total field strength) is directly encoded in the topology of the bundle to which $\boldsymbol{\mathcal{A}}$ extends as its Chern number. Since the Yang-Mills action is determined by the asymptotic behavior of the field strength $\boldsymbol{\mathcal{F}}$ as $|x| \longrightarrow \infty$ in ^{4}, we find that it is this asymptotic behavior that determines the bundle to which $\boldsymbol{\mathcal{A}}$ extends. This phenomenon of boundary conditions on physical fields manifesting themselves as topology will be a recurrent theme here. Notice, in particular, that, since the Chern number of an $SU(2)$-bundle over S^4 is an integer, the possible asymptotic boundary conditions for finite action, ASD potential on ^{4} fall into countably many, discrete "topological types."

Remark: We have already seen that all of the BPST potentials $\boldsymbol{\mathcal{A}}_{\lambda,n}$ are ASD and have Yang-Mills action $8\,\pi^2$. They must, of course, have the same Yang-Mills action since they all extend to the same bundle, i.e., the Hopf bundle (which we now see has Chern number 1).

In order to establish contact with the Bogomolny monopole equations we consider an arbitrary potential $\hat{\boldsymbol{\mathcal{A}}} = \hat{\mathcal{A}}_\alpha \, dx^\alpha$ on ^{4} with field strength $\hat{\boldsymbol{\mathcal{F}}} = \frac{1}{2} \hat{\mathcal{F}}_{\alpha\beta} \, dx^\alpha \wedge dx^\beta$ and use (6.4.4) of **[N4]** to write the components of the Hodge dual ${}^*\hat{\boldsymbol{\mathcal{F}}}$ as

$$ {}^*\hat{\mathcal{F}}_{\alpha\beta} = \frac{1}{2} \sum_{\gamma,\delta=1}^{4} \epsilon_{\alpha\beta\gamma\delta}\, \hat{\mathcal{F}}_{\alpha\beta}, \quad \alpha, \beta = 1, 2, 3, 4, $$

where $\epsilon_{\alpha\beta\gamma\delta}$ is the Levi-Civita symbol (totally anti-symmetric in $\alpha\,\beta\,\gamma\,\delta$ with $\epsilon_{1234} = 1$). Then the (anti-) self-duality equations take the form

$$ \hat{\mathcal{F}}_{\alpha\beta} = \mp\frac{1}{2} \sum_{\gamma,\delta=1}^{4} \epsilon_{\alpha\beta\gamma\delta}\, \hat{\mathcal{F}}_{\alpha\beta}, \quad \alpha, \beta = 1, 2, 3, 4. \tag{2.5.30} $$

Using i, j and k for indices taking the values 1, 2 and 3 one finds (by just writing them out) that all of these equations are contained in the following:

$$ \hat{\mathcal{F}}_{ij} = \pm \sum_{k=1}^{3} \epsilon_{ijk} \hat{\mathcal{F}}_{k4}, \quad i, j = 1, 2, 3, \tag{2.5.31} $$

where ϵ_{ijk} is totally anti-symmetric in ijk and $\epsilon_{123} = 1$. For example, taking $\alpha = 3$ and $\beta = 4$ in (2.5.30) gives

$$ \begin{aligned} \hat{\mathcal{F}}_{34} &= \mp\frac{1}{2} \sum_{\gamma,\delta=1}^{4} \epsilon_{34\gamma\delta} \hat{\mathcal{F}}_{\gamma\delta} = \mp\frac{1}{2}\left[\epsilon_{3412}\hat{\mathcal{F}}_{12} + \epsilon_{3421}\hat{\mathcal{F}}_{21}\right] \\ &= \mp\frac{1}{2}\left[2\epsilon_{3412}\hat{\mathcal{F}}_{12}\right] = \mp\,\epsilon_{3412}\hat{\mathcal{F}}_{12} = \pm\,\epsilon_{4123}\hat{\mathcal{F}}_{12} \\ &= \pm\,\epsilon_{123}\hat{\mathcal{F}}_{12} \end{aligned} $$

which is equivalent to

$$ \hat{\mathcal{F}}_{12} = \pm\,\epsilon_{123}\hat{\mathcal{F}}_{34} $$

and this is (2.5.31) with $i = 1$ and $j = 2$. Similarly, taking $\alpha = 1$ and $\beta = 2$ in (2.5.30) gives

$$ \hat{\mathcal{F}}_{12} = \mp\,\epsilon_{1234}\hat{\mathcal{F}}_{34} = \pm\,\epsilon_{123}\hat{\mathcal{F}}_{34} $$

which is also (2.5.31) with $i = 1$ and $j = 2$.

Thus, the (anti-) self-duality equations on ^{4} take the form (2.5.31). Finite action solutions to these equations are the instantons discussed earlier. We now wish to seek solutions to (2.5.31) that are *static*, i.e., for which the $\hat{\mathcal{A}}_\alpha$ are independent of x^4. Naturally, no such solution can have finite action on ^{4} (unless it is zero) and so cannot be an instanton. For $\hat{\mathcal{A}}_\alpha$'s that are independent of x^4, (2.5.31) becomes

$$\begin{aligned}
\hat{\mathcal{F}}_{ij} &= \pm \sum_{k=1}^{3} \epsilon_{ijk} \hat{\mathcal{F}}_{k4} \\
&= \pm \sum_{k=1}^{3} \epsilon_{ijk} \left(\partial_k \hat{\mathcal{A}}_4 - \partial_4 \hat{\mathcal{A}}_k + \left[\hat{\mathcal{A}}_k, \hat{\mathcal{A}}_4 \right] \right) \\
\hat{\mathcal{F}}_{ij} &= \pm \sum_{k=1}^{3} \epsilon_{ijk} \left(\partial_k \hat{\mathcal{A}}_4 + \left[\hat{\mathcal{A}}_k, \hat{\mathcal{A}}_4 \right] \right), \quad i,j = 1,2,3
\end{aligned} \tag{2.5.32}$$

(you may wish to glance back at (2.5.15) if you're wondering where all of this is going).

Now we "reduce to ^{3}" as follows: Fix some value x_0^4 of x^4 and consider the submanifold $^3 \times \{x_0^4\}$ of ^{4} (which we henceforth identify with ^{3}). Restrict our trivial $SU(2)$-bundle over ^{4} to this ^{3} and obtain a trivial $SU(2)$-bundle over ^{3}. Let $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ be the restrictions to ^{3} of $\hat{\mathcal{A}}_1, \hat{\mathcal{A}}_2$ and $\hat{\mathcal{A}}_3$ (all of which are assumed independent of x^4). Then

$$\boldsymbol{\mathcal{A}} = \mathcal{A}_1 dx^1 + \mathcal{A}_2 dx^2 + \mathcal{A}_3 dx^3 = \mathcal{A}_i dx^i$$

is a gauge potential on ^{3} and the corresponding field strength $\boldsymbol{\mathcal{F}} = \frac{1}{2}\mathcal{F}_{ij} dx^i \wedge dx^j$ has components $\mathcal{F}_{ij} = \partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i + [\mathcal{A}_i, \mathcal{A}_j]$ that are just the restrictions to ^{3} of the $\hat{\mathcal{F}}_{ij}$, $i,j = 1,2,3$. Note that $\hat{\mathcal{A}}_4$ does not enter into either the potential or the field strength on ^{3}. However, if we define $\phi :\ ^3 \longrightarrow su(2)$ by

$$\phi = \hat{\mathcal{A}}_4 \mid\ ^3,$$

then the time independent, (anti-) self-duality equations (2.5.32) on ^{4} become (when restricted to ^{3}) the Bogomolny monopole equations

$$\mathcal{F}_{ij} = \pm \sum_{k=1}^{3} \epsilon_{ijk} (\partial_k \phi + [\mathcal{A}_k,\ \phi]), \quad i,j = 1,2,3. \tag{2.5.33}$$

Thus, an $SU(2)$ Bogomolny monopole on ^{3} is essentially just a static, (anti-) self-dual gauge potential on ^{4}. We intend to describe the geometry and topology of these monopoles in more detail, but first we will exhibit a concrete example that will, in some sense, bring us full circle. We began (in Chapter 0, [**N4**]) by thinking about Dirac's magnetic monopoles and finding that they were most naturally modeled by connections on $U(1)$-bundles over S^2. In particular, the monopole of lowest strength was identified with the natural connection on the complex Hopf bundle. Seeking generalizations we looked at the natural connection on the quaternionic Hopf bundle and found lurking there a BPST instanton. Now we find that a "static instanton" is to be identified with a particular type of Yang-Mills-Higgs field on ^{3} which we have (for reasons that no doubt remain obscure) called a monopole. To understand the terminology and to see Dirac monopoles in an entirely new light we will describe now the **t'Hooft-Polyakov-Prasad-Sommerfleld monopole,** which is an exact solution to the equations (2.5.33). For this we will need some notation. First, in $\ ^3$ we will write

$$\begin{aligned} x &= (x^1, x^2, x^3) \\ r &= |x| = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2} \\ n^i &= x^i/r, \quad i = 1, 2, 3 \quad (r \neq 0) \\ d\vec{x} &= (dx^1, dx^2, dx^3) \end{aligned}$$

and, in $su(2)$,

$$\begin{aligned} T_a &= -\frac{1}{2}\, \boldsymbol{i}\, \sigma_a, \quad a = 1, 2, 3 \\ \vec{T} &= (T_1, T_2, T_3). \end{aligned}$$

In addition we set

$$\begin{aligned} \vec{n} \cdot \vec{T} &= n^1\, T_1 + n^2\, T_2 + n^3\, T_3 = n^a\, T_a \\ \vec{n} \times \vec{T} &= (n^2\, T_3 - n^3\, T_2, n^3\, T_1 - n^1\, T_3, n^1\, T_2 - n^2\, T_1) \end{aligned}$$

and

$$\begin{aligned} \left(\vec{n} \times \vec{T}\right) \cdot d\vec{x} = {} & (n^2\, T_3 - n^3\, T_2)\, dx^1 + (n^3\, T_1 - n^1\, T_3)\, dx^2 \\ & + (n^1\, T_2 - n^2\, T_1)\, dx^3. \end{aligned}$$

Our objective is to find $(\boldsymbol{\mathcal{A}}(x), \phi(x))$ which satisfies (2.5.33) and has the required asymptotic behavior ($\|\phi\| \longrightarrow 1$ as $r \longrightarrow \infty$).

Although the monopole equations (2.5.33) are substantially less complicated that the full Yang-Mills-Higgs equations, they are still far beyond the means of elementary techniques. To reduce the level of difficulty a bit more requires a guess (physicists prefer the term *Ansatz*) as to the form one might expect for a solution. Here's the one that worked for t'Hooft and Polyakov (some rationale for the *Ansatz* is discussed in Section 4.2 of [**GO**]): We will seek functions $f(r)$ and $h(r)$ satisfying

$$\frac{f(r)}{r} \longrightarrow 1 \quad \text{as} \quad r \longrightarrow \infty \quad \text{and} \quad f(0) = 0, \tag{2.5.34}$$

and

$$h\,(r) \longrightarrow 0 \quad \text{as} \quad r \longrightarrow \infty \quad \text{and} \quad h(0) = 1, \tag{2.5.35}$$

such that

$$\phi\,(x) = \frac{f(r)}{r}\left(\vec{n} \cdot \vec{T}\right) = \frac{f(r)}{r^2}\, x^a\, T_a \tag{2.5.36}$$

and

$$\begin{aligned} \boldsymbol{\mathcal{A}}(x) &= \frac{1-h(r)}{r}\left(\vec{n}\times\vec{T}\right)\cdot d\vec{x} \\ &= \frac{1-h(r)}{r^2}\sum_{a=1}^{3}\epsilon_{aij}\,x^j\,dx^i\,T_a \end{aligned} \tag{2.5.37}$$

give a solution $(\boldsymbol{\mathcal{A}}, \phi)$ to (2.5.33).

Remark: The conditions $f(0) = 0$ and $h(0) = 1$ are intended to give $\phi(x)$ and $\boldsymbol{\mathcal{A}}(x)$ a chance of being smooth at the origin.

Substituting (2.5.36) and (2.5.37) into (2.5.33) (with the minus sign) gives the following system of ordinary differential equations for $f(r)$ and $h(r)$:

$$\begin{cases} r\,\dfrac{dh}{dr} = -hf \\ r\,\dfrac{df}{dr} = f - (h^2 - 1). \end{cases} \tag{2.5.38}$$

Now make the change of variable

$$-f = 1 + rF, \quad h = rH$$

to obtain the new system

$$\begin{cases} \dfrac{dH}{dr} = HF \\ \dfrac{dF}{dr} = H^2 \end{cases}$$

for F and H. Observe that it follows from these equations that $F^2 - H^2$ is constant since

$$\frac{d}{dr}\,(F^2 - H^2) = 2F\,\frac{dF}{dr} - 2H\,\frac{dH}{dr} = 2FH^2 - 2H(HF) = 0.$$

But since, for $r > 0$, $F^2 - H^2 = \dfrac{(1+f)^2}{r^2} - \dfrac{h^2}{r^2} = \dfrac{1}{r^2} + \dfrac{2}{r}\left(\dfrac{f}{r}\right) + \left(\dfrac{f}{r}\right)^2 - \dfrac{1}{r^2}\,h^2$ and this is to approach 1 as $r \longrightarrow \infty$, we must have

$$F^2 - H^2 = 1.$$

To get solutions f and h satisfying $f(0) = 0$ and $h(0) = 1$ we take for F and H satisfying $F^2 - H^2 = 1$ the following:

$$F(r) = -\coth r \quad \text{and} \quad H(r) = \operatorname{csch} r.$$

Then

$$f(r) = r\coth r - 1 \quad \text{and} \quad h(r) = r\operatorname{csch} r$$

and one can verify directly that these are smooth (even at $r = 0$, where they take the required boundary values) and satisfy (2.5.38). Thus, our field configuration $(\boldsymbol{\mathcal{A}}, \phi)$ is given by

$$\phi(x) = \left(\coth r - \frac{1}{r}\right) \vec{n} \cdot \vec{T} = \frac{1}{r}\left(\coth r - \frac{1}{r}\right) x^a \, T_a \tag{2.5.39}$$

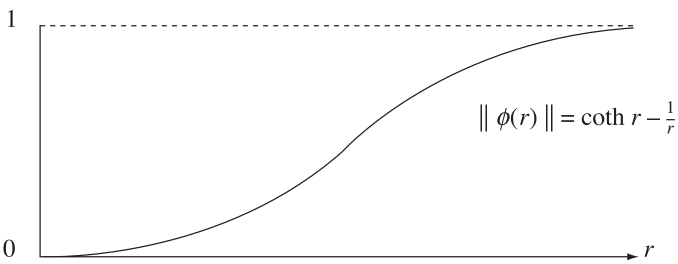
$$\begin{aligned} \boldsymbol{\mathcal{A}}(x) &= \left(\frac{1}{r} - \operatorname{csch} r\right)\left(\vec{n} \times \vec{T}\right) \cdot d\vec{x} \\ &= \frac{1}{r}\left(\frac{1}{r} - \operatorname{csch} r\right) \sum_{a=1}^{3} \epsilon_{aij} \, x^j \, dx^i \, T_a \end{aligned} \tag{2.5.40}$$

Remark: The same calculation for (2.5.33) with the plus sign gives the same result except that the sign of ϕ is changed.

The exact solution $(\boldsymbol{\mathcal{A}}, \phi)$ given by (2.5.39) and (2.5.40) is remarkable for a number of reasons (quite aside from the fact that it *is* an exact solution which is more than one generally has a right to expect). The form of the thing itself is extraordinary for its "mixing" of the spatial and internal directions. For example, (2.5.39) describes a Higgs field which, in the x^a-direction in ^{3}, $a = 1, 2, 3$, has only an internal T^a-component. It is, in some strange way, "radial" (Polyakov called it a "hedgehog" solution). Another feature, and one that will be particularly significant quite soon, is that, despite appearances, the component functions of ϕ and $\boldsymbol{\mathcal{A}}$ are smooth (in fact, real analytic), even at the origin. For example,

$$\begin{aligned} \coth r - \frac{1}{r} = \frac{1}{r}\left(\frac{r \cosh r}{\sinh r} - 1\right) &= \frac{1}{r}\left(\frac{r + \frac{1}{2!} r^3 + \frac{1}{4!} r^5 + \cdots}{r + \frac{1}{3!} r^3 + \frac{1}{5!} r^5 + \cdots} - 1\right) \\ &= \frac{1}{r}\left(\frac{1 + \frac{1}{2!} r^2 + \frac{1}{4!} r^4 + \cdots}{1 + \frac{1}{3!} r^2 + \frac{1}{5!} r^4 + \cdots} - 1\right) \\ &= \frac{1}{r}\left(1 + \left(\frac{1}{2!} - \frac{1}{3!}\right) r^2 + \cdots - 1\right) \end{aligned}$$

(by long division) and this is indeed analytic at $r = 0$. Computing the derivative of $\coth r - \frac{1}{r}$ one finds that it is positive. Moreover, $\lim_{r \longrightarrow \infty}(\coth r - \frac{1}{r}) = 1$ so one obtains the following picture of $\|\phi\|$:



The configuration $(\boldsymbol{\mathcal{A}}, \phi)$ is a globally defined, smooth object on all of ^{3}.

But how did configurations such as this come to be called "monopoles"? This is not entirely clear from the "hedgehog" form in which we currently have the solution written, but will become clear after an appropriate (local) gauge transformation. We intend to define a gauge transformation on an open subset of ^{3} which, at each point in space, rotates the Higgs field (in the internal space $su(2)$ at that point) from its "radial" direction to a direction parallel to the internal T_3-axis. One should keep in mind that a gauge transformation is assumed to change only the appearance, not the physics of field configurations. Our gauge transformation will be a map g from an open subset of ^{3} into $SU(2)$ and its effect on ϕ and $\boldsymbol{\mathcal{A}}$ will be, as usual,

$$\phi \longrightarrow \phi^g = g^{-1}\phi g$$
$$\boldsymbol{\mathcal{A}} \longrightarrow \boldsymbol{\mathcal{A}}^g = g^{-1}\,\boldsymbol{\mathcal{A}} g + g^{-1} dg.$$

Specifically, in terms of spherical coordinates (r, φ, θ) on ^{3}, we let

$$g(x) = \begin{pmatrix} \cos\dfrac{\varphi}{2} & -e^{-\boldsymbol{i}\theta}\sin\dfrac{\varphi}{2} \\ e^{\boldsymbol{i}\theta}\sin\dfrac{\varphi}{2} & \cos\dfrac{\varphi}{2} \end{pmatrix}. \tag{2.5.41}$$

To compute ϕ^g we proceed as follows:

$$\phi^g = g^{-1}\phi g = g^{-1}\left(\frac{f(r)}{r^2}x^a T_a\right) g = \frac{f(r)}{r^2}\left(g^{-1}\left(x^a T_a\right) g\right)$$
$$= \frac{f(r)}{r^2}x^a \left(g^{-1}T_a g\right) = -\frac{1}{2}\boldsymbol{i}\frac{f(r)}{r^2}x^a \left(g^{-1}\sigma_a g\right).$$

Now, notice that, for any $g = \left(\begin{smallmatrix} \alpha & \beta \\ -\bar{\beta} & \alpha \end{smallmatrix}\right)$, $\alpha \in$, in $SU(2)$, $g^{-1} = \left(\begin{smallmatrix} \alpha & -\beta \\ \bar{\beta} & \alpha \end{smallmatrix}\right)$ and

$$g^{-1}\sigma_1 g = g^{-1}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g = \begin{pmatrix} -\alpha\bar{\beta} - \alpha\beta & \alpha^2 - \beta^2 \\ -\bar{\beta}^2 + \alpha^2 & \alpha\bar{\beta} + \alpha\beta \end{pmatrix}$$

$$g^{-1}\sigma_2 g = \boldsymbol{i}g^{-1}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} g = \boldsymbol{i}\begin{pmatrix} \alpha\bar{\beta} - \alpha\beta & -\alpha^2 - \beta^2 \\ \bar{\beta}^2 + \alpha^2 & -\alpha\bar{\beta} + \alpha\beta \end{pmatrix}$$

$$g^{-1}\sigma_3 g = g^{-1}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g = \begin{pmatrix} \alpha^2 - \beta\bar{\beta} & 2\alpha\beta \\ 2\alpha\bar{\beta} & \beta\bar{\beta} - \alpha^2 \end{pmatrix}$$

Now, with $g = \left(\begin{smallmatrix} \alpha & \beta \\ -\bar{\beta} & \alpha \end{smallmatrix}\right) = \begin{pmatrix} \cos\frac{\varphi}{2} & -e^{-\boldsymbol{i}\theta}\sin\frac{\varphi}{2} \\ e^{\boldsymbol{i}\theta}\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} \end{pmatrix}$ we have

$$
\begin{aligned}
-\alpha\bar{\beta} - \alpha\beta &= -\alpha(\bar{\beta} + \beta) = -\cos\frac{\varphi}{2}(2\mathrm{Re}(\beta)) \\
&= 2\cos\frac{\varphi}{2}\sin\frac{\varphi}{2}\cos\theta = \sin\varphi\cos\theta = \frac{x^1}{r} \\
\alpha^2 - \beta^2 &= \cos^2\frac{\varphi}{2} - e^{-2\boldsymbol{i}\theta}\sin^2\frac{\varphi}{2} \\
&= \cos^2\frac{\varphi}{2} - (\cos 2\theta - \boldsymbol{i}\sin 2\theta)\sin^2\frac{\varphi}{2} \\
&= \left(\cos^2\frac{\varphi}{2} - \cos 2\theta\sin^2\frac{\varphi}{2}\right) + \boldsymbol{i}\sin 2\theta\sin^2\frac{\varphi}{2} \\
\boldsymbol{i}\left(\alpha\bar{\beta} - \alpha\beta\right) &= \boldsymbol{i}\alpha\left(\bar{\beta} - \beta\right) = \boldsymbol{i}\alpha(-2\boldsymbol{i}\,\mathrm{Im}(\beta)) \\
&= 2\alpha\,\mathrm{Im}\,(\beta) = 2\cos\frac{\varphi}{2}(\sin\theta\sin\frac{\varphi}{2}) \\
&= \sin\varphi\sin\theta = \frac{x^2}{r} \\
-\alpha^2 - \beta^2 &= (-\cos^2\frac{\varphi}{2} - \cos 2\theta\sin^2\frac{\varphi}{2}) + \boldsymbol{i}\sin 2\theta\sin^2\frac{\varphi}{2} \\
\alpha^2 - \beta\bar{\beta} &= \cos^2\frac{\varphi}{2} - \sin^2\frac{\varphi}{2} \\
&= \cos\varphi = \frac{x^3}{r}
\end{aligned}
$$

$$
\begin{aligned}
2\alpha\beta &= 2\cos\frac{\varphi}{2}\left(-e^{-\boldsymbol{i}\theta}\sin\frac{\varphi}{2}\right) = -e^{-\boldsymbol{i}\theta}\sin\varphi \\
&= -\cos\theta\sin\varphi + \boldsymbol{i}\sin\theta\sin\varphi.
\end{aligned}
$$

We need just the (1,1) and (1,2) entries of ϕ^g so we compute as follows: The (1,1) entry of $x^a(g^{-1}\sigma_a g)$ is

$$
\begin{aligned}
&x^1(\sin\varphi\cos\theta) + x^2(\sin\varphi\sin\theta) + x^3\cos\varphi \\
&= x^1\left(\frac{x^1}{r}\right) + x^2\left(\frac{x^2}{r}\right) + x^3\left(\frac{x^3}{r}\right) = \frac{r^2}{r} = r.
\end{aligned}
$$

The real part of the (1,2) entry of $x^a(g^{-1}\sigma_a g)$ is

$$\begin{aligned}
&x^1(\cos^2\frac{\varphi}{2}-\cos 2\theta\sin^2\frac{\varphi}{2})+x^2(-\sin 2\theta\sin^2\frac{\varphi}{2})+x^3(-\cos\theta\sin\varphi)\\
&\quad=(r\sin\varphi\cos\theta)\cos^2\frac{\varphi}{2}-(r\sin\varphi\cos\theta)\cos 2\theta\sin^2\frac{\varphi}{2}\\
&\qquad-(r\sin\varphi\sin\theta)\sin 2\theta\sin^2\frac{\varphi}{2}-r\cos\varphi\cos\theta\sin\varphi\\
&\quad= r\sin\varphi\cos\theta\left(\frac{1}{2}+\frac{1}{2}\cos\varphi\right)\\
&\qquad-r\sin\varphi\cos\theta(1-2\sin^2\theta)\left(\frac{1}{2}-\frac{1}{2}\cos\varphi\right)\\
&\qquad-2r\sin\varphi\sin^2\theta\cos\theta\left(\frac{1}{2}-\frac{1}{2}\cos\varphi\right)-r\cos\varphi\cos\theta\sin\varphi\\
&\quad= \frac{1}{2}r\sin\varphi\cos\theta+\frac{1}{2}r\sin\phi\cos\varphi\cos\theta\\
&\qquad-(r\sin\varphi\cos\theta-2r\sin\varphi\cos\theta\sin^2\theta)\left(\frac{1}{2}-\frac{1}{2}\cos\varphi\right)\\
&\qquad-2r\sin\varphi\cos\theta\sin^2\theta\left(\frac{1}{2}-\frac{1}{2}\cos\varphi\right)-r\cos\varphi\cos\theta\sin\varphi\\
&\quad= \frac{1}{2}r\sin\varphi\cos\theta+\frac{1}{2}r\sin\varphi\cos\varphi\cos\theta\\
&\qquad-(r\sin\varphi\cos\theta)\left(\frac{1}{2}-\frac{1}{2}\cos\varphi\right)-r\sin\varphi\cos\varphi\cos\theta\\
&\quad= \frac{1}{2}r\sin\varphi\cos\theta-\frac{1}{2}r\sin\varphi\cos\varphi\cos\theta\\
&\qquad-\frac{1}{2}r\sin\varphi\cos\theta+\frac{1}{2}r\sin\varphi\cos\varphi\cos\theta\\
&\quad= 0.
\end{aligned}$$

The imaginary part of the (1,2) entry of $x^a(g^{-1}\sigma_a g)$ is shown to be zero in the same way so the (1,2) entry itself is zero. Since ϕ^g takes values in $su(2)$ we have

$$\phi^g=-\frac{1}{2}\boldsymbol{i}\frac{f(r)}{r^2}\begin{pmatrix} r & 0\\ 0 & -r\end{pmatrix}=-\frac{1}{2}\boldsymbol{i}\frac{f(r)}{r}\begin{pmatrix} 1 & 0\\ 0 & -1\end{pmatrix}$$

$$\phi^g=\left(\coth r-\frac{1}{r}\right)T_3. \tag{2.5.42}$$

Thus, as promised, our Higgs field is now aligned along the third isospin axis T_3 in the internal space at each point of ^{3}. As to the gauge potential $\boldsymbol{\mathcal{A}}^g$, we will again need the (1,1) and (1,2) entries in $g^{-1}\boldsymbol{\mathcal{A}}g+g^{-1}dg$. First observe that

$$
\begin{aligned}
g^{-1}\boldsymbol{\mathcal{A}}g &= g^{-1}\left(\frac{1-h(r)}{r^2}\sum_{a=1}^{3}\epsilon_{aij}x^j dx^i T_a\right)g \\
&= \frac{1-h(r)}{r^2}\sum_{a=1}^{3}(\epsilon_{aij}x^j dx^i)\left(g^{-1}T_a g\right) \\
&= -\frac{1}{2}\boldsymbol{i}\frac{1-h(r)}{r^2}\sum_{a=1}^{3}(\epsilon_{aij}x^j dx^i)\left(g^{-1}\sigma_a g\right) \\
&= -\frac{1}{2}\boldsymbol{i}\frac{1-h(r)}{r^2}\Big[(x^3dx^2-x^2dx^3)\left(g^{-1}\sigma_1 g\right) \\
&\qquad + (x^1dx^3-x^3dx^1)(g^{-1}\sigma_2 g) \\
&\qquad + \ (x^2dx^1-x^1dx^2)(g^{-1}\sigma_3 g)\Big].
\end{aligned}
$$

Notice that the (1,1) entry of $g^{-1}\boldsymbol{\mathcal{A}}g$ is zero:

$$
\begin{aligned}
&-\frac{1}{2}\boldsymbol{i}\frac{1-h(r)}{r^2}\left[(x^3dx^2-x^2dx^3)\left(\frac{x^1}{r}\right)+(x^1dx^3-x^3dx^1)\left(\frac{x^2}{r}\right)\right. \\
&\qquad\qquad \left. +(x^2dx^1-x^1dx^2)\left(\frac{x^3}{r}\right)\right] \\
&= -\frac{1}{2}\boldsymbol{i}\frac{1-h(r)}{r^3}\Big[x^1x^3dx^2-x^1x^2dx^3+x^1x^2dx^3-x^2x^3dx^1 \\
&\qquad\qquad +x^2x^3dx^1-x^1x^3dx^2\Big] \\
&= 0.
\end{aligned}
$$

Thus, the (1,1) entry of $\boldsymbol{\mathcal{A}}^g$ is the same as the (1,1) entry of $g^{-1}dg$ (and so, in particular, does not depend on the gauge potential $\boldsymbol{\mathcal{A}}$).
Let $g = \begin{pmatrix}\alpha & \beta\\ -\bar{\beta} & \alpha\end{pmatrix}$ with α real and depending only on φ and β depending only on φ and θ. Then $g^{-1} = \begin{pmatrix}\alpha & -\beta\\ \bar{\beta} & \alpha\end{pmatrix}$ and

$$
dg = \begin{pmatrix}\alpha_\varphi d\varphi & \beta_\varphi d\varphi+\beta_\theta d\theta \\ -\bar{\beta}_\varphi d\varphi-\bar{\beta}_\theta d\theta & \alpha_\varphi d\varphi\end{pmatrix}
$$

so

$$
g^{-1}dg = \begin{pmatrix}\alpha & -\beta\\ \bar{\beta} & \alpha\end{pmatrix}\begin{pmatrix}\alpha_\varphi d\varphi & \beta_\varphi d\varphi+\beta_\theta d\theta \\ -\bar{\beta}_\varphi d\varphi-\bar{\beta}_\theta d\theta & \alpha_\varphi d\varphi\end{pmatrix}
$$

The (1,1) entry of $g^{-1}dg$ is

$$\begin{aligned}
&\alpha\alpha_\varphi d\varphi + \beta\bar{\beta}_\varphi d\varphi + \beta\bar{\beta}_\theta d\theta \\
&\qquad = \cos\frac{\varphi}{2}\left(-\frac{1}{2}\sin\frac{\varphi}{2}\right)d\varphi \\
&\qquad\quad + \left(-e^{-\boldsymbol{i}\theta}\sin\frac{\varphi}{2}\right)\left(-e^{\boldsymbol{i}\theta}\frac{1}{2}\cos\frac{\varphi}{2}\right)d\varphi \\
&\qquad\quad + \left(-e^{-\boldsymbol{i}\theta}\sin\frac{\varphi}{2}\right)\left(-\boldsymbol{i}e^{\boldsymbol{i}\theta}\sin\frac{\varphi}{2}\right)d\theta \\
&\qquad = -\frac{1}{2}\sin\frac{\varphi}{2}\cos\frac{\varphi}{2}d\varphi + \frac{1}{2}\sin\frac{\varphi}{2}\cos\frac{\varphi}{2}d\varphi + \boldsymbol{i}\sin^2\frac{\varphi}{2}d\theta \\
&\qquad = \boldsymbol{i}\sin^2\frac{\varphi}{2}d\theta \\
&\qquad = \frac{1}{2}\boldsymbol{i}(1-\cos\varphi)d\theta.
\end{aligned}$$

Thus, $\boldsymbol{\mathcal{A}}^g$ has the form

$$\begin{aligned}
\boldsymbol{\mathcal{A}}^g &= \begin{pmatrix} \frac{1}{2}\boldsymbol{i}(1-\cos\varphi)d\theta & \text{——} \\ \text{——} & \text{——} \end{pmatrix} \\
&= -\frac{1}{2}\begin{pmatrix} (-(1-\cos\varphi)d\theta)\boldsymbol{i} & \text{——} \\ \text{——} & \text{——} \end{pmatrix}
\end{aligned}$$

which we write as

$$\boldsymbol{\mathcal{A}}^g = -\frac{1}{2}\begin{pmatrix} -\frac{1}{2}(2)\boldsymbol{i}(1-\cos\varphi)d\theta & \text{——} \\ \text{——} & \text{——} \end{pmatrix}. \tag{2.5.43}$$

A similar calculation for the (1,2) entry gives $(\boldsymbol{\mathcal{A}}^g)^1$ and $(\boldsymbol{\mathcal{A}}^g)^2$ as shown below.

$$\begin{aligned}
(\phi^g)^1 &= (\phi^g)^2 = 0 \\
(\phi^g)^3 &= \frac{f(r)}{r} = \coth r - \frac{1}{r} \\
(\boldsymbol{\mathcal{A}}^g)^1 &= -h(r)(\sin\theta\, d\varphi + \cos\theta\sin\varphi\, d\theta) \\
&= -r\,\mathrm{csch}\, r(\sin\theta\, d\varphi + \cos\theta\sin\varphi\, d\theta) \qquad (2.5.44) \\
(\boldsymbol{\mathcal{A}}^g)^2 &= h(r)(\cos\theta\, d\varphi - \sin\theta\sin\varphi\, d\theta) \\
&= r\,\mathrm{csch}\, r(\cos\theta\, d\varphi - \sin\theta\sin\varphi\, d\theta) \\
(\boldsymbol{\mathcal{A}}^g)^3 &= -(1-\cos\varphi)d\theta.
\end{aligned}$$

Now for the good part. We have already observed that $(\boldsymbol{\mathcal{A}}, \phi)$ is a globally defined, smooth field configuration on all of ^{3}. As $r \longrightarrow \infty$ (i.e., as seen from a distance), the Higgs field approaches the constant value T_3 (because $\frac{f(r)}{r} \longrightarrow 1$), whereas $(\boldsymbol{\mathcal{A}}^g)^1$ and $(\boldsymbol{\mathcal{A}}^g)^2$ approach 0 (because $h(r) \longrightarrow 0$), while $(\boldsymbol{\mathcal{A}}^g)^3$, which does not depend on r, remains fixed at $-(1-\cos\varphi)d\theta$. As seen from infinity the potential function $\boldsymbol{\mathcal{A}}$ in this gauge assumes the form

$$-\frac{1}{2}\begin{pmatrix} -\frac{1}{2}(2)\boldsymbol{i}(1-\cos\varphi)d\theta & 0 \\ 0 & \frac{1}{2}(2)\boldsymbol{i}(1-\cos\varphi)d\theta \end{pmatrix}$$

which is essentially just the Im -valued 1-form

$$-\frac{1}{2}(2)\boldsymbol{i}(1-\cos\varphi)d\theta$$

and this is precisely the potential for a Dirac monopole of magnetic charge 2 (see page 68). Seen from afar, the (nonsingular) t' Hooft-Polyakov monopole looks like a Dirac monopole. The "string singularity" (page 4, [**N4**]) of the Dirac monopole now shows up in the fact that our gauge transformation g (defined in terms of spherical coordinates on ^{3}) is singular on the nonpositive x^3-axis.

Remark: The essential point to keep in mind here is that, in classical electrodynamics, monopoles are (but certainly need not be) "put in by hand," whereas, in $SU(2)$ Yang-Mills-Higgs theory, they arise of their own accord as solutions to the field equations.

Since the t'Hooft-Polyakov monopole satisfies $\boldsymbol{\mathcal{F}} = {}^*d^{\mathcal{A}}\phi$ one can compute the field strength $\boldsymbol{\mathcal{F}}$ entirely from $d^{\mathcal{A}}\phi$. In gauge s^g,

$$\begin{aligned} d^{\mathcal{A}}\phi &= d\phi + [\boldsymbol{\mathcal{A}}, \phi] \\ &= d\left(\frac{f(r)}{r}T_3\right) + \left[\boldsymbol{\mathcal{A}}^a T_a, \frac{f(r)}{r}T_3\right] \\ &= d\left(\frac{f(r)}{r}\right)T_3 + \frac{f(r)}{r}\boldsymbol{\mathcal{A}}^a[T_a, T_3] \\ &= d\left(\frac{f(r)}{r}\right)T_3 + \frac{f(r)}{r}\left[\boldsymbol{\mathcal{A}}^1[T_1, T_3] + \boldsymbol{\mathcal{A}}^2[T_2, T_3] + \boldsymbol{\mathcal{A}}^3[T_3, T_3]\right] \\ &= d\left(\frac{f(r)}{r}\right)T_3 + \frac{f(r)}{r}[-\boldsymbol{\mathcal{A}}^1 T_2 + \boldsymbol{\mathcal{A}}^2 T_1] \\ &= \frac{f(r)}{r}(\boldsymbol{\mathcal{A}}^2 T_1 - \boldsymbol{\mathcal{A}}^1 T_2) + \left(\frac{\partial}{\partial r}\frac{f(r)}{r}\right)dr\, T_3 \\ &= \left(\coth r - \frac{1}{r}\right)(\boldsymbol{\mathcal{A}}^2 T_1 - \boldsymbol{\mathcal{A}}^1 T_2) + \left(\frac{1}{r^2} - \operatorname{csch}^2 r\right)dr\, T_3 \end{aligned}$$

$$d^{\mathcal{A}}\phi = \left(\coth r - \frac{1}{r}\right)(r \operatorname{csch} r)\big[(\cos\theta\, d\varphi - \sin\theta \sin\varphi\, d\theta)T_1 \\ + (\sin\theta\, d\varphi + \cos\,\theta\,\sin\,\varphi\, d\theta)T_2\big] \\ + \left(\frac{1}{r^2} - \operatorname{csch}^2 r\right) dr\, T_3 \tag{2.5.45}$$

Let us return to an arbitrary configuration $(\boldsymbol{\mathcal{A}}, \phi)$ satisfying the monopole equations $\boldsymbol{\mathcal{F}} = \pm{}^* d^{\mathcal{A}}\phi$. We have already seen (page 134) that for such a configuration

$$\|\, \boldsymbol{\mathcal{F}} \,\|^2 + \left\|\, d^{\mathcal{A}}\phi \,\right\|^2 = \pm 2 \left\langle \boldsymbol{\mathcal{F}}, {}^* d^{\mathcal{A}}\phi \right\rangle$$

so

$$\begin{aligned} A(\boldsymbol{\mathcal{A}}, \phi) &= \frac{1}{2}\int_{\ 3} \left(\|\, \boldsymbol{\mathcal{F}} \,\|^2 + \|\, d^{\mathcal{A}}\phi\|^2\right)\, dx^1 \wedge dx^2 \wedge dx^3 \\ &= \pm \int_{\ 3} \left\langle \boldsymbol{\mathcal{F}}, {}^* d^{\mathcal{A}}\phi \right\rangle dx^1 \wedge dx^2 \wedge dx^3 \\ &= \mp \int_{\ 3} 2\,\text{trace}\big(\boldsymbol{\mathcal{F}} \wedge {}^{**} d^{\mathcal{A}}\phi\big) \\ &= \mp \int_{\ 3} 2\,\text{trace}\big(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}}\phi\big) \end{aligned}$$

which we now write as

$$A(\boldsymbol{\mathcal{A}}, \phi) = \mp \int_{\ 3} \text{Tr}\big(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}}\phi\big) \qquad \text{(monopole)}, \tag{2.5.46}$$

where $\text{Tr} = 2\,\text{trace}$. Computing this integral for the t'Hooft-Polyakov monopole (by writing it as $\int_{\ 3} \text{Tr}({}^* d^{\mathcal{A}}\phi \wedge d^{\mathcal{A}}\phi) = -\int_{\ 3} \|\, d^{\mathcal{A}}\phi \,\|^2\, dx^1 \wedge dx^2 \wedge dx^3$ and using (2.5.45)) gives a value of 4π. For any configuration $(\boldsymbol{\mathcal{A}}, \phi) \in \mathcal{C}$ satisfying the monopole equations (2.5.14) we define the **monopole number** $N(\boldsymbol{\mathcal{A}}, \phi)$ by

$$N(\boldsymbol{\mathcal{A}}, \phi) = \frac{1}{4\pi}\int_{\ 3} \text{Tr}(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}}\phi). \tag{2.5.47}$$

There are alternative ways of computing $N(\boldsymbol{\mathcal{A}}, \phi)$ (several of which are discussed below) that make it clear that $N(\boldsymbol{\mathcal{A}}, \phi)$ is actually an integer. Indeed, such an integer-valued monopole number can be defined in a much more general context. In **[JT]** it is shown that $\frac{1}{4\pi}\int_{\ 3} \text{Tr}(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}}\phi)$ is well-defined and integer-valued for any $(\boldsymbol{\mathcal{A}}, \phi) \in \mathcal{C}$ that is a critical point for the action A given by (2.5.10). Then **[Groi2]** shows that it is not even necessary to assume $(\boldsymbol{\mathcal{A}}, \phi)$ is a critical point. More precisely, if

$$A(\boldsymbol{\mathcal{A}}, \phi) = \frac{1}{2} \int_{\ ^3} \left(\| \boldsymbol{\mathcal{F}} \|^2 + \| d^{\mathcal{A}} \phi \|^2 \right) dx^1 \wedge dx^2 \wedge dx^3$$

and

$$C = \left\{ (\boldsymbol{\mathcal{A}}, \phi) : A(\boldsymbol{\mathcal{A}}, \phi) < \infty, \lim_{R \longrightarrow \infty} \sup_{|x| \geq R} \big| 1 - \| \phi \| \big| = 0 \right\},$$

then, for any $(\boldsymbol{\mathcal{A}}, \phi) \in \mathcal{C}$,

$$N(\boldsymbol{\mathcal{A}}, \phi) = \frac{1}{4\pi} \int_{\ ^3} \mathrm{Tr} \left(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}} \phi \right)$$

is a well-defined integer called the **monopole number** of $(\boldsymbol{\mathcal{A}}, \phi)$ (even though we do not assume that $(\boldsymbol{\mathcal{A}}, \phi)$ is a solution to the monopole equations).

Remark: The results of **[Groi2]** are actually much more general than this, but these will do for our purposes. Much of what we have to say also extends to the $\lambda > 0$ case which we abandoned on page 133 (see **[JT]** and **[Groi1]**).

We will now describe some alternative ways of computing $N(\boldsymbol{\mathcal{A}}, \phi)$ that shed much light on its topological significance. First we think of $\int_{\ ^3}$ as $\lim_{R \longrightarrow \infty} \int_{|x| \leq R}$ and apply Stokes' Theorem to each $\int_{|x| \leq R}$. For this we first note that

$$\mathrm{Tr} \left(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}} \phi \right) = d \big(\mathrm{Tr}(\phi \boldsymbol{\mathcal{F}}) \big), \tag{2.5.48}$$

where $\phi \boldsymbol{\mathcal{F}}$ is a matrix product. To see this note that $\mathrm{Tr}(d^{\mathcal{A}} \phi) = \mathrm{Tr}(d\phi + [\boldsymbol{\mathcal{A}}, \phi]) = \mathrm{Tr}(d\phi) + \mathrm{Tr}([\boldsymbol{\mathcal{A}}, \phi]) = \mathrm{Tr}(d\phi)$ because any commutator has trace zero ($\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$). Now, using a few simple computational facts that we will establish in Chapter 4,

$$\begin{aligned}
d\,(\mathrm{Tr}(\phi \boldsymbol{\mathcal{F}})) &= \mathrm{Tr}\big(d(\phi \boldsymbol{\mathcal{F}}) \big) = \mathrm{Tr}\big(d^{\mathcal{A}} (\phi \boldsymbol{\mathcal{F}}) \big) \\
&= \mathrm{Tr}(\phi d^{\mathcal{A}} \boldsymbol{\mathcal{F}} + d^{\mathcal{A}} \phi \wedge \boldsymbol{\mathcal{F}}) && \text{(“Product Rule”)} \\
&= \mathrm{Tr}(d^{\mathcal{A}} \phi \wedge \boldsymbol{\mathcal{F}}) && \text{(Bianchi identity)} \\
&= \mathrm{Tr}(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}} \phi),
\end{aligned}$$

where the last equality follows from the fact that $d^{\mathcal{A}} \phi \wedge \boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}} \phi$ differ by a bracket, which has trace zero. This proves (2.5.48) and Stokes' Theorem (Section 4.7) gives

$$\int_{|x| \leq R} \mathrm{Tr} \Big(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}} \phi \Big) = \int_{|x| \leq R} d\,(\mathrm{Tr}(\phi \boldsymbol{\mathcal{F}})) = \int_{|x| = R} \mathrm{Tr}(\phi \boldsymbol{\mathcal{F}}).$$

Writing S_R^2 for the set of points in ^{3} with $|\,x\,| = R$ we obtain

$$N(\boldsymbol{\mathcal{A}}, \phi) = \frac{1}{4\pi}\int_{\ {}^3} \operatorname{Tr}\,(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}}\phi) = \lim_{R \longrightarrow \infty} \frac{1}{4\pi}\int_{|\,x\,| \leq R} \operatorname{Tr}\,(\boldsymbol{\mathcal{F}} \wedge d^{\mathcal{A}}\phi)$$

$$N\,(\boldsymbol{\mathcal{A}}, \phi) = \lim_{R \longrightarrow \infty} \frac{1}{4\pi}\int_{S_R^2} \operatorname{Tr}\,(\phi \boldsymbol{\mathcal{F}}). \tag{2.5.49}$$

Since $\lim_{R \longrightarrow \infty} \sup_{|\,x\,| \geq R} |\,1 - \|\,\phi\,\|\,| = 0$, there exists an $R_0 < \infty$ such that $\|\,\phi(x)\,\| > \frac{1}{2}$ for all x with $|\,x\,| > R_0$. For any such x we define

$$\hat{\phi}\,(\,x\,) = \|\,\phi\,(x)\,\|^{-1}\phi\,(x)$$

and, if $R > R_0$,

$$\hat{\phi}_R = \hat{\phi}|_{S_R^2}.$$

These are smooth maps on their domains and they map into $S^2_{su(2)}$ (the unit 2-sphere in $su(2) \cong {}^3$). One can show (see **[JT]** and **[Groi2]**) that ϕ can be replaced by $\hat{\phi}$ in (2.5.49), i.e.,

$$\begin{aligned} N(\boldsymbol{\mathcal{A}}, \phi) \;&=\; \lim_{R \longrightarrow \infty} \frac{1}{4\pi}\int_{S_R^2} \operatorname{Tr}(\,\hat{\phi}\boldsymbol{\mathcal{F}}\,) \\ &=\; \lim_{R \longrightarrow \infty} \frac{1}{4\pi}\int_{S_R^2} \|\phi\|^{-1}\operatorname{Tr}(\phi\boldsymbol{\mathcal{F}}). \end{aligned} \tag{2.5.50}$$

The reason for preferring the maps $\hat{\phi}$ and $\hat{\phi}_R$ can be seen as follows: Each $\hat{\phi}_R$ can be regarded as a map from S^2 to S^2 and so determines an element $[\hat{\phi}_R]$ of the homotopy group $\pi_2(S^2)$. Since $\hat{\phi}$ is smooth for $R > R_0$, $\hat{\phi}_R$ varies smoothly with $R > R_0$ so this homotopy class is *independent of* $R > R_0$ and we will denote it simply $[\hat{\phi}]$. We claim that $[\hat{\phi}]$ is also *gauge invariant*, i.e., that if $g : {}^3 \longrightarrow SU(2)$ is a gauge transformation and $\phi^g = g^{-1}\phi g$, then, on $|\,x\,| > R_0$, $\|\,\phi^g\,\|^{-1}\phi^g$ is well-defined and homotopic to $\|\,\phi\,\|^{-1}\phi$. It is well-defined because $\|\,\phi^g\,\| = \|\,g^{-1}\phi g\,\| = \|\,\phi\,\|$ which is nonzero on $|\,x\,| > R_0$. On the other hand, since ^{3} is contractible, g is homotopic to the map that sends all of ^{3} the identity e in $SU(2)$ (Exercise 2.3.6, **[N4]**). Thus, on $|\,x\,| > R_0$, $\|\,\phi^g\,\|^{-1}\phi^g = \|\,\phi\,\|^{-1}(g^{-1}\phi g)$ is homotopic to $\|\,\phi\,\|^{-1}(e^{-1}\phi e) = \|\,\phi\,\|^{-1}\phi$ so $[\phi^g] = [\phi]$ as required.

Each map $\hat{\phi}_R$ can be regarded as a map from S^2 to S^2 and therefore has a Brouwer degree $\deg(\hat{\phi}_R)$ (see Section 5.7 or Section 3.4 of **[N4]**). Since the various maps $\hat{\phi}_R, R > R_0$, determine the same homotopy class in $\pi_2(S^2)$, they have the same degree. Remarkably, this degree actually coincides with the monopole number, i.e.,

$$N(\boldsymbol{\mathcal{A}},\phi) = \deg\left(\hat{\phi}_R\right) = \deg\left(\|\,\phi\,\|^{-1}\phi|_{S^2_R}\right) \quad (R > R_0) \tag{2.5.51}$$

(see **[JT]** and **[Groi2]**). Notice that $\deg(\|\,\phi\,\|^{-1}\phi|_{S^2_R})$ depends only on ϕ and, indeed, only on its asymptotic behavior. Even more, it depends only on the "homotopy type of its asymptotic behavior" (if you get my drift). The monopole number distinguishes "homotopy classes" of Higgs fields. These classes are stable in the sense that a continuous perturbation of the field cannot change the class (physicists would say that an infinite potential barrier separates fields with different monopole numbers). Mathematically, there is a natural topology on the configuration space $\mathcal{C}$ with path components labeled by the integers and such that two configurations lie in the same path component if and only if they have the same monopole number (see **[Groi2]**).

We point out that there is an explicit integral formula for calculating the degrees (monopole numbers) in (2.5.51) that is sometimes more manageable than those in earlier formulas:

$$N(\boldsymbol{\mathcal{A}},\phi) = -\frac{1}{4\pi}\int_{S^2_R} \mathrm{Tr}\left(\hat{\phi}\,d\hat{\phi}\wedge d\hat{\phi}\right) \quad (R > R_0). \tag{2.5.52}$$

We'll do this calculation for the t' Hooft-Polyakov monopole. From (2.5.39) and the fact that $\|\,\phi(x)\,\| = \coth r - \frac{1}{r}$ we conclude that, on $S^2_R\,(R > R_0)$,

$$\hat{\phi}(x) = \vec{n}\cdot\vec{T} = \frac{x^a}{R}T_a = -\frac{1}{2R}\begin{pmatrix} x^3\boldsymbol{i} & x^2 + x^1\boldsymbol{i} \\ -x^2 + x^1\boldsymbol{i} & -x^3\boldsymbol{i}\end{pmatrix}.$$

Thus,

$$\begin{aligned} &d\hat{\phi}\wedge d\hat{\phi} \\ &= \frac{1}{4R^2}\begin{pmatrix} dx^3\boldsymbol{i} & dx^2 + dx^1\boldsymbol{i} \\ -dx^2 + dx^1\boldsymbol{i} & -dx^3\boldsymbol{i}\end{pmatrix}\begin{pmatrix} dx^3\boldsymbol{i} & dx^2 + dx^1\boldsymbol{i} \\ -dx^2 + dx^1\boldsymbol{i} & -dx^3\boldsymbol{i}\end{pmatrix} \\ &= -\frac{1}{2R^2}\begin{pmatrix} dx^1\wedge dx^2\boldsymbol{i} & -dx^1\wedge dx^3 + dx^2\wedge dx^3\boldsymbol{i} \\ dx^1\wedge dx^3 + dx^2\wedge dx^3\boldsymbol{i} & -dx^1\wedge dx^2\boldsymbol{i}\end{pmatrix}\end{aligned}$$

and so

$$\begin{aligned}\hat{\phi}\,d\hat{\phi}\wedge d\hat{\phi} =& \frac{1}{4R^3}\begin{pmatrix} x^3\boldsymbol{i} & x^2 + x^1\boldsymbol{i} \\ -x^2 + x^1\boldsymbol{i} & -x^3\boldsymbol{i}\end{pmatrix} \\ &\times\begin{pmatrix} dx^1\wedge dx^2\boldsymbol{i} & -dx^1\wedge dx^3 + dx^2\wedge dx^3\boldsymbol{i} \\ dx^1\wedge dx^3 + dx^2\wedge dx^3\boldsymbol{i} & -dx^1\wedge dx^2\boldsymbol{i}\end{pmatrix}.\end{aligned}$$

The (1,1) entry is

$$\frac{1}{4R^3}\left[\left(-x^3dx^1\wedge dx^2 + x^2dx^1\wedge dx^3 - x^1dx^2\wedge dx^3\right)\right.$$
$$\left. + \left(x^1dx^1\wedge dx^3 + x^2dx^2\wedge dx^3\right)\boldsymbol{i}\right].$$

and the (2,2) entry is

$$\tfrac{1}{4R^3}\left[\left(x^2dx^1\wedge dx^3 - x^1dx^2\wedge dx^3 - x^3dx^1\wedge dx^2\right)\right.$$
$$\left. - \left(x^1dx^1\wedge dx^3 + x^2dx^2\wedge dx^3\right)\boldsymbol{i}\right].$$

Thus,

$$\begin{aligned}\operatorname{Tr}\left(\hat{\phi}d\hat{\phi}\wedge d\hat{\phi}\right) &= 2\operatorname{trace}\left(\hat{\phi}d\hat{\phi}\wedge d\hat{\phi}\right)\\ &= -\frac{1}{R^3}\left(x^1dx^2\wedge dx^3 - x^2dx^1\wedge dx^3 + x^3dx^1\wedge dx^2\right).\end{aligned}$$

We will learn how to integrate such a 2-form over S^2_R in Chapter 4 (indeed, we will find that the restriction of $x^1dx^2\wedge dx^3 - x^2dx^1\wedge dx^3 + x^3dx^1\wedge dx^2$ to S^2 is just the standard volume (i.e., area) form on S^2). Once the machinery is all in hand we will find that one can calculate such things by simply doing what comes natural. In this case, one introduces spherical coordinates

$$\begin{aligned}x^1 &= R\sin\varphi\cos\theta\\ x^2 &= R\sin\varphi\sin\theta\\ x^3 &= R\cos\varphi\end{aligned}$$

with $0\le\theta\le 2\pi$ and $0\le\varphi\le\pi$, computes

$$\begin{aligned}x^1dx^2\wedge dx^3 &= (R\sin\varphi\cos\theta)\,d\,(R\sin\varphi\sin\theta)\wedge d\,(R\cos\varphi)\\ &= R^3\sin\varphi\cos\theta(\cos\varphi\sin\theta d\varphi + \sin\varphi\cos\theta d\theta)\wedge(-\sin\varphi d\varphi)\\ &= -R^3\sin^3\varphi\cos^2\theta d\theta\wedge d\varphi\end{aligned}$$

and similarly for the remaining terms. The result is

$$\begin{aligned}\int_{S^2_R}\operatorname{Tr}\left(\hat{\phi}\,d\hat{\phi}\wedge d\hat{\phi}\right) &= -\int_{(0,\pi)\times(0,2\pi)}\sin\phi\,d\phi\wedge d\theta\\ &= -\int_0^{2\pi}\int_0^{\pi}\sin\phi\,d\phi d\theta\\ &= -4\pi.\end{aligned}$$

Thus, for the t' Hooft-Polyakov monopole,

$$N(\boldsymbol{\mathcal{A}},\phi) = -\frac{1}{4\pi}\int_{S_R^2} \operatorname{Tr}\left(\hat{\phi}\, d\hat{\phi} \wedge d\hat{\phi}\right) = 1.$$

Remark: Changing the sign of ϕ (but leaving $\boldsymbol{\mathcal{A}}$ alone) gives a configuration which is a solution to $\boldsymbol{\mathcal{F}} = {}^*d^{\mathcal{A}}\phi$ (see page 144) and has monopole number -1.

The behavior of the Higgs field on large 2-spheres S_R^2 therefore captures the topological type of the configuration. There is yet another way of seeing this that is reminiscent of our earlier experience with instantons (where the topological type was to be found in the Chern class of a certain bundle). To see this we again fix

$$(\boldsymbol{\mathcal{A}},\phi) \in \mathcal{C} = \left\{(\boldsymbol{\mathcal{A}},\phi) : A(\boldsymbol{\mathcal{A}},\phi) < \infty,\ \lim_{R\longrightarrow\infty}\ \sup_{|x|\geq R}\ |1 - \|\phi\|\,| = 0\right\}$$

and select $R_0 < \infty$ such that $\|\phi(x)\| > \frac{1}{2}$ on $|x| > R_0$. Define

$$\hat{\phi} = \|\phi\|^{-1}\phi : \quad ^3 - \{x : |x| \leq R_0\} \longrightarrow S^2_{su(2)}$$

and, for each $R > R_0$,

$$\hat{\phi}_R = \hat{\phi}\,\Big|_{S_R^2} : S_R^2 \longrightarrow S^2_{su(2)}.$$

Now we fix some $R > R_0$. ϕ is the pullback by the standard cross-section of an equivariant map $\boldsymbol{\Phi} : \ ^3 \times SU(2) \longrightarrow su\ (2)$. The restriction of the trivial $SU(2)$-bundle over ^{3} to S_R^2 is the trivial $SU(2)$-bundle over S_R^2:

$$SU(2) \hookrightarrow S_R^2 \times SU(2) \overset{\mathcal{P}}{\longrightarrow} S_R^2.$$

Let $\boldsymbol{\Phi}_R = \boldsymbol{\Phi}_{S_R^2 \times SU(2)}$ and $\hat{\boldsymbol{\Phi}}_R = \|\boldsymbol{\Phi}_R\|^{-1}\boldsymbol{\Phi}_R$. Both are equivariant and $\hat{\boldsymbol{\Phi}}_R$ takes values in $S^2_{su(2)}$. Furthermore, $\hat{\phi}_R$ is the pullback to S_R^2 by the standard cross-section of $\hat{\boldsymbol{\Phi}}_R$. Thus, $\hat{\phi}_R$ is the standard gauge representation for a Higgs field on the trivial $SU(2)$-bundle over S_R^2 with values in $S^2_{su(2)}$.

Now, select some $\phi_0 \in S^2_{su(2)}$ (a "ground state" for the "virtual potential"; see pages 132–133). The isotropy subgroup of ϕ_0 (with respect to the adjoint action of $SU(2)$ on $su(2)$) is a copy of $U(1)$ in $SU(2)$ (pages 132–133). One can show that $\hat{\boldsymbol{\Phi}}_R^{-1}(\phi_0)$ is a submanifold of $S_R^2 \times SU(2)$ (because $\hat{\boldsymbol{\Phi}}_R$ is a submersion at each point of $\hat{\boldsymbol{\Phi}}_R^{-1}(\phi_0)$) and, furthermore

(i) for each $x \in S_R^2$, $\mathcal{P}^{-1}(x) \cap \hat{\boldsymbol{\Phi}}_R^{-1}(\phi_0) \neq \emptyset$, and

(ii) for $p \in \hat{\mathbf{\Phi}}_R^{-1}(\phi_0)$ and $g \in SU(2)$,

$$p \cdot g \in \hat{\mathbf{\Phi}}_R^{-1}(\phi_0) \quad \text{iff} \quad g \in U(1) \qquad \text{(isotropy subgroup of } \phi_0\text{)}.$$

From these it follows that

$$\mathcal{P}\Big|\hat{\mathbf{\Phi}}_R^{-1}(\phi_0) : \hat{\mathbf{\Phi}}_R^{-1}(\phi_0) \longrightarrow S_R^2$$

is a principal U(l)-bundle over S_R^2 (where the action of $U(1)$ on $\hat{\mathbf{\Phi}}_R^{-1}(\phi_0)$ is just the original $SU(2)$-action on $S_R^2 \times SU(2)$, but with p restricted to $\hat{\mathbf{\Phi}}_R^{-1}(\phi_0)$ and g restricted to $U(1) \subseteq SU(2)$). This $U(1)$-bundle over S_R^2 is called a **reduction** of the structure group of $SU(2) \hookrightarrow S_R^2 \times SU(2) \stackrel{\mathcal{P}}{\longrightarrow} S_R^2$ to $U(1)$. Recall that principal $U(1)$-bundles over spheres are characterized up to equivalence by their 1st Chern number (see pages 63–64 and 68 of Section 2.2). The result of interest to us is the following: *The 1st Chern number of*

$$U(1) \hookrightarrow \hat{\mathbf{\Phi}}_R^{-1}(\phi_0) \xrightarrow{\mathcal{P}\left|\hat{\mathbf{\Phi}}_R^{-1}(\phi_0)\right.} S_R^2$$

is the monopole number $N(\boldsymbol{\mathcal{A}}, \phi)$ of the configuration $(\boldsymbol{\mathcal{A}}, \phi)$.

Since the Chern class can be computed from any connection on the bundle, the proof amounts to finding such a connection that arises naturally from the original Yang-Mills-Higgs potential. We will briefly illustrate how this is done (the following is a special case of Proposition 6.4, Chapter II, of [**KN1**]): We have a connection on $SU(2) \hookrightarrow {}^3 \times SU(2) \longrightarrow {}^3$. Its restriction to $SU(2) \hookrightarrow S_R^2 \times SU(2) \stackrel{\mathcal{P}}{\longrightarrow} S_R^2$ is a connection which we will denote $\boldsymbol{\omega}$. Since $U(1)$ is a subgroup of $SU(2)$, $u(1)$ is a subalgebra of $su(2)$. Now, we have an ad-invariant, positive definite inner product

$$\langle A, B \rangle = -2\mathrm{trace}(AB)$$

on $su(2)$. Let $u(1)^\perp$ be the $\langle\ ,\ \rangle$-orthogonal complement of $u(1)$ in $su(2)$. Then $su(2) = u(1) \oplus u(1)^\perp$ and $u(1)^\perp$ is also ad-invariant (we let $\mu : SU(2) \longrightarrow GL(u(1)^\perp)$ be the induced representation). If $\iota : \hat{\mathbf{\Phi}}_R^{-1}(\phi_0) \hookrightarrow S_R^2 \times SU(2)$ is the inclusion map, then $\iota^*\boldsymbol{\omega}$ is an $su(2)$-valued 1-form on $\hat{\mathbf{\Phi}}_R^{-1}(\phi_0)$ and therefore splits

$$\iota^*\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \boldsymbol{\gamma},$$

where $\boldsymbol{\omega}_0$ is $u(1)$-valued and $\boldsymbol{\gamma}$ is $u(1)^\perp$-valued. It's easy to see that $\boldsymbol{\omega}_0$ is a connection 1-form on

$$U(1) \hookrightarrow \hat{\mathbf{\Phi}}_R^{-1}(\phi_0) \xrightarrow{\mathcal{P}\left|\hat{\mathbf{\Phi}}_R^{-1}(\phi_0)\right.} S_R^2$$

(and γ is "tensorial of type μ"). Computing the 1st Chern number of this bundle from ω_0 gives the expression (2.5.52) for $N(\mathcal{A}, \phi)$.

2.6 Epilogue

We hope by now to have satisfied the curiosity of those who may have wondered how such apparently abstruse mathematical notions as spinor structures and characteristic classes might arise in the study of the world around us. We will have one more serious encounter with this in the Appendix, but it is time now to put aside the informal, heuristic, discussions that have characterized this chapter and deal honestly with these notions for their own sake. The remainder of the book is intended to do just that. Certainly, one need not demand any physical motivation to study and appreciate the rather beautiful mathematics to follow. Nevertheless, it is pro-foundly satisfying that the physical motivation exists and, as the concepts are made precise and the theorems are rigorously proved, we recommend a periodic dip in the murkier waters of physics (the journal *Communications in Mathematical Physics* is fine for browsing). It lends perspective.

3
Frame Bundles and Spacetimes

3.1 Partitions of Unity, Riemannian Metrics and Connections

An n-dimensional smooth manifold X is locally diffeomorphic to ^{n} and the study of such manifolds generally necessitates piecing together well-understood local information about ^{n} into global information about X. The basic tool for this piecing together operation is a partition of unity. In this section we will prove that these exist in abundance on any manifold and use them to establish the global existence of two useful objects that clearly always exist locally (Riemannian metrics and connection forms). The same technique will be used in Section 4.3 to obtain a convenient reformulation of the notion of orientability while, in Section 4.6, a theory of integration is constructed on any orientable manifold by piecing together Lebesgue integrals on coordinate neighborhoods.

We begin with a purely topological lemma. A family $\{A_\alpha : \alpha \in \mathcal{A}\}$ of subsets of a topological space X is said to be **locally finite** if each $x \in X$ has a neighborhood U such that $U \cap A_\alpha \neq \emptyset$ for at most finitely many $\alpha \in \mathcal{A}$.

Exercise 3.1.1 Show that, if $\{A_\alpha : \alpha \in \mathcal{A}\}$ is locally finite and $\bar{A}_\alpha$ denotes the closure of A_α in X, then $\{\bar{A}_\alpha : \alpha \in \mathcal{A}\}$ is also locally finite.

Recall (Section 1.4, [**N4**]) that a subset U of X is relatively compact if its closure $\bar{U}$ is compact and that X itself is said to be locally compact if it is Hausdorff and each point in X has a relatively compact open neighborhood. If X is also second countable (i.e., has a countable basis for its topology), then it has a countable basis of relatively compact open sets (Lemma 1.4.9, [**N4**]). Manifolds are, in particular, locally compact, second countable topological spaces. Finally, recall (Section 4.3, [**N4**]) that, if $\mathcal{U}$ and $\mathcal{V}$ are two covers of X, then $\mathcal{V}$ is said to be a **refinement** of $\mathcal{U}$ if every $V \in \mathcal{V}$ is contained in some $U \in \mathcal{U}$.

Lemma 3.1.1 *Let X be a locally compact, second countable topological space. Then any open cover of X has a countable, locally finite refinement consisting of relatively compact open sets.*

Proof: Let $\{B_1, B_2, \ldots\}$ be a basis for the open sets of X with each $\bar{B}_i$ compact. We begin by constructing an auxiliary countable cover $\{W_1, W_2, \ldots\}$ of X consisting of relatively compact open sets and with the property that $\bar{W}_i \subseteq W_{i+1}$ for each $i = 1, 2, \ldots$. Set $W_1 = B_1$ and assume (for the inductive construction) that $W_1, \ldots, W_n$ have been defined so that W_i is open

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0_3,

and relatively compact for $i \leq n$ and $\bar{W}_i \subseteq W_{i+1}$ for $i \leq n-1$. Since $\bar{W}_n$ is compact it is contained in a finite union of the B_i. Let i_n be the least positive integer greater than or equal to n such that $\bar{W}_n \subseteq B_1 \cup \cdots \cup B_{i_n}$. Set $W_{n+1} = B_1 \cup \cdots \cup B_{i_n}$. Then W_{n+1} is open, $\bar{W}_n \subseteq W_{n+1}$ and $\bar{W}_{n+1} = \overline{B_1 \cup \cdots \cup B_{i_n}} = \bar{B}_1 \cup \cdots \cup \bar{B}_{i_n}$ is compact so the induction is complete and we have a sequence $\{W_1, W_2, \ldots\}$ of relatively compact open sets and with $\bar{W}_i \subseteq W_{i+1}$ for each $i = 1, 2, \ldots$. Since W_i contains the union of the first i elements of $\{B_1, B_2, \ldots\}$, $\bigcup_{i=1}^{\infty} W_i = X$ and $\{W_1, W_2, \ldots\}$ is a cover of X.

Exercise 3.1.2 Show that each of the sets $\bar{W}_2,\ \bar{W}_3 - W_2,\ \bar{W}_4 - W_3, \ldots,$ $\bar{W}_i - W_{i-1}, \ldots$ is compact, each of the sets $W_3,\ W_4 - \bar{W}_1,\ W_5 - \bar{W}_2, \ldots,$ $W_{i+1} - \bar{W}_{i-2}, \ldots$ is open, and

$$\begin{gathered}
\bar{W}_2 \subseteq W_3 \\
\bar{W}_3 - W_2 \subseteq W_4 - \bar{W}_1 \\
\bar{W}_4 - W_3 \subseteq W_5 - \bar{W}_2 \\
\vdots \\
\bar{W}_i - W_{i-1} \subseteq W_{i+1} - \bar{W}_{i-2} \\
\vdots
\end{gathered}$$

With this auxiliary cover in hand we can now prove the lemma. Let $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ be an arbitrary open cover of X. Define $\mathcal{U}_2 = \{U_\alpha \cap W_3 : \alpha \in \mathcal{A}\}$ and, for $i \geq 3$, $\mathcal{U}_i = \{U_\alpha \cap (W_{i+1} - \bar{W}_{i-2}) : \alpha \in \mathcal{A}\}$. Then $\mathcal{U}_2$ covers $\bar{W}_2$ and, for $i \geq 3$, $\mathcal{U}_i$ covers $\bar{W}_i - W_{i-1}$. Let $\mathcal{V}_2$ be a finite collection of elements of $\mathcal{U}_2$ that cover $\bar{W}_2$ and, for $i \geq 3$, let $\mathcal{V}_i$ be a finite collection of elements of $\mathcal{U}_i$ that cover $\bar{W}_i - W_{i-1}$. Finally, let $\mathcal{V} = \bigcup_{i=2}^{\infty} \mathcal{V}_i$. Then $\mathcal{V}$ is a countable family of open sets in X. Each element of $\mathcal{V}$ is contained in some U_α as well as in some $\bar{W}_j$ so that $\mathcal{V}$ refines $\mathcal{U}$ and the elements of $\mathcal{V}$ have compact closure. To see that $\mathcal{V}$ is a cover, let x be an arbitrary element of X. Select the least integer $i \geq 1$ such that $x \in W_i$. If $i = 1$, or 2, then $x \in \bar{W}_2$ and therefore in some element of $\mathcal{V}_2$. If $i \geq 3$, then $x \in \bar{W}_i - W_{i-1}$ so x is in some element of $\mathcal{V}_i$. This also implies that $\mathcal{V}$ is locally finite since any $x \in X$ is in one of the compact sets $\bar{W}_2,\ \bar{W}_3 - W_2, \ldots$ and the corresponding open set $W_3,\ W_4 - \bar{W}_1, \ldots$ is an open neighborhood of x which, by construction, intersects only finitely many elements of $\mathcal{V}$. ■

For any real-valued function f on a topological space X we define the **support** of f, denoted $\operatorname{supp} f$, to be the closure in X of the set of points on which f is nonzero, i.e.,

$$\operatorname{supp} f = \overline{\{x \in X : f(x) \neq 0\}}.$$

Thus, f is identically zero on $X - \operatorname{supp} f$, but may be zero at some points of $\operatorname{supp} f$ as well. Now we return to the manifold setting. A (**smooth**) **partition**

of unity on the n-dimensional smooth manifold X is a collection $\{\phi_\gamma : \gamma \in \mathcal{C}\}$ of smooth, non-negative real-valued functions on X such that

1. Each $\operatorname{supp} \phi_\gamma$ is compact.
2. $\{\operatorname{supp} \phi_\gamma : \gamma \in \mathcal{C}\}$ is locally finite.
3. For each $x \in X, \sum_{\gamma \in \mathcal{C}} \phi_\gamma(x) = 1$.

Remark: Notice that, by #2, each $x \in X$ has a neighborhood on which at most finitely many of the ϕ_γ are nonzero so the sum in #3 is a finite sum. The fact that the sum is 1 implies that the sets $\operatorname{supp} \phi_\gamma, \gamma \in \mathcal{C}$, cover X. Moreover, since the ϕ_γ are assumed non-negative, we must have $0 \leq \phi_\gamma(x) \leq 1$ for each $x \in X$.

If $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ is an open cover of X, then the partition of unity $\{\phi_\gamma : \gamma \in \mathcal{C}\}$ is said to be **subordinate** to $\mathcal{U}$ if each $\operatorname{supp} \phi_\gamma$ is contained in some U_α. We wish to prove that any open cover of any smooth manifold has a (countable) partition of unity subordinate to it. For this we will need to use the fact that manifolds have rich supplies of smooth real-valued functions. More precisely, we will appeal to the following two results, both proved in [**N4**].

Lemma 3.1.2 *Let X be a smooth manifold, U a coordinate neighborhood in X and A_0 and A_1 disjoint closed subsets of U. Then there exists a smooth function $f : U \longrightarrow$ satisfying $0 \leq f(x) \leq 1$ for all $x \in U$, $A_0 = f^{-1}(0)$ and $A_1 = f^{-1}(1)$.*

This is Exercise 5.4.4 of [**N4**]. The following is Lemma 5.4.2 of [**N4**].

Lemma 3.1.3 *Let W be an open subset of the smooth manifold X and p a point in W. Then there exists a non-negative, C^∞ function g on X that is 1 on an open neighborhood of p in W and 0 outside W (g is called a* **bump function** *at p in W).*

With these we can prove our major result.

Theorem 3.1.4 *Let X be a smooth manifold and $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ an open cover of X. Then there exists a countable partition of unity $\{\phi_k\}_{k=1,2,\ldots}$ subordinate to $\mathcal{U}$.*

Proof: It will clearly suffice to find such a partition of unity subordinate to any refinement of $\mathcal{U}$ so, by intersecting with the coordinate neighborhoods in some atlas for X, we might as well assume at the outset that each U_α is a coordinate neighborhood. For each $x \in X$ we can select an open neighborhood of x whose closure is contained in some U_α. Let $\mathcal{U}'$ be the open cover of X consisting of all of these neighborhoods. By Lemma 3.1.1, there is a countable, locally finite, open refinement $\mathcal{V} = \{V_k\}_{k=1,2,\ldots}$ of $\mathcal{U}'$ with each $\bar{V}_k$ compact. Thus, $\mathcal{V}$ is also a refinement of $\mathcal{U}$ and, moreover, each $\bar{V}_k$ is contained in some element of $\mathcal{U}$. By Exercise 3.1.1, $\{\bar{V}_k\}_{k=1,2\ldots}$ is also locally finite so, to complete the proof, it will suffice to find a partition of unity $\{\phi_k\}_{k=1,2\ldots}$ with $\operatorname{supp}\phi_k = \bar{V}_k$ for each $k = 1, 2, \ldots$.

For each $k = 1, 2, \ldots$ select $\alpha(k) \in \mathcal{A}$ such that $\bar{V}_k \subseteq U_{\alpha(k)}$. Then $U_{\alpha(k)} - V_k$ is a nonempty, proper closed subset of $U_{\alpha(k)}$ so, by Lemma 3.1.2, there is a non-negative smooth function f'_k on $U_{\alpha(k)}$ with $(f'_k)^{-1}(0) = U_{\alpha(k)} - V_k$. From this we construct a non-negative smooth function f_k on X with $f_k^{-1}(0) = X - V_k$ as follows: For each $p \in \bar{V}_k$ select a non-negative bump function g_p which is 1 on a neighborhood U_p of p in $U_{\alpha(k)}$ and 0 on $X - U_{\alpha(k)}$. Cover $\bar{V}_k$ by finitely many of these neighborhoods $U_{p_1}, \ldots, U_{p_j}$ and let $g = g_{p_1} + \cdots + g_{p_j}$. Then g is C^∞ on X, nonzero on $\bar{V}_k$ and 0 on $X - U_{\alpha(k)}$. Define f_k on X by

$$f_k(x) = \begin{cases} f'_k(x)g(x), & x \in U_{\alpha(k)} \\ 0, & x \in X - U_{\alpha(k)} \end{cases} .$$

Exercise 3.1.3 Show that f_k is a well-defined smooth map on X with $f_k^{-1}(0) = X - V_k$.

From Exercise 3.1.3 we conclude that $\operatorname{supp} f_k = \bar{V}_k$. Thus, $\{\operatorname{supp} f_k\}_{k=1,2\ldots}$ is locally finite so we can define a smooth real-valued function f on X by

$$f(x) = \sum_{k=1}^{\infty} f_k(x)$$

for each $x \in X$. Moreover, since every $x \in X$ is in some V_k, $f(x) > 0$ for each $x \in X$. Thus, for each $k = 1, 2, \ldots$, the function $\phi_k : X \longrightarrow$ defined by

$$\phi_k(x) = \frac{f_k(x)}{f(x)}$$

is non-negative, C^∞, has $\operatorname{supp}\phi_k = \operatorname{supp} f_k = \bar{V}_k$ and satisfies $\sum_{k=1}^{\infty} \phi_k(x) = 1$ for each $x \in X$ as required. ∎

The functions in a partition of unity are required to have compact support. On occasion it is convenient to drop this requirement.

Corollary 3.1.5 *Let X be a smooth manifold and $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ an open cover of X. Then there exists a family $\{\phi_\alpha : \alpha \in \mathcal{A}\}$ of smooth, non-negative real-valued functions on X such that* $\operatorname{supp}\phi_\alpha \subseteq U_\alpha$ *for each $\alpha \in \mathcal{A}$, $\{\operatorname{supp}\phi_\alpha : \alpha \in \mathcal{A}\}$ it is locally finite and, for each $x \in X$, $\sum_{k=1}^{\infty}$ $\phi_k(x) = 1$.*

Proof: Choose a countable partition of unity $\{\phi'_k\}_{k=1,2,\ldots}$ on X subordinate to $\mathcal{U}$. Define ϕ_α to be identically zero if no ϕ'_k has support in U_α and, otherwise, let ϕ_α be the sum of the ϕ'_k with support in U_α.

Exercise 3.1.4 Show that $\{\phi_\alpha : \alpha \in \mathcal{A}\}$ has the required properties. ∎

Exercise 3.1.5 Let X be a smooth manifold, U an open subset of X and A a closed subset of X with $A \subseteq U$. Show that there exists a smooth, real-valued function ϕ on X satisfying $0 \le \phi(x) \le 1$ for each $x \in X$, $\phi(x) = 1$ for each $x \in A$ and $\phi(x) = 0$ for each $x \in X - U$. **Hint**: Consider the open cover $\{U,\ X - A\}$ of X and use Corollary 3.1.5.

We conclude this section with two important applications of Theorem 3.1.4.

Theorem 3.1.6 *Any smooth manifold X admits a Riemannian metric.*

Proof: Observe first that if (U, φ) is a chart on X, then the open submanifold U of X admits a Riemannian metric (e.g., $\varphi^* \bar{\boldsymbol{g}}$, where $\bar{\boldsymbol{g}}$ is the standard metric on $\varphi(U) \subseteq \ ^n$). Now, let $\{(U_\alpha, \varphi_\alpha) : \alpha \in \mathcal{A}\}$ be an atlas for X. By Theorem 3.1.4 there is a countable partition of unity $\{\phi_k\}_{k=1,2,\ldots}$ on X subordinate to $\{U_\alpha : \alpha \in \mathcal{A}\}$. For each $k = 1, 2, \ldots$ choose $\alpha(k) \in \mathcal{A}$ such that $\operatorname{supp} \phi_k \subseteq U_{\alpha(k)}$. Then $\{U_{\alpha(k)}\}_{k=1,2,\ldots}$ is a countable subcover of $\{U_\alpha : \alpha \in \mathcal{A}\}$. On each $U_{\alpha(k)}$ select a Riemannian metric $\boldsymbol{g}_k$ and define $\boldsymbol{g}$ by

$$\boldsymbol{g} = \sum_{k=1}^{\infty} \phi_k \boldsymbol{g}_k.$$

More precisely, we define $\boldsymbol{g}$ as follows: For each $p \in X$ and all $\boldsymbol{v},\ \boldsymbol{w} \in T_p(X)$ we let

$$\boldsymbol{g}(p)(\boldsymbol{v}, \boldsymbol{w}) = \sum_{k=1}^{\infty} \phi_k(p) \boldsymbol{g}_k(p)(\boldsymbol{v}, \boldsymbol{w}), \tag{3.1.1}$$

where it is understood that, if $p \notin \operatorname{supp} \phi_k$, then the term $\phi_k(p)\boldsymbol{g}_k(p)(\boldsymbol{v}, \boldsymbol{w})$ is taken to be zero ($\phi_k(p)$ is zero, of course, but $\boldsymbol{g}_k(p)$ will not be defined if $p \notin U_{\alpha(k)}$). Thus, the sum is finite.

Each $\boldsymbol{g}(p)$ defined by (3.1.1) is clearly symmetric and bilinear. Moreover, $\boldsymbol{g}(p)(\boldsymbol{v}, \boldsymbol{v}) \ge 0$ and, if $\boldsymbol{v} \ne \boldsymbol{0}$, $\boldsymbol{g}(p)(\boldsymbol{v}, \boldsymbol{v}) > 0$ since no term in the sum (3.1.1) is negative and at least one must be positive if $\boldsymbol{w} = \boldsymbol{v}$. Thus, each $\boldsymbol{g}(p)$ is a positive definite inner product on $T_p(X)$. All that remains is to prove smoothness and this we may do locally. Fix a $p \in X$. Then p has an open neighborhood that intersects only finitely many of the sets $\operatorname{supp} \phi_k$. Intersect this neighborhood with all of the corresponding $U_{\alpha(k)}$ containing p to obtain a coordinate neighborhood U of p. Restrict to U the coordinates $x^1, \ldots, x^n$

of any one of these coordinate neighborhoods $U_{\alpha(k)}$. Then, on U, the sum $\boldsymbol{g} = \sum_{k=1}^{\infty} \phi_k \boldsymbol{g}_k$ defining $\boldsymbol{g}$ is finite, each ϕ_k is a C^∞ function of $x^1, \ldots, x^n$ and each $\boldsymbol{g}_k(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$ is C^∞ so $\boldsymbol{g}(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$ is C^∞ for $i, j = 1, \ldots, n$. Thus, $\boldsymbol{g}$ is smooth on U as required. ■

An analogous argument establishes the existence of a connection form on any smooth principal bundle.

Theorem 3.1.7 *Any smooth principal bundle* $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ *admits a connection.*

Proof: First note that if (U, φ) is a chart for which U is a trivializing neighborhood, then the trivial bundle $G \hookrightarrow \mathcal{P}^{-1}(U) \xrightarrow{\mathcal{P}} U$ admits a connection. Indeed, if $\Psi : \mathcal{P}^{-1}(U) \longrightarrow U \times G$ is a trivialization, $\pi : U \times G \longrightarrow G$ is the projection and $\boldsymbol{\Theta}$ is the Cartan 1-form for G, then $\pi^*\boldsymbol{\Theta}$ is a (flat) connection form on $U \times G$ (Exercise 6.2.12, [**N4**]) so $\Psi^*(\pi^*\boldsymbol{\Theta}) = (\pi \circ \Psi)^*\boldsymbol{\Theta}$ is a connection form on $\mathcal{P}^{-1}(U)$ (Theorem 6.1.3, [**N4**]).

Now let $\{(U_\alpha, \varphi_\alpha) : \alpha \in \mathcal{A}\}$ be an atlas for X with each U_α a trivializing neighborhood for $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$. Let $\{\phi_k\}_{k=1,2,\ldots}$ be a countable partition of unity subordinate to $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$. For each k, choose $\alpha(k) \in \mathcal{A}$ such that $\operatorname{supp} \phi_k \subseteq U_{\alpha(k)}$. Then $\{U_{\alpha(k)}\}_{k=1,2,\ldots}$ is a countable subcover of $\mathcal{U}$. For each k, let $\boldsymbol{\omega}_k$ be a connection form on $\mathcal{P}^{-1}(U_k)$ and define

$$\boldsymbol{\omega} = \sum_{k=1}^{\infty} (\phi_k \circ \mathcal{P})\, \boldsymbol{\omega}_k.$$

More precisely, for each $p \in P$ and all $\boldsymbol{v} \in T_p(P)$ we set

$$\boldsymbol{\omega}(p)(\boldsymbol{v}) = \sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p)) \boldsymbol{\omega}_k(p)(\boldsymbol{v}), \tag{3.1.2}$$

where it is understood that, if $\mathcal{P}(p) \notin \operatorname{supp} \phi_k$, then the term $\phi_k(\mathcal{P}(p))\boldsymbol{\omega}_k(p)(\boldsymbol{v})$ is taken to be zero. Note that (3.1.2) does, indeed, define a $\mathcal{G}$-valued 1-form on P since, at each p, the sum is finite and $\mathcal{G}$ is a real vector space.

Exercise 3.1.6 Show that $\boldsymbol{\omega}$ is smooth.

We show that $\boldsymbol{\omega}$ is a connection form by proving

$$\sigma_g^* \boldsymbol{\omega} = ad_{g^{-1}} \circ \boldsymbol{\omega} \tag{3.1.3}$$

for all $g \in G$ and

$$\boldsymbol{\omega}(A^{\#}) = A \tag{3.1.4}$$

for all $A \in G$. (3.1.4) is easy because

$$\begin{aligned}\boldsymbol{\omega}(p)(A^{\#}(p)) &= \sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\boldsymbol{\omega}_k(p)(A^{\#}(p)) \\ &= \sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\, A = \left(\sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\right) A \\ &= 1A = A.\end{aligned}$$

(3.1.3) follows from the linearity of $ad_{g^{-1}}$. Indeed, for $g \in G$, $p \in P$ and $\boldsymbol{v} \in T_{p\cdot g^{-1}}(P)$,

$$\begin{aligned}\left(\sigma_g^*\boldsymbol{\omega}\right)(p\cdot g^{-1})(\boldsymbol{v}) &= \boldsymbol{\omega}(p)\Big((\sigma_g)_{*p\cdot g^{-1}}(\boldsymbol{v})\Big) \\ &= \sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\,\boldsymbol{\omega}_k(p)\Big((\sigma_g)_{*p\cdot g^{-1}}(\boldsymbol{v})\Big) \\ &= \sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\left(\sigma_g^*\boldsymbol{\omega}_k\right)_{p\cdot g^{-1}}(\boldsymbol{v}) \\ &= \sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\, ad_{g^{-1}}\Big(\boldsymbol{\omega}_k(p\cdot g^{-1})(\boldsymbol{v})\Big) \\ &= ad_{g^{-1}}\left(\sum_{k=1}^{\infty} \phi_k(\mathcal{P}(p))\,\boldsymbol{\omega}_k\,(p\cdot g^{-1})(\boldsymbol{v})\right) \\ &= ad_{g^{-1}}\Big(\boldsymbol{\omega}\,(p\cdot g^{-1})(\boldsymbol{v})\Big)\end{aligned}$$

as required. ■

3.2 Continuous Versus Smooth

We begin by using the material in Section 3.1 to prove that continuous maps between open subsets of Euclidean spaces can be arbitrarily well approximated by smooth maps. Indeed, we prove a bit more.

Theorem 3.2.1 *Let U be an open subset of ^{n}, V an open subset of ^{m} and $h : U \longrightarrow V$ a continuous map. Suppose U_0 is an open subset of U, $A \subseteq U_0$ is closed in U and $h\,|\,U_0$ is smooth. Then, for any continuous function $\epsilon : U \longrightarrow (0,\infty)$, there exists a smooth map $f : U \longrightarrow V$ such that*

1. *$\|f(x) - h(x)\| < \epsilon(x)$ for every $x \in U$.*
2. *$f(x) = h(x)$ for all $x \in A$.*

Proof: We ask the reader to take the first step.

Exercise 3.2.1 Show that it is enough to prove that, for every continuous $\epsilon : U \longrightarrow (0, \infty)$, there is a smooth map $f : U \longrightarrow \quad ^m$ satisfying (1) and (2). **Hint**: Assume this has been proved and consider

$$\epsilon'(x) = \min\left\{\epsilon(x),\ \frac{1}{2}\operatorname{dist}(h(x), \quad ^m - V)\right\},$$

where

$$\operatorname{dist}(h(x), \quad ^m - V) = \inf\{\|h(x) - y\| : y \in \quad ^m - V\}.$$

Now, h and ϵ are continuous on U and $U - A$ is open. For each fixed $p \in U - A$, the function $\|h(p) - h(x)\|$ is continuous and takes the value 0 at p. Since $\epsilon(p) > 0$ there is an open neighborhood U_p of p in $U - A$ on which $\|h(p) - h(x)\| < \epsilon(x)$. Consider the open cover of U consisting of U_0 together with all of the U_p, $p \in U - A$. Corollary 3.1.5 gives a family $\{\phi_0\} \cup \{\phi_p : p \in U - A\}$ of smooth, non-negative functions on U such that $\operatorname{supp}\phi_0 \subseteq U_0$, $\operatorname{supp}\phi_p \subseteq U_p$ for each $p \in U - A$, $\{\operatorname{supp}\phi_0\} \cup \{\operatorname{supp}\phi_p : p \in U - A\}$ is locally finite and, for each $x \in U$, $\phi_0(x) + \sum_{p \in U-A}\phi_p(x) = 1$. Since h is smooth on U_0 and $\operatorname{supp}\phi_0 \subseteq U_0$ we can define a smooth function $f : U \longrightarrow \quad ^m$ by

$$f(x) = \phi_0(x)\, h(x) + \sum_{p \in U-A} \phi_p(x)\, h(p).$$

Notice that

$$\begin{aligned} h(x) &= \left(\phi_0(x) + \sum_{p \in U-A} \phi_p(x)\right) h(x) \\ &= \phi_0(x)\, h(x) + \sum_{p \in U-A} \phi_p(x)\, h(x) \end{aligned}$$

so

$$f(x) - h(x) = \sum_{p \in U-A} \phi_p(x)\, (h(p) - h(x)).$$

Exercise 3.2.2 Show that f satisfies (1) and (2). ∎

Corollary 3.2.2 *Let U be an open subset of $\quad ^n$, V an open subset of $\quad ^m$ and $h : U \longrightarrow V$ a continuous map. Then h is homotopic to a smooth map $f : U \longrightarrow V$.*

Proof: For each $x \in U$ let $\epsilon(x) = \frac{1}{2}\operatorname{dist}(h(x), \quad ^m - V)$ (see Exercise 3.2.1). Then the open ball $U_{\epsilon(x)}(h(x))$ of radius $\epsilon(x)$ about $h(x)$ is contained in V for each $x \in U$. By Theorem 3.2.1, we can select a smooth map $f : U \longrightarrow V$ with $\|f(x) - h(x)\| < \epsilon(x)$ for each $x \in U$. In particular, since $f(x) \in U_{\epsilon(x)}(h(x))$ and this ball is convex, $(1 - t)h(x) + t\, f(x)$ is in $U_{\epsilon(x)}(h(x))$ and therefore in V for each t in $[0, 1]$. Thus,

$$F(x,\ t) = (1-t)h(x) + t\,f(x), \quad x \in U, \quad t \in [0,1],$$

defines a homotopy from h to f. ■

Corollary 3.2.3 *Let U be an open subset of ^{n}, V an open subset of ^{m} and $f_0,\ f_1 : U \longrightarrow V$ two smooth maps that are homotopic. Then there exists a smooth map $F : U \times \quad \longrightarrow V$ with $F(x,0) = f_0(x)$ and $F(x,1) = f_1(x)$ for each $x \in U$ (F is called a* **smooth homotopy** *from f_0 to f_1).*

Proof: By assumption, there exists a continuous map $G : U \times [0,1] \longrightarrow V$ with $G(x,0) = f_0(x)$ and $G(x,1) = f_1(x)$ for all $x \in U$. By Exercise 3.1.5, there exists a smooth function $\phi : \quad \longrightarrow \quad$ with $0 \leq \phi(t) \leq 1$ for each $t \in \ $, $\phi(t) = 0$ for each $t \in (-\infty, \frac{1}{3}]$ and $\phi(t) = 1$ for each $t \in [\frac{2}{3}, \infty)$. Now define $H : U \times \quad \longrightarrow V$ by

$$H(x,\ t) = G(x,\ \phi(t))$$

for each $(x,t) \in U \times \ $. Note that H is continuous and satisfies $H(x,t) = f_0(x)$ for all $t \leq \frac{1}{3}$ and $H(x,t) = f_1(x)$ for all $t \geq \frac{2}{3}$ and for all $x \in U$. In particular, H is smooth on the open set $U_0 = U \times (-\infty, \frac{1}{3}) \cup U \times (\frac{2}{3}, \infty)$. Furthermore, $A = U \times \{0,1\}$ is a closed subset of the subspace U with $A \subseteq U_0$. According to Theorem 3.2.1, there exists a smooth map $F : U \times \quad \longrightarrow V$ that agrees with H on $U \times \{0,1\}$ and this clearly has the required properties. ■

These last few results have interesting applications, a few of which we will see in Chapter 5.

Remarks: Each of the results we have proved in this section for open submanifolds of Euclidean space can be generalized to arbitrary smooth manifolds, but the proofs often require substantial machinery. For example, in proving that any continuous map $h : X \longrightarrow Y$ between smooth manifolds is homotopic to a smooth map $f : X \longrightarrow Y$, the family of balls $U_{\epsilon(x)}(h(x))$ covering the image of h which appeared in the proof of Corollary 3.2.2 must be replaced by a "tubular neighborhood" and the existence of these is nontrivial. A good reference for this and many analogous approximation theorems is [**Hir**]. Another important result of this same sort concerns cross-sections of principal bundles. Suppose $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is a smooth principal bundle and suppose there exists a continuous map $s_c : X \longrightarrow P$ such that $\mathcal{P} \circ s_c = id_X$ (i.e., a "continuous cross-section" of the bundle). Then one can show (Theorem, Section 6.7, Part I, of [**St**]) that there must exist a smooth cross-section $s : X \longrightarrow P$ ("arbitrarily close" to s_c and, if s_c is smooth on an open set $U \subseteq X$ and A is closed in U, agreeing with s_c on A). The proof requires only tools that we now have at our disposal, but the argument given in [**St**] is so thorough that we will simply refer the reader to this classic exposition for details. An immediate consequence is that, if a smooth principal bundle is trivial as a C^0-bundle, then it is trivial as a C^∞-bundle. Here is another: Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a C^0-principal bundle in which it just so happens that G

is a Lie group, X and P are smooth manifolds and $\mathcal{P}$ as well as the action of G on P are smooth. Then $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is a smooth principal bundle (i.e., has smooth local trivializations). One need only apply the result from [**St**] to each trivialization.

We conclude with two more items of this same sort. Given a smooth manifold X, a Lie group G, an open cover $\{V_j\}_{j \in J}$ of X and a family $\{g_{ji}\}$ of smooth maps from nonempty intersections $V_i \cap V_j$ into G satisfying the cocycle condition, it is a simple exercise to trace through the proof of the (C^0) Reconstruction Theorem (Theorem 4.3.4, [**N4**]) and show that the resulting bundle admits a natural smooth structure (if such an exercise does not appeal to you, see Proposition 5.2, Chapter I, of [**KN1**]). One thereby obtains the version of the Reconstruction Theorem recorded on page 34. Similarly, the proof of the (C^0) Classification Theorem for principal G-bundles over S^n (Theorem 4.4.3, [**N4**]) shows that the equivalence class of such a bundle is uniquely determined by the homotopy type of its characteristic map $T = g_{12} \,|\, S^{n-1}$ from the equator $S^{n-1} \subseteq S^n$ into G. If G is a Lie group one can choose a smooth map homotopic to T and use this as a characteristic map to build a G-bundle over S^n (Lemma 4.4.1, [**N4**]). This latter bundle is smooth (by the smooth Reconstruction Theorem) and equivalent to the original bundle so every equivalence class contains a smooth bundle. In this way one arrives at the smooth version of the Classification Theorem, as stated on page 34.

3.3 Frame Bundles

We begin by constructing, for any smooth manifold X, a principal bundle whose fiber above any $x \in X$ consists of all the ordered bases for the tangent space $T_x(X)$ at x. The group is $GL(n, \;\;)$ and the action is the natural one that carries one basis onto another. Additional structures on X (e.g., a Riemannian metric) distinguish certain bases for the tangent spaces (e.g., orthonormal) and "reduce" the structure group $GL(n, \;\;)$ to some subgroup (e.g., $O(n)$).

If X is an n-dimensional smooth manifold, then a **frame** at $x \in X$ is an ordered basis $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ for $T_x(X)$. Any such frame p gives rise to a linear isomorphism

$$\underset{\sim}{p} : \;\; {}^n \longrightarrow T_x(X)$$

defined by

$$\underset{\sim}{p}(e_i) = \boldsymbol{b}_i, \qquad i = 1, \ldots, n\,,$$

where $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, 0, \ldots, 1)$ is the standard basis for $\;\;{}^n$. Conversely, every isomorphism $\underset{\sim}{p} : \;\; {}^n \longrightarrow T_x(X)$ gives rise to a frame $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) = (\underset{\sim}{p}(e_1), \ldots, \underset{\sim}{p}(e_n))$. It is often convenient to identify a frame p with its corresponding isomorphism $\underset{\sim}{p}$. We denote by $L(X)_x$ the set of all frames at $x \in X$ and let $L(X) = \bigcup_{x \in X} L(X)_x$. For each $p \in L(X)_x \subseteq L(X)$

we define $\mathcal{P}_L(p) = x$ and thereby obtain a surjective map $\mathcal{P}_L : L(X) \longrightarrow X$. Next we define a right action $\sigma : L(X) \times GL(n,\ \ ) \longrightarrow L(X)$ of $GL(n,\ \ )$ on $L(X)$ as follows: For each $(p, g) \in L(X) \times GL(n,\ \ )$, with $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \in L(X)_x \subseteq L(X)$ and $g = (g^i{}_j) \in GL(n,\ \ )$, we let $\sigma(p, g) = p \cdot g$ be the frame $(\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_n)$ at x, where

$$\hat{\boldsymbol{b}}_j = \boldsymbol{b}_i \boldsymbol{g}^i{}_j, \qquad j = 1, \ldots, n. \tag{3.3.1}$$

Formally, we may write

$$(\hat{\boldsymbol{b}}_1 \cdots \hat{\boldsymbol{b}}_n) = (\boldsymbol{b}_1 \cdots \boldsymbol{b}_n) \begin{pmatrix} g^1{}_1 & \cdots & g^1{}_n \\ \vdots & & \vdots \\ g^n{}_1 & \cdots & g^n{}_n \end{pmatrix} \tag{3.3.2}$$

and from this it is clear that $p \cdot (g_1 \, g_2) = (p \cdot g_1) \cdot g_2$ and $p \cdot id = p$. Moreover, by definition, $\mathcal{P}_L(p \cdot g) = \mathcal{P}_L(p)$ for all $p \in L(X)$ and $g \in GL(n,\ \ )$. It will be useful to have a description of this action when a frame is identified with an isomorphism $\underset{\sim}{p} : \ {}^n \longrightarrow T_x(X)$. Thus, we identify any $g \in GL(n,\ \ )$ with an isomorphism $\underset{\sim}{g} : \ {}^n \longrightarrow \ {}^n$ given by

$$\underset{\sim}{g}(e_j) = e_i \, g^i{}_j, \qquad j = 1, \ldots, n, \tag{3.3.3}$$

so that

$$\Big(\underset{\sim}{g}(e_i) \cdots \underset{\sim}{g}(e_n)\Big) = (e_1 \cdots e_n) \begin{pmatrix} g^1{}_1 & \cdots & g^1{}_n \\ \vdots & & \vdots \\ g^n{}_1 & \cdots & g^n{}_n \end{pmatrix}. \tag{3.3.4}$$

Then the frame $p \cdot g$ corresponds to the isomorphism

$$\underset{\sim}{p \cdot g} = \underset{\sim}{p} \circ \underset{\sim}{g} : \ {}^n \longrightarrow T_x(X) \tag{3.3.5}$$

since

$$\Big(\underset{\sim}{p} \circ \underset{\sim}{g}\Big)(e_j) = \underset{\sim}{p}(e_i g^i{}_j) = \underset{\sim}{p}(e_i) g^i{}_j = \boldsymbol{b}_i g^i{}_j = \hat{\boldsymbol{b}}_j$$

for $j = 1, \ldots, n$.

Our objective now is to provide $L(X)$ with a topology and manifold structure in such a way that, with the action σ described above,

$$GL(n,\ \ ) \hookrightarrow L(X) \stackrel{\mathcal{P}_L}{\longrightarrow} X$$

is a smooth principal $GL(n,\ \ )$-bundle over X, called the (**linear**) **frame bundle** of X. Toward this end we let (U, φ) be any chart on X with coordinate functions $x^1, \ldots, x^n$. Define $\tilde{\varphi} : \mathcal{P}_L^{-1}(U) \longrightarrow \varphi(U) \times GL(n,\ \ )$ as follows: Let

$p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ be a frame at $x = \mathcal{P}_L(p) \in U$. For each $j = 1, \ldots, n$, write

$$\boldsymbol{b}_j = \frac{\partial}{\partial x^i}\bigg|_x A^i{}_j(p).$$

The matrix

$$A(p) = \left(A^i{}_j(p)\right)$$

is nonsingular so we may set

$$\tilde{\varphi}(p) = (\varphi(x),\ A(p)). \tag{3.3.6}$$

Now, $\tilde{\varphi}$ is clearly one-to-one and any $A \in GL(n,\ )$ corresponds to some frame at each $x \in U$ so $\tilde{\varphi}$ is a bijection of $\mathcal{P}_L^{-1}(U)$ onto $\varphi(U) \times GL(n,\ )$. Moreover, $\varphi(U) \times GL(n,\ )$ is an open set in ${}^n \times {}^{n^2} \cong {}^{n+n^2}$. We define a topology on $L(X)$ by declaring that a subset $\mathcal{U}$ of $L(X)$ is open if and only if, for each chart (U, φ) on X, $\tilde{\varphi}(\mathcal{U} \cap \mathcal{P}_L^{-1}(U))$ is open in $\tilde{\varphi}(\mathcal{P}_L^{-1}(U)) = \varphi(U) \times GL(n,\ )$.

Exercise 3.3.1 Show that the collection of all such subsets $\mathcal{U}$ of $L(X)$ does, indeed, define a topology for $L(X)$ and that, if (V, ψ) is any chart for X, then $\mathcal{P}_L^{-1}(V)$ is open in $L(X)$.

Next we consider two charts (U, φ) and (V, ψ) for X with coordinate functions $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$, respectively, and with $U \cap V \neq \emptyset$. Then $\tilde{\varphi}$ and $\tilde{\psi}$ are both defined on $\mathcal{P}_L^{-1}(U \cap V)$, which is open in $L(X)$. We compute

$$\tilde{\psi} \circ \tilde{\varphi}^{-1} : \varphi(U \cap V) \times GL(n,\ ) \longrightarrow \psi(U \cap V) \times GL(n,\ ).$$

Let $((x^1, \ldots, x^n), (A^i{}_j))$ be in $\varphi(U \cap V) \times GL(n,\ )$. Then $\tilde{\varphi}^{-1}\ ((x^1, \ldots, x^n), (A^i{}_j))$ is the frame $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ at $x = \varphi^{-1}(x^1, \ldots, x^n)$ satisfying

$$\boldsymbol{b}_j = \frac{\partial}{\partial x^i}\bigg|_x A^i{}_j\,, \quad j = 1, \ldots, n.$$

Moreover, $\tilde{\psi}(p) = (\psi(x), (B^i{}_j)) = ((\psi \circ \varphi^{-1})(x^1, \ldots, x^n), (B^i{}_j))$, where

$$\boldsymbol{b}_j = \frac{\partial}{\partial y^i}\bigg|_x B^i{}_j\,, \quad j = 1, \ldots, n.$$

But

$$\boldsymbol{b}_j = \frac{\partial}{\partial x^k}\bigg|_x A^k{}_j = \left(\frac{\partial}{\partial y^i}\bigg|_x \frac{\partial y^i}{\partial x^k}(x)\right) A^k{}_j = \frac{\partial}{\partial y^i}\bigg|_x \left(\frac{\partial y^i}{\partial x^k}(x) A^k{}_j\right)$$

so

$$B^i{}_j = \frac{\partial y^i}{\partial x^k}(x)\, A^k{}_j, \quad i, j = 1, \ldots, n.$$

Thus,

$$\left(\tilde{\psi} \circ \tilde{\varphi}^{-1}\right)((x^1, \ldots, x^n),\ (A^i{}_j))$$
$$= \left((\psi \circ \varphi^{-1})(x^1, \ldots, x^n),\ \left(\frac{\partial y^i}{\partial x^k}(x)\, A^k{}_j\right)\right), \tag{3.3.7}$$

which, in particular, is C^∞. Reversing the roles of $\tilde{\psi}$ and $\tilde{\varphi}$ we find that $\tilde{\psi} \circ \tilde{\varphi}^{-1}$ and $\tilde{\varphi} \circ \tilde{\psi}^{-1}$ are inverse diffeomorphisms and we will use this fact to provide $L(X)$ with a differentiable structure.

First we show that, for any chart (U, φ) on X, the map $\tilde{\varphi} : \mathcal{P}_L^{-1}(U) \longrightarrow \varphi(U) \times GL(n,\ )$ is a homeomorphism (so that, in particular, $L(X)$ is locally Euclidean). We need only show that $\tilde{\varphi}$ is continuous and an open map. For the latter, we let W be an open set in $P_L^{-1}(U)$. Since $\mathcal{P}_L^{-1}(U)$ is open in $L(X)$, W is open in $L(X)$ so $\tilde{\varphi}(W \cap \mathcal{P}_L^{-1}(U)) = \tilde{\varphi}(W)$ is open in $\varphi(U) \times GL(n,\ )$, as required. To prove continuity we let Z be open in $\varphi(U) \times GL(n,\ )$. To show that $\tilde{\varphi}^{-1}(Z)$ is open in $\mathcal{P}_L^{-1}(U)$ we let (V, ψ) be an arbitrary chart for X with $U \cap V \neq \emptyset$. We must show that $\tilde{\psi}(\tilde{\varphi}^{-1}(Z) \cap \mathcal{P}_L^{-1}(V))$ is open in $\psi(V) \times GL(n,\ )$. But

$$\begin{aligned}
\tilde{\psi}(\tilde{\varphi}^{-1}(Z) \cap \mathcal{P}_L^{-1}(V)) &= \tilde{\psi}\left(\tilde{\varphi}^{-1}(Z) \cap \mathcal{P}_L^{-1}(U) \cap \mathcal{P}_L^{-1}(V)\right) \\
&= \tilde{\psi}\left(\tilde{\varphi}^{-1}(Z) \cap \mathcal{P}_L^{-1}(U \cap V)\right) \\
&= \tilde{\psi}\left(\tilde{\varphi}^{-1}(Z) \cap \tilde{\varphi}^{-1}(\varphi(U \cap V) \times GL(n,\ ))\right) \\
&= \tilde{\psi}\left(\tilde{\varphi}^{-1}(Z \cap (\varphi(U \cap V) \times GL(n,\ )))\right) \\
&= (\tilde{\psi} \circ \tilde{\varphi}^{-1})\left(Z \cap (\varphi(U \cap V) \times GL(n,\ ))\right)
\end{aligned}$$

which is open because $\tilde{\psi} \circ \tilde{\varphi}^{-1}$ is a diffeomorphism.

Exercise 3.3.2 Show that $L(X)$ is Hausdorff and second countable.

We have thus far shown that $L(X)$ is a topological manifold of dimension $n + n^2$ and that $\{(\mathcal{P}_L^{-1}(U), \tilde{\varphi}) : (U, \varphi)$ is a chart for $X\}$ is an atlas for $L(X)$. We provide $L(X)$ with the differentiable structure determined by this atlas. Next we show that, with this differentiable structure, the projection $\mathcal{P}_L : L(X) \longrightarrow X$ is smooth. For this it will suffice to show that, for any chart (U, φ) on X, the coordinate expression $\varphi \circ \mathcal{P}_L \circ \tilde{\varphi}^{-1}$ is C^∞. But $\varphi \circ \mathcal{P}_L \circ \tilde{\varphi}^{-1}((x^1, \ldots, x^n), (A^i{}_j)) = (x^1, \ldots, x^n)$ so this is clear.

Exercise 3.3.3 Show that the right action $\sigma : L(X) \times GL(n,\ ) \longrightarrow L(X)$ defined above is smooth.

To complete the proof that $GL(n,\ ) \hookrightarrow L(X) \xrightarrow{\mathcal{P}_L} X$ is a smooth principal $GL(n,\ )$-bundle we need only exhibit local trivializations. For each chart

(U, φ) on X we define

$$\mathbf{\Phi} : \mathcal{P}_L^{-1}(U) \longrightarrow U \times GL(n, \quad)$$

as follows: For $p \in L(X)_x \subseteq \mathcal{P}_L^{-1}(U)$,

$$\begin{aligned} \mathbf{\Phi}(p) &= \left(\varphi^{-1} \times id_{GL(n, \quad)}\right) \circ \tilde{\varphi}(p) \\ &= \left(\varphi^{-1} \times id_{GL(n, \quad)}\right) (\varphi(x),\ A(p)) \\ &= (x, A(p)) \\ &= (\mathcal{P}_L(p), A(p)). \end{aligned} \tag{3.3.8}$$

Then $\mathbf{\Phi}$ is a diffeomorphism and we need only check that $A(p \cdot g) = A(p)g$ for all $p \in \mathcal{P}_L^{-1}(U)$ and $g \in GL(n, \quad)$. Let $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \in L(X)_x$, where $x \in U$. Then $A(p) = (A^i{}_j(p))$, where

$$\boldsymbol{b}_j = \frac{\partial}{\partial x^i}\bigg|_x A^i{}_j(p), \quad j = 1, \ldots, n.$$

Now, $p \cdot g = (\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_n)$, where $\hat{\boldsymbol{b}}_j = \boldsymbol{b}_i g^i{}_j$, $j = 1, \ldots, n$, so $A(p \cdot g) = (A^i{}_j(p \cdot g))$, where

$$\hat{\boldsymbol{b}}_j = \frac{\partial}{\partial x^i}\bigg|_x A^i{}_j(p \cdot g).$$

But then

$$\hat{\boldsymbol{b}}_j = \boldsymbol{b}_k g^k{}_j = \frac{\partial}{\partial x^i}\bigg|_x A^i{}_k(p) g^k{}_j$$

implies

$$A^i{}_j(p \cdot g) = A^i{}_k(p) g^k{}_j, \qquad i, j = 1, \ldots, n.$$

Since the right-hand side of this last equality is the (i, j)-entry in the matrix product $A(p)g$ we have $A(p \cdot g) = A(p)g$, as required.

Notice that, if Φ is the trivialization arising from (U, φ), with coordinate functions $x^1, \ldots, x^n$ and Ψ is the trivialization arising from (V, ψ) with coordinate functions $y^1, \ldots, y^n$ and if $U \cap V \neq \emptyset$, then it follows from (3.3.7) that the transition function $g_{VU} : U \cap V \longrightarrow GL(n, \quad)$ is just the Jacobian of the coordinate transformation, i.e.,

$$g_{VU}(x) = \left(\frac{\partial y^i}{\partial x^k}(x)\right)$$

(Lemma 4.2.1, [**N4**]).

A local cross-section $s : U \longrightarrow L(X)$ of the frame bundle assigns to each $x \in U$ a frame $s(x) = (\boldsymbol{b}_1(x), \ldots, \boldsymbol{b}_n(x))$ at x and is called a (**linear**) **frame field**, or **moving frame**, on U.

Exercise 3.3.4 Show that each $\boldsymbol{b}_i(x)$, $i = 1, \ldots, n$, defines a smooth vector field on U and conclude that X is parallelizable (Exercise 5.8.17, [**N4**]) if and only if its frame bundle is trivial (admits a global cross-section).

Any Lie group is parallelizable (Exercise 5.8.17, [**N4**]) so, in particular, S^1 and S^3 have trivial frame bundles. S^7 is not a Lie group, but is close enough (via Cayley multiplication) and turns out to be parallelizable. It is a deep result of Bott and Milnor that S^1, S^3 and S^7 are the only parallelizable spheres.

Next we consider a number of vector bundles associated with the frame bundle by various representations of $GL(n,\ )$. First consider the natural representation $\rho : GL(n,\ ) \longrightarrow GL(\ ^n)$ of $GL(n,\ )$ on $\ ^n$ given by $\rho(g)(v) = g \cdot v = gv$, where $v \in\ ^n$ is written as a column matrix and gv denotes matrix multiplication. The associated vector bundle $L(X) \times_\rho\ ^n$ is called the **tangent bundle** of X and is denoted $T(X)$. The terminology would seem to suggest a bundle whose fibers are the tangent spaces to X and we would like to spend a moment showing that $T(X)$ is just that. First select (arbitrarily) some frame at each $x \in X$ and thereby identify each tangent space $T_x(X)$ with $\ ^n$. Any other frame at x then corresponds to a basis for $\ ^n$ and any two such bases are related by a unique element of $GL(n,\ )$. Now, the vector bundle associated to $GL(n,\ ) \hookrightarrow L(X) \xrightarrow{\mathcal{P}_L} X$ by ρ is a quotient of $L(X) \times\ ^n$ (see the construction in Section 6.7, [**N4**]). An element of $L(X) \times\ ^n$ is a pair (p, v), where $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ is a frame at $\mathcal{P}(p) = x$ and $v = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ is now identified with the tangent vector at x whose components relative to $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ are $v^1, \ldots, v^n$ ($v = v^i \boldsymbol{b}_i$). The action of $GL(n,\ )$ on $L(X) \times\ ^n$ whose orbit space is the vector bundle $T(X)$ is given by

$$(p, v) \cdot g = (p \cdot g,\ g^{-1} \cdot v).$$

Now, $p \cdot g$ is the new frame at x given by

$$p \cdot g = (\hat{\boldsymbol{b}}_1 \cdots \hat{\boldsymbol{b}}_n) = (\boldsymbol{b}_1 \cdots \boldsymbol{b}_n) \begin{pmatrix} g^1{}_1 & \cdots & g^1{}_n \\ \vdots & & \vdots \\ g^n{}_1 & \cdots & g^n{}_n \end{pmatrix}$$

and $g^{-1} \cdot v$ is given by

$$\begin{pmatrix} \hat{v}^1 \\ \vdots \\ \hat{v}^n \end{pmatrix} = \begin{pmatrix} g^1{}_1 & \cdots & g^1{}_n \\ \vdots & & \vdots \\ g^n{}_1 & \cdots & g^n{}_n \end{pmatrix}^{-1} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$$

$$= \begin{pmatrix} g_1{}^1 & \cdots & g_n{}^1 \\ \vdots & & \vdots \\ g_1{}^n & \cdots & g_n{}^n \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}.$$

Observe that $\hat{v}^1, \ldots, \hat{v}^n$ are precisely the components of $v^i \boldsymbol{b}_i$ relative to $\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_n$ since

$$\hat{v}^j \hat{\boldsymbol{b}}_j = (g_\alpha{}^j v^\alpha)(\boldsymbol{b}_\beta g^\beta{}_j) = (g^\beta{}_j g_\alpha{}^j) v^\alpha \boldsymbol{b}_\beta = \delta^\beta{}_\alpha v^\alpha \boldsymbol{b}_\beta = v^\alpha \boldsymbol{b}_\alpha.$$

Thus, an equivalence class

$$[p, v] = \{(p \cdot g, g^{-1} \cdot v) : g \in GL(n, \ )\}$$

is just the set of all possible descriptions of some fixed tangent vector at x. Note that, despite the ultramodern attire, this is just a dressed-up version of the "old-fashioned" view of a vector as a collection of n-tuples, one for each basis, related by the transformation law ρ.

A vector field on an open subset U of X can now be identified with a local cross-section $\boldsymbol{V} : U \longrightarrow T(X)$ of the tangent bundle. Equivalently (Section 6.8, [**N4**]), one can identify $\boldsymbol{V}$ with an $\ ^n$-valued map on $\mathcal{P}_L^{-1}(U)$ that is equivariant, i.e., satisfies $\boldsymbol{V}(p \cdot g) = g^{-1} \boldsymbol{V}(p)$ for each $p \in \mathcal{P}_L^{-1}(U)$ and $g \in GL(n, \ )$.

Similarly, if one defines a representation $\rho \colon GL(n, \ ) \longrightarrow GL(\ ^n)$ by $\rho(g)(\theta) = g \cdot \theta = (g^T)^{-1}\theta$, where $\theta \in \ ^n$ is written as a column matrix and $(g^T)^{-1}\theta$ denotes matrix multiplication, then the associated vector bundle is called the **cotangent bundle** and denoted $T^*(X)$. A 1-form can then be identified with either a cross-section of $T^*(X)$ or an $\ ^n$-valued map on $L(X)$ that is equivariant with respect to this representation ($\boldsymbol{\theta}(p \cdot g) = g^T \boldsymbol{\theta}(p)$). **Tensor bundles** (and their cross-sections, or equivariant maps, called **tensor fields**) arise in exactly the same way by making other choices for the representation ρ.

If an n-dimensional manifold X has a Riemannian metric $\boldsymbol{g}$ defined on it, then a frame $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ at $x \in X$ is orthonormal if $\boldsymbol{g}(\boldsymbol{b}_i, \boldsymbol{b}_j) = \delta_{ij}, i, j = 1, \ldots, n$. Such frames are related by elements of the orthogonal group $O(n) \subseteq GL(n, \ )$ and we wish to build an "orthonormal frame bundle" with group $O(n)$ analogous to the linear frame bundle $GL(n, \ ) \hookrightarrow L(X) \xrightarrow{\mathcal{P}_L} X$. However, we will need the construction in the indefinite case as well so we begin by generalizing what we know about the positive definite case. Most of the proofs are virtually identical and so will be left to the reader in a sequence of exercises.

On $\ ^n$ we will denote by $\langle\ ,\ \rangle_k, 0 \leq k \leq n$, the standard inner product of index $n-k$. Thus, if $\{e_1, \ldots, e_n\}$ is the standard basis for $\ ^n$, $x = x^i e_i$ and $y = y^j e_j$, then

$$\langle x,\ y\rangle_k = x^1 y^1 + \cdots + x^k y^k - x^{k+1} y^{k+1} - \cdots - x^n y^n = \eta_{ij} x^i y^j,$$

where

$$\eta = (\eta_{ij}) = \begin{pmatrix} id_{k\times k} & 0 \\ 0 & -id_{(n-k)\times(n-k)} \end{pmatrix}.$$

Equipped with this (indefinite) inner product, $\ ^n$ will be denoted $\ ^{k,n-k}$. When $k = n$ we naturally write $\ ^{n,0} = \ ^n$. The case of particular interest to us will be $k = 1,\ n \geq 2$. $\ ^{1,n-1}$ is called ***n*-dimensional Minkowski space.** For reasons that we will discuss later, $\ ^{1,3}$ is called **Minkowski spacetime.**

Exercise 3.3.5 Show that a linear transformation $A : \ ^{k,n-k} \longrightarrow \ ^{k,n-k}$ satisfies $\langle Ax, Ay\rangle_k = \langle x, y\rangle_k$ for all $x, y \in \ ^{k,n-k}$ if and only if its matrix with respect to $\{e_1, \ldots, e_n\}$, also denoted A, satisfies $A^T \eta\, A = \eta$. **Hint**: Mimic the proof of (1.1.24) in [**N4**].

Motivated by Exercise 3.3.5 we define the **semi-orthogonal group** $O(k, n-k)$ by

$$O(k, n-k) = \left\{A \in \mathcal{GL}(n,\ ) : A^T \eta A = \eta\right\}.$$

Exercise 3.3.6 Show that $O(k, n-k)$ is, indeed, a group under matrix multiplication and that $\det A = \pm 1$ for every $A \in O(k, n-k)$.

The subgroup

$$SO(k, n-k) = \left\{A \in O(k, n-k) : \det A = 1\right\}$$

is called the **special semi-orthogonal group**. Of course, when $k = n$ we write $O(n, 0) = O(n)$ and $SO(n, 0) = SO(n)$. We will have more to say later about $O(1, 3)$, called the **general Lorentz group** and denoted $\mathcal{L}$, and $SO(1, 3)$, called the **proper Lorentz group** and denoted $\mathcal{L}_+$.

Exercise 3.3.7 Show that, for any $\theta \in\ $, the matrix

$$L(\theta) = \begin{pmatrix} \cosh\theta & 0 & 0 & -\sinh\theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\theta & 0 & 0 & \cosh\theta \end{pmatrix}$$

is in $\mathcal{L}_+$.

Exercise 3.3.8 Let S_n denote the set of symmetric elements of $\mathcal{GL}(n, \ \ )$ and define a map $f : \mathcal{GL}(n, \ \ ) \longrightarrow S_n$ by $f(A) = A^T \eta \, A$. By mimicing the arguments for $O(n)$ on page 257 of [**N4**], show that f is a smooth map, $\eta \in S_n$ is a regular value of f and

$$O(k, n-k) = f^{-1}(\eta).$$

From Exercise 3.3.8 (and Corollary 5.6.7, [**N4**]) we conclude that $O(k, n-k)$ is a submanifold of $\mathcal{GL}(n, \ \ ) \cong \ \ ^{n^2}$ of dimension $n^2 - \frac{1}{2}n(n+1) = \frac{1}{2}n(n-1)$. Since matrix multiplication on $\mathcal{GL}(n, \ \ )$ is smooth, $O(k, n-k)$ is a Lie group.

Remark: In the positive definite case ($k = n$), the orthogonal group $O(n)$ is compact (because the rows of any $A \in O(n)$ must form a Euclidean orthonormal basis for $\ \ ^n$ so $O(n) \subseteq \ \ ^{n^2}$ is bounded). This is not the case when $0 < k < n$, e.g., the elements of $O(1,3)$ described in Exercise 3.3.7 form an unbounded set.

Since the determinant function on $GL(n, \ \ )$ is continuous and $SO(k, n-k) = O(k, n-k) \cap \det^{-1}(0, \infty)$, the special semi-orthogonal group is an open submanifold of $O(k, n-k)$ and therefore is also a Lie group.

Since $SO(k, n-k)$ is an *open* submanifold of $O(k, n-k)$, the Lie algebras $so(k, n-k)$ and $o(k, n-k)$ are the same. One determines this Lie algebra in precisely the same way as for $O(n)$ and $SO(n)$.

Exercise 3.3.9 Mimic the arguments on pages 279–280 of [**N4**] to show that

$$so(k, n-k) = o(k, n-k) = \{A \in \mathcal{GL}(n, \ \ ) : A^T = -\eta A \eta\}.$$

Now, a metric $\boldsymbol{g}$ on an n-manifold X is said to be **semi-Riemannian of index** $n - k$ if, at each $x \in X$, the inner product $\boldsymbol{g}_x$ has index $n-k$. Then a frame $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ at x is **orthonormal** if

$$\boldsymbol{g}_x(\boldsymbol{b}_i, \boldsymbol{b}_j) = \eta_{ij} = \begin{cases} 1, & i = j = 1, \ldots, k \\ -1, & i = j = k+1, \ldots, n \, . \\ 0, & i \neq j \end{cases}$$

We denote by $F(X)_x$ the set of all orthonormal frames at x and let $F(X) = \bigcup_{x \in X} F(X)_x$. For each $p \in F(X)_x \subseteq F(X)$ we let $\mathcal{P}_F(p) = x$ and thereby obtain a surjective map

$$\mathcal{P}_F : F(X) \longrightarrow X.$$

Next we define a right action $\sigma : F(X) \times O(k, n-k) \longrightarrow F(X)$ of $O(k, n-k)$ on $F(X)$ as follows: For each $(p, g) \in F(X) \times O(k, n-k)$, with $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \in F(X)_x \subseteq F(X)$ and $g = (g^i{}_j) \in O(k, n-k)$, we let $\sigma(p, g) = p \cdot g \in F(X)_x$ be the frame $(\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_n)$ at x, where

$$\hat{\boldsymbol{b}}_j = \boldsymbol{b}_i g^i{}_j, \quad j = 1, \ldots, n.$$

Notice that this frame is, indeed, orthonormal by Exercise 3.3.4 and that σ is a right action for precisely the same reason that the analogous map on $L(X)\times GL(n, \ \ )$ is a right action (see (3.3.2)). By definition, $\mathcal{P}_F(p\cdot g) = \mathcal{P}_F(p)$ for all $(p, g) \in F(X)\times O(k, n-k)$. Just as for $L(X)$ it will sometimes be convenient to identify an orthonormal frame p with the corresponding isomorphism $\underset{\sim}{p} : \ \ ^{k,n-k} \longrightarrow T_x(X)$.

We wish to show that

$$O(k, n-k) \hookrightarrow F(X) \xrightarrow{\mathcal{P}_F} X$$

is a smooth principal $O(k, n-k)$-bundle over X, called the **orthonormal frame bundle** of X (corresponding to the metric $\boldsymbol{g}$). The arguments are essentially identical to those given for the linear frame bundle $GL(n, \ \ ) \hookrightarrow L(X) \xrightarrow{\mathcal{P}_L} X$ except that the local coordinate frame fields $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\}$ must be replaced by local orthonormal frame fields $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n\}$. The existence of these is proved in the positive definite case in Proposition 5.11.3 of [**N4**]. We will leave it to the reader to supply the necessary refinements of the argument in the indefinite case.

Exercise 3.3.10 Let X be a smooth n-manifold and $\boldsymbol{g}$ a semi-Riemannian metric of index $n-k, 0 \leq k \leq n$, on X. Show that, for each $x_0 \in X$, there exists a connected open neighborhood U of x_0 in X and smooth vector fields $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n$ on U such that

$$\boldsymbol{g}_{x_0}(\boldsymbol{E}_i(x), \boldsymbol{E}_j(x)) = \eta_{ij}, \quad i, j = 1, \ldots, n,$$

for all $x \in U$. Show, furthermore, that, if X has an orientation μ, then one can ensure $\{\boldsymbol{E}_1(x), \ldots, \boldsymbol{E}_n(x)\} \in \mu_x$ for each $x \in U$. $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n\}$ is referred to as an (**oriented**) **orthonormal frame field** on U.

Now, given a chart (U, φ) for X we may shrink U if necessary and assume that it has defined on it an orthonormal frame field $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n\}$. Define $\tilde{\varphi} : \mathcal{P}_F^{-1}(U) \longrightarrow \varphi(U) \times O(k, n-k)$ as follows: Let $p = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ be an orthonormal frame at $x = \mathcal{P}_F(p) \in U$. For each $j = 1, \ldots, n$ we write

$$\boldsymbol{b}_j = \boldsymbol{E}_i(x) A^i{}_j(p).$$

Then the matrix $A(p) = (A^i{}_j(p))$ is in $O(k, n-k)$ so we may set

$$\tilde{\varphi}(p) = (\varphi(x), A(p)).$$

$\tilde{\varphi}$ is clearly one-to-one and any $A \in O(k, n-k)$ corresponds to some orthonormal frame at each $x \in U$ so $\tilde{\varphi}$ is a bijection of $\mathcal{P}_F^{-1}(U)$ onto the open set $\varphi(U) \times O(k, n-k)$ in $\ \ ^n \times O(k, n-k)$. Now, define a topology on $F(X)$ by declaring that a subset $\mathcal{U}$ of $F(X)$ is open if and only if, for each (U, φ), $\tilde{\varphi}(\mathcal{U} \cap \mathcal{P}_F^{-1}(U))$ is open in $\tilde{\varphi}(\mathcal{P}_F^{-1}(U)) = \varphi(U) \times O(k, n-k)$.

Exercise 3.3.11 Show that the collection of all such subsets $\mathcal{U}$ of $F(X)$ does, indeed, define a topology on $F(X)$ and that, if (V,ψ) is any chart for X, then $\mathcal{P}_F^{-1}(V)$ is open in $F(X)$.

Next let (U,φ) and (V,ψ) be two charts on X with $U\cap V\neq\emptyset$ and with orthonormal frame fields $\{\boldsymbol{E}_1,\ldots,\boldsymbol{E}_n\}$ and $\{\boldsymbol{F}_1,\ldots,\boldsymbol{F}_n\}$, respectively. For each $x\in U\cap V$ we write

$$\boldsymbol{E}_i(x)=\Lambda_{\boldsymbol{i}}{}^j(x)\boldsymbol{F}_j(x),\quad i=1,\ldots,n,$$

where $\Lambda_i{}^j(x)=\boldsymbol{g}_x(\boldsymbol{E}_i(x),\boldsymbol{F}_j(x))$.

Exercise 3.3.12 Show that

$$\tilde{\psi}\circ\tilde{\varphi}^{-1}:\varphi(U\cap V)\times O(k,n-k)\longrightarrow\psi(U\cap V)\times O(k,n-k)$$

is given by

$$\begin{aligned}&\left(\tilde{\psi}\circ\tilde{\varphi}^{-1}\right)\left((x^1,\ldots,x^n),(A^i{}_j)\right)\\&=\left((\psi\circ\varphi^{-1})(x^1,\ldots,x^n),\ \left(\Lambda_k{}^i(x)A^k{}_j\right)\right),\end{aligned}$$

where $x=\varphi^{-1}(x^1,\ldots,x^n)$.

In particular, it follows from Exercise 3.3.12 that $\tilde{\psi}\circ\tilde{\varphi}^{-1}$ is C^∞ so, reversing the roles of $\tilde{\psi}$ and $\tilde{\varphi}$, we find that $\tilde{\psi}\circ\tilde{\varphi}^{-1}$ and $\tilde{\varphi}\circ\tilde{\psi}^{-1}$ are inverse diffeomorphisms.

Exercise 3.3.13 Use this fact to show that $\tilde{\varphi}:\mathcal{P}_F^{-1}(U)\longrightarrow\varphi(U)\times O(k,n-k)$ is a homeomorphism.

Since $O(k,n-k)$ is locally Euclidean, it follows from Exercise 3.3.13 that the same is true of $F(X)$.

Exercise 3.3.14 Show that $F(X)$ is Hausdorff and second countable.

Exercise 3.3.15 Provide $F(X)$ with the differentiable structure determined by the atlas $\{(\mathcal{P}_F^{-1}(U),\tilde{\varphi}):(U,\varphi)\text{ is a chart on }X\}$ and show that $\mathcal{P}_F:F(X)\longrightarrow X$ and $\sigma:F(X)\times O(k,n-k)\longrightarrow F(X)$ are smooth.

Exercise 3.3.16 For each chart (U,φ) show that the map $\Phi:\mathcal{P}_F^{-1}(U)\longrightarrow U\times O(k,n-k)$ defined by

$$\Phi=\left(\varphi^{-1}\circ id_{O(k,n-k)}\right)\circ\tilde{\varphi}$$

is a local trivialization and that, if Ψ is the analogous trivialization arising from (V,ψ) with $U\cap V\neq\emptyset$, then the transition function $g_{VU}:U\cap V\longrightarrow O(k,n-k)$ is given by $g_{VU}(x)=(\Lambda_i{}^j(x))$.

This completes the construction of the orthonormal frame bundle

$$O(k, n-k) \hookrightarrow F(X) \xrightarrow{\mathcal{P}_F} X$$

for any semi-Riemannian manifold X and we now have every confidence that the reader can take the next step without assistance.

Exercise 3.3.17 Let X be an oriented semi-Riemannian manifold of index $n-k$. Construct, in detail, the **oriented**, **orthonormal frame bundle**

$$SO(k, n-k) \hookrightarrow F_+(X) \xrightarrow{\mathcal{P}_{F_+}} X.$$

Being a Lie group S^3 has a trivial linear frame bundle. It is instructive, and will be useful somewhat later, to explicitly construct a cross-section of the oriented, orthonormal frame bundle, thereby showing that it is trivial as well.

Exercise 3.3.18 Show that it will suffice to define smooth vector fields $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ and $\boldsymbol{V}_3$ on S^3 with the property that, for each $p \in S^3, \{\boldsymbol{V}_1(p)$, $\boldsymbol{V}_2(p)$, $\boldsymbol{V}_3(p)\}$ is an oriented, orthonormal basis for $T_p(S^3)$.

The procedure for constructing the vector fields described in Exercise 3.3.18 is quite simple. Regard $S^3 \subseteq {}^4 =$ as the unit quaternions. For each $p \in S^3$, the tangent space $T_p(S^3)$ can be identified with a subspace of $T_p({}^4)$ and this, in turn, is canonically identified with ^{4} itself. Viewed in this way, $T_p(S^3)$ is just the subspace of $T_p({}^4) = {}^4$ consisting of those $v = (v^0, v^1, v^2, v^3)$ with $\langle v, p \rangle = 0$, i.e., with$v^0 p^0 + v^1 p^1 + v^2 p^2 + v^3 p^3 = 0$. Now, select a point, say $p_0 = (1, 0, 0, 0)$, in S^3 and let

$$e_1 = (0, 1, 0, 0),\ e_2 = (0, 0, 1, 0),\ e_3 = (0, 0, 0, 1)$$

be the standard basis for $T_{p_0}(S^3)$. We "rotate" this basis to each $p \in S^3$ by quaternion multiplication, i.e., we define

$$\boldsymbol{V}_i(p) = e_i \cdot p, \quad i = 1, 2, 3.$$

Evaluating these quaternion products gives the components of $\boldsymbol{V}_i(p)$ in $T_p(S^3) \subseteq T_p({}^4) = {}^4$ in terms of the coordinates of p.

$$\begin{aligned} \boldsymbol{V}_1(p) &= (-p^1,\, p^0,\, -p^3,\, p^2) \\ \boldsymbol{V}_2(p) &= (-p^2,\, p^3,\, p^0,\, -p^1) \\ \boldsymbol{V}_3(p) &= (-p^3,\, -p^2,\, p^1,\, p^0) \end{aligned} \tag{3.3.9}$$

Notice that $\langle \boldsymbol{V}_i(p), p \rangle = 0, i = 1, 2, 3$, so these are indeed vectors in $T_p(S^3)$. Furthermore,

$$\langle \boldsymbol{V}_i(p),\, \boldsymbol{V}_j(p) \rangle = \delta_{ij}, \quad i,\, j = 1, 2, 3$$

for each p so $\{\boldsymbol{V}_1(p), \boldsymbol{V}_2(p), \boldsymbol{V}_3(p)\}$ is an orthonormal basis for $T_p(S^3)$ (the metric on S^3 is the restriction to S^3 of the standard metric on ^{4}). Now, an ordered basis for $T_p(S^3)$ is in the standard orientation for S^3 if and only if one obtains an oriented basis for $\ ^4$ by adjoining to it (at the beginning) the "position vector" p (this is a special case of Exercise 4.3.7 which may be consulted, and solved, at this point). Thus, all that remains is to show that the determinant of the matrix

$$\begin{pmatrix} p^0 & p^1 & p^2 & p^3 \\ -p^1 & p^0 & -p^3 & p^2 \\ -p^2 & p^3 & p^0 & -p^1 \\ -p^3 & -p^2 & p^1 & p^0 \end{pmatrix} \tag{3.3.10}$$

is positive. One can compute this directly or argue indirectly as follows: Since the rows of (3.3.10) are orthogonal unit vectors in $\ ^4$ the matrix itself is an orthogonal matrix and so has determinant ± 1 at each $p \in S^3$. But the determinant function is continuous and S^3 is connected so the determinant of (3.3.10) is either 1 for all $p \in S^3$ or -1 for all $p \in S^3$. Since this determinant is obviously 1 when $p = (1, 0, 0, 0) \in S^3$ the result follows.

We will have occasion somewhat later to introduce yet one more frame bundle (the "oriented, time oriented, orthonormal frame bundle" of a "spacetime" manifold). For the present we will conclude this discussion with a few remarks on some topics we will not pursue in any depth here. According to Theorem 3.1.7, any principal bundle has connections defined on it. and this is true, in particular, for the frame bundles we have constructed. A connection on the linear frame bundle $GL(n, \ ) \hookrightarrow L(X) \longrightarrow X$ is called a **linear connection** on X and these are of fundamental importance in the study of the geometry of X (see Chapter III of [**KN1**]). If X admits a Riemannian or semi-Riemannian metric, then there is, among these linear connections, a distinguished one called the **Levi-Civita connection** (or **Riemannian connection**) that is adapted to the metric structure. The study of this connection is the vast and beautiful subject of (**semi-**) **Riemannian geometry**. Although this is not our subject here we intend to borrow one of its results. Moreover, the Levi-Civita connection plays a role in defining the Seiberg-Witten equations which we will sketch in the Appendix so a brief tour of the definition is probably in order (for details one can consult [**Bl**], [**O'N**], or [**KN1**]). We consider a smooth n-manifold X and its linear frame bundle $GL(n, \ ) \hookrightarrow L(X) \longrightarrow X$. A connection $\boldsymbol{\omega}$ on $L(X)$ is a $gl(n, \mathbb{R})$-valued 1-form on $L(X)$ which we describe as follows. Let E_i^j be the $n \times n$ matrix for which the entry in the i^{th} row and j^{th} column is 1 and all other entries are 0. Then $\boldsymbol{\omega}$ can be written as $\boldsymbol{\omega} = \boldsymbol{\omega}^i_j E_i^j$, where each $\boldsymbol{\omega}^i_j$ is a real-valued 1-form on $L(X)$. Similarly, the curvature $\boldsymbol{\Omega} = d\boldsymbol{\omega} + \frac{1}{2}[\boldsymbol{\omega}, \boldsymbol{\omega}]$ can be written $\boldsymbol{\Omega} = \boldsymbol{\Omega}^i_j E_i^j$, where each $\boldsymbol{\Omega}^i_j$ is a real-valued 2-form on $L(X)$. Thus,

$$\boldsymbol{\Omega}^i_j = d\boldsymbol{\omega}^i_j + \boldsymbol{\omega}^i_k \wedge \boldsymbol{\omega}^k_j.$$

If U is a coordinate neighborhood in X with coordinates $x^1, \ldots, x^n$ and $S_U : U \longrightarrow L(X)$ is the corresponding cross-section

$$S_U(x) = (\partial_1(x), \ldots, \partial_n(x)),$$

then the pullback $\boldsymbol{\omega}_U = (S_U)^*\boldsymbol{\omega}$ can be written

$$\boldsymbol{\omega}_U = (S_U)^*\boldsymbol{\omega} = ((S_U)^*\boldsymbol{\omega}^i_k)E^k_i = (\Gamma^i_{jk}dx^j)E^k_i$$

for some smooth functions Γ^i_{jk} on U (called the **Christoffel symbols** for $\boldsymbol{\omega}$ in the coordinate neighborhood U). Similarly,

$$\boldsymbol{\Omega}_U = (S_U)^*\boldsymbol{\Omega} = ((S_U)^*\boldsymbol{\Omega}^i_j)E^j_i = \left(\frac{1}{2}(R^i_{jkl}dx^k \wedge dx^l)E^j_i\right)$$

for some smooth functions R^i_{jkl} on U. Unraveling the definitions gives

$$R^i_{jkl} = \partial_k\Gamma^i_{lj} - \partial_l\Gamma^i_{kj} + \Gamma^m_{lj}\Gamma^i_{km} - \Gamma^m_{kj}\Gamma^i_{lm}$$

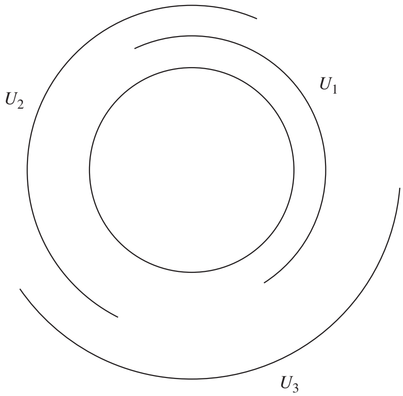
for $i, j, k, l = 1, \ldots, n$.

Now suppose that X has a metric $\boldsymbol{g}$ of index $n - k$ and consider the orthonormal frame bundle $O(k, n-k) \hookrightarrow F(X) \longrightarrow X$. If X is also oriented we have an oriented, orthonormal frame bundle $SO(k, n-k) \hookrightarrow F_+(X) \longrightarrow X$ as well. It is not difficult to see that any connection on $F(X)$ or $F_+(X)$ extends uniquely to a connection on $L(X)$ (page 158 of [**KN1**]), but it is not true that every connection on $L(X)$ restricts to a connection on $F(X)$ or $F_+(X)$ because a connection on $L(X)$ takes values in $gl(n,\ )$ and not necessarily in $o(k, n-k) = so(k, n-k)$; those that do are called **metric connections** on $L(X)$. The so-called **Fundamental Theorem of Riemannian Geometry** states that there is a unique metric connection on $L(X)$ that is also *symmetric* in the sense that $\Gamma^i_{jk} = \Gamma^i_{kj}$ for all $i, j, k = 1, \ldots, n$ in any local coordinate system. This is called the **Levi-Civita connection** and denoted $\boldsymbol{\omega}_{LC}$ (the same terminology and notation is used for its restriction to either $F(X)$ or $F_+(X)$). One can show that $\boldsymbol{\omega}_{LC}$ is characterized by the fact that, in any local coordinate system, the Christoffel symbols are given in terms of the metric components by

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}(\partial_k g_{jl} + \partial_j g_{kl} - \partial_l g_{jk})$$

for $i, j, k, l = 1, \ldots, n$. All of the usual objects of study in (semi-) Riemannian geometry are defined in terms of the Levi-Civita connection. For example, R^i_{jkl} are the local components of the *Riemann curvature tensor*, $R_{ij} = R^k_{ikj}$ are the components of the *Ricci tensor*, and $R = g^{ik}R_{ik}$ is the *scalar curvature*.

When we define the Čech cohomology groups of a manifold in Chapter 6 we will require the notion (and the existence) of "simple covers" for a smooth manifold. The sort of cover we have in mind is illustrated for the circle S^1 below.



Here we have an open cover $\mathcal{U} = \{U_1, U_2, U_3\}$ that is locally finite (actually, finite in this case), each element of which has compact closure and which has the property that any nonempty finite intersection of its elements is diffeomorphic to the Euclidean space of the same dimension as the manifold ($U_1, U_2, U_3, U_1 \cap U_2, U_1 \cap U_3$, and $U_2 \cap U_3$ are all diffeomorphic to). In general, a **simple cover** of an n-manifold X is a locally finite open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in \mathcal{A}}$ such that any nonempty finite intersection of its elements is diffeomorphic to ^{n}. Simple covers are generally not that easy to find, but they always exist. Indeed, the following is essentially Theorem 3.7, Chapter IV, of [**KN1**].

Theorem 3.3.1 *Let X be an n-dimensional smooth manifold. Then any open cover for X has a refinement that is a countable simple cover of X.*

Exercise 3.3.19 Find a simple cover for the torus $S^1 \times S^1$.

If X has a finite simple cover, then it is said to be of **finite type**. Certainly, any compact manifold is of finite type, but so is ^{n}.

3.4 Minkowski Spacetime

As we mentioned earlier the indefinite inner product space ^{1,3} is called Minkowski spacetime and the associated semi-orthogonal groups $O(1,3)$ and $SO(1,3)$, called the Lorentz group and the proper Lorentz group, respectively, are denoted $\mathcal{L}$ and $\mathcal{L}_+$. In this section we will examine these objects in somewhat more detail and have a bit to say about their physical interpretation (also see Chapter 2).

Remark: A much more thorough study of both the mathematics and the physics is available in [**N3**].

We will bow to the generally accepted conventions of physics and use (x^0, x^1, x^2, x^3), rather than (x^1, x^2, x^3, x^4), for the standard coordinates in ^{1,3} and will drop the subscript 1 on the inner product $\langle\ ,\ \rangle_1$. Thus, if $\{e_0, e_1, e_2, e_3\}$ is the standard basis with $x = x^\alpha e_\alpha$ and $y = y^\beta e_\beta$, then

$$\langle x, y\rangle = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3 = \eta_{\alpha\beta} x^\alpha y^\beta$$

where

$$\eta_{\alpha\beta} = \begin{cases} 1, & \alpha = \beta = 0 \\ -1, & \alpha = \beta = 1, 2, 3\,. \\ 0, & \alpha \neq \beta \end{cases}$$

In particular, $\{e_0, e_1, e_2, e_3\}$ is an orthonormal basis for ^{1,3}. We will denote by $Q :\ ^{1,3} \longrightarrow$ the quadratic form $Q(x) = \langle x, x\rangle$ associated with $\langle\ ,\ \rangle$.

The elements of ^{1,3} are called **events** and should be identified intuitively with idealized physical events having no spatial extension and no duration, e.g., an instantaneous collision, or explosion, or an instant in the history of some point particle. The coordinates (x^0, x^1, x^2, x^3) of such an event are to be regarded as the spatial (x^1, x^2, x^3) and time (x^0) coordinates assigned to that event by some inertial observer with x^0 measured in units of distance, i.e., light travel time (1 meter of time, or 1 light meter, being the time required by a photon to travel 1 m *in vacuo*). The observer can then be identified with $\{e_0, e_1, e_2, e_3\}$. Another inertial observer is identified with another orthonormal basis $\{\hat{e}_0, \hat{e}_1, \hat{e}_2, \hat{e}_3\}$ (we assume the basis elements are always numbered in such a way that $\langle \hat{e}_\alpha, \hat{e}_\beta\} = \eta_{\alpha\beta}$). For any such basis there will exist a linear transformation $L :\ ^{1,3} \longrightarrow\ ^{1,3}$ with $L\hat{e}_\alpha = e_\alpha$ for $\alpha = 0, 1, 2, 3$ so that L is an orthogonal transformation of ^{1,3}, i.e.,

$$\langle Lx, Ly\rangle = \langle x, y\rangle$$

for all $x, y \in\ ^{1,3}$. Writing

$$e_\beta = \Lambda^\alpha{}_\beta \hat{e}_\alpha, \quad \beta = 0, 1, 2, 3,$$

we obtain a matrix $\Lambda = (\Lambda^\alpha{}_\beta)_{\alpha,\beta=0,1,2,3}$ (the matrix of L^{-1} relative to $\{\hat{e}_\alpha\}$) which must be in $\mathcal{L}$. If $x \in\ ^{1,3}$ and $x = x^\alpha e_\alpha = \hat{x}^\alpha \hat{e}_\alpha$, then

$$\hat{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta, \quad \alpha = 0, 1, 2, 3, \tag{3.4.1}$$

is the coordinate transformation between the two observers. The condition $\langle e_\alpha, e_\beta\rangle = \eta_{\alpha\beta}$ implies that Λ must satisfy

$$\Lambda^a{}_\alpha \Lambda^b{}_\beta \eta_{ab} = \eta_{\alpha\beta}, \quad \alpha, \beta = 0, 1, 2, 3. \tag{3.4.2}$$

In particular, when $\alpha = \beta = 0$, this gives

$$(\Lambda^0{}_0)^2 - (\Lambda^1{}_0)^2 - (\Lambda^2{}_0)^2 - (\Lambda^3{}_0)^2 = 1$$

so

$$\Lambda^0{}_0 \geq 1 \quad \text{or} \quad \Lambda^0{}_0 \leq -1. \tag{3.4.3}$$

Those elements of $\mathcal{L}$ for which $\Lambda^0{}_0 \geq 1$ are called **orthochronous** and the set

$$\mathcal{L}_+^{\uparrow} = \{\Lambda \in \mathcal{L}_+ : \Lambda^0{}_0 \geq 1\}$$

is called the **proper orthochronous Lorentz group**. For physical reasons that we will discuss shortly we will consider only those inertial observers (orthonormal bases) related to the standard basis by an element of $\mathcal{L}_+^{\uparrow}$. We call such bases **admissible**.

If gravitational effects are assumed negligible, then the **Minkowski inner product** $\langle\ ,\ \rangle$ on ^{1,3} has something physically significant to say about relationships between events. Consider, for example, two events $x_0, x_1 \in {}^{1,3}$ and the displacement vector $x = x_1 - x_0$ between them. Suppose that $Q(x) = 0$. Then, if we write $x = (\Delta x^\alpha)e_\alpha$,

$$(\Delta x^1)^2 + (\triangle x^2)^2 + (\Delta x^3)^2 = (\Delta x^0)^2 \tag{3.4.4}$$

and the same equation is satisfied in any other orthonormal basis. Consequently, the spatial separation of the two events x_0 and x_1 is numerically equal to the distance light would travel during the time lapse between them. The two events are "connectible by a light ray." In this case, x is said to be **null**, or **lightlike**. With x_0 held fixed the set of all x_1 for which $x = x_1 - x_0$ is null is called the **null cone**, or **light cone**, at x_0 because of the formal resemblance of (3.4.4) to the equation of a right circular cone in ^{3}. A straight line which lies entirely on such a null cone is called a **worldline of a photon** and is thought of as the set of all events in the history of some "particle of light."

Suppose instead that $Q(x) = Q(x_1 - x_0) > 0$. In this case we say that x is **timelike** and, in any orthonormal basis, we have

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 < (\Delta x^0)^2 \tag{3.4.5}$$

(x is "inside" the null cone at x_0). Then the spatial separation of x_0 and x_1 is less than the distance light would travel during the elapsed time between them so it is possible for some material particle (or observer) to experience both events without violating the speed limit imposed on such particles by special relativity. A straight line joining two such events is called a **timelike straight line**, or a **worldline of a free material particle** and can be identified intuitively with the set of all events in the history of an unaccelerated material particle that experiences both events.

If $Q(x) = Q(x_1 - x_0) < 0$, then $x = x_1 - x_0$ is said to be **spacelike** and lies outside the null cone at x_0. No known physical agency can experience both x_0 and x_1 since, to do so, such an agency would have to be transmitted at a speed greater than that of light. A smooth curve in ^{1,3} is said to be **spacelike**, **timelike**, or **null** if its tangent vector at each point is spacelike, timelike, or null, respectively.

Theorem 3.4.1 *Suppose x is timelike and y is either timelike, or null and nonzero. Let $\{e_0, e_1, e_2, e_3\}$ be any orthonormal basis for ^{1,3} with $x = x^\alpha e_\alpha$ and $y = y^\beta e_\beta$. Then either*

1. *$x^0 y^0 > 0$ in which case $\langle x, y \rangle > 0$, or*
2. *$x^0 y^0 < 0$ in which case $\langle x, y \rangle < 0$.*

Proof: By assumption, $\langle x, x \rangle = (x^0)^2 - \vec{x} \cdot \vec{x} > 0$ and $\langle y, y \rangle = (y^0)^2 - \vec{y} \cdot \vec{y} \geq 0$, where we have written $\vec{x} \cdot \vec{x}$ for $(x^1)^2 + (x^2)^2 + (x^3)^2$ and similarly for y. Thus, $(x^0 y^0)^2 > (\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y})$ so

$$|x^0 y^0| > \sqrt{(\vec{x} \cdot \vec{x})\ (\vec{y} \cdot \vec{y})}. \tag{3.4.6}$$

Now, for any real number $t, 0 \leq (ty^1 + x^1)^2 + (ty^2 + x^2)^2 + (ty^3 + x^3)^2$, i.e., $(\vec{y} \cdot \vec{y})t^2 + 2(\tilde{x} \cdot \vec{y})t + (\vec{x} \cdot \vec{x}) \geq 0$. The discriminant of this quadratic in t must therefore be less than or equal to zero and this gives $(\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y}) \geq (\vec{x} \cdot \vec{y})^2$. Combining this with (3.4.6) yields

$$|x^0 y^0| > |\vec{x} \cdot \vec{y}| = |x^1 y^1 + x^2 y^2 + x^3 y^3|. \tag{3.4.7}$$

In particular, $x^0 y^0 \neq 0$. Suppose that $x^0 y^0 > 0$. Then $x^0 y^0 = |x^0 y^0| > |x^1 y^1 + x^2 y^2 + x^3 y^3| \geq x^1 y^1 + x^2 y^2 + x^3 y^3$ so $\langle x, y \rangle > 0$. On the other hand, if $x^0 y^0 < 0$, then $\langle x, -y \rangle > 0$ so $\langle x, y \rangle < 0$. ■

Exercise 3.4.1 Show that if a nonzero vector in ^{1,3} is orthogonal to a timelike vector, then it must be spacelike.

If $Q(x_1 - x_0) = 0$, then x_0 and x_1 lie on the worldline of a photon so one of them can be regarded as the emission of a photon and the other as its subsequent reception somewhere else. We assume that admissible observers are related by *orthochronous* elements of $\mathcal{L}$ so that all of them will agree on which is which.

Exercise 3.4.2 Let $\Lambda = (\Lambda^\alpha{}_\beta)_{\alpha,\beta=0,1,2,3}$ be an element of $\mathcal{L}$. Show that the following are equivalent.

(a) Λ is orthochronous (i.e., $\Lambda^0{}_0 \geq 1$).
(b) Λ preserves the time orientation of all nonzero null vectors, i.e., if $x = x^\alpha e_\alpha$ is nonzero and null, then x^0 and $\hat{x}^0 = \Lambda^0{}_\beta x^\beta$ have the same sign.
(c) Λ preserves the time orientation of all timelike vectors.

Notice that Exercise 3.4.2 does not assert that admissible observers agree on the temporal order of x_0 and x_1 if $x_1 - x_0$ is spacelike and, indeed, they need not.

Exercise 3.4.3 Show that if $\Lambda \in \mathcal{L}$ is orthochronous, but $\det \Lambda = -1$ (see Exercise 3.3.6), then

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \Lambda$$

is in $\mathcal{L}_+^\uparrow$.

Since the matrix in Exercise 3.4.3 simply reverses the orientation of the spatial axes, our assumption that admissible observers are related by elements of $\mathcal{L}_+^\uparrow$ essentially amounts to the requirement that no one's clock runs backwards and no one uses "left-handed" spatial coordinates. Henceforth, we will refer to the elements of $\mathcal{L}_+^\uparrow$ simply as **Lorentz transformations**.

Exercise 3.4.4 Show that if $(R^i{}_j)_{i,j=1,2,3}$ is an element of $SO(3)$, then

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & (R^i{}_j) & \\ 0 & & & \end{pmatrix}$$

is in $\mathcal{L}_+^\uparrow$. Show also that the collection of all such elements of $\mathcal{L}_+^\uparrow$ is a subgroup $\mathcal{R}$ of $\mathcal{L}_+^\uparrow$ (called the **rotation subgroup** of $\mathcal{L}_+^\uparrow$).

If two observers are related by an element of $\mathcal{R}$, then they differ only in having spatial coordinate axes that are rotated relative to each other.

Exercise 3.4.5 Let $\Lambda = (\Lambda^\alpha{}_\beta)_{\alpha,\beta=0,1,2,3}$ be an element of $\mathcal{L}_+^\uparrow$. Show that the following are equivalent.

(a) $\Lambda \in \mathcal{R}$.
(b) $\Lambda^1{}_0 = \Lambda^2{}_0 = \Lambda^3{}_0 = 0$.
(c) $\Lambda^0{}_1 = \Lambda^0{}_2 = \Lambda^0{}_3 = 0$.
(d) $\Lambda^0{}_0 = 1$.

From the point of view of physics, the elements of $\mathcal{R}$ are rather dull. On the other hand, the matrices $L(\theta), \theta \in$, described in Exercise 3.3.7 are clearly in $\mathcal{L}_+^\uparrow (\cosh\theta \geq 1)$ and these are not at all dull. Physically, they correspond to inertial observers whose spatial coordinate axes are parallel and whose relative motion is along their common x^1-, $\hat{x}^1$-axis with speed $\beta = \tanh\theta$ (see Section 1.3 of [**N3**]). These Lorentz transformations $L(\theta)$ are called **boosts** and, in some sense, contain all of the interesting kinematical information in $\mathcal{L}_+^\uparrow$.

Theorem 3.4.2 *Let Λ be an element of $\mathcal{L}_+^\uparrow$. Then there exists a real number θ and two rotations R_1 and R_2 in $\mathcal{R} \subseteq \mathcal{L}_+^\uparrow$ such that*

$$\Lambda = R_1 L(\theta) R_2.$$

The intuitive content of this result is quite simple. The Lorentz transformation Λ from the frame of reference of observer $\mathcal{O}_1$ to the frame of reference of observer $\mathcal{O}_2$ can be accomplished in three stages: Rotate $\mathcal{O}_1$'s spatial axes so that the x^1-axis coincides with the line along which the relative motion takes place. Boost to a new frame whose spatial axes are parallel to the rotated axes of $\mathcal{O}_1$ and at rest relative to $\mathcal{O}_2$. Finally, rotate these new spatial axes so that they coincide with those of $\mathcal{O}_2$. For a detailed algebraic proof see Theorem 1.3.5 of [**N3**].

Remark: One can use the decomposition in Theorem 3.4.2 to define a deformation retraction of $\mathcal{L}_+^\uparrow$ onto $SO(3)$ and conclude that these two have the same homotopy type (Lemma 2.4.9, [**N4**]). Indeed, one can show more. $\mathcal{L}_+^\uparrow$ is actually homeomorphic to $SO(3) \times \ ^3$ (there is a nice proof of this on pages 73–74 of [**Bl**]). From either of these it follows that $\pi_1(\mathcal{L}_+^\uparrow) \cong \pi_1(SO(3)) \cong \mathbb{Z}_2$ (Appendix B of [**N3**]). We will need this fact only for motivational purposes in Section 6.5 and so will not give the details here. However, we make the following observation. On pages 92–93 we exhibited two smooth loops $R_1(t)$ and $R_2(t)$ in $SO(3)$ corresponding to a continuous rotation about the x-axis through 2π and 4π, respectively, and showed that they represent the two equivalence classes in $\pi_1(SO(3)) \cong \mathbb{Z}_2$. The same two curves, now thought of as loops in $\mathcal{R} \subseteq \mathcal{L}_+^\uparrow$ as in Exercise 3.4.4 represent the two classes in $\pi_1(\mathcal{L}_+^\uparrow) \cong \mathbb{Z}_2$.

We have already described (on pages 101–103) an alternative model of Minkowski spacetime as the real linear space $\mathcal{H}$ of 2×2 complex Hermitian matrices with Q given by the determinant (and $\langle\,,\,\rangle$ thereby determined via the Polarization Identity). This view of $\ ^{1,3}$ is particularly convenient for describing the double cover

$$\text{Spin} : SL(2, \ \ ) \longrightarrow \mathcal{L}_+^\uparrow$$

of $\mathcal{L}_+^\uparrow$. This, the reader may recall, was required to produce a ("2-valued") representation of $\mathcal{L}_+^\uparrow$ that is the appropriate transformation law (under change of inertial frame of reference) for the wavefunction of a spin $\frac{1}{2}$ particle (Section 2.4). We also briefly alluded (on pages 113–115) to the difficulties involved in generalizing these considerations to the situation in which gravitational fields are not negligible so that $\ ^{1,3}$ is no longer an accurate model of the "event world" and must be replaced by a spacetime manifold. It is to these objects that we now turn our attention.

3.5 Spacetime Manifolds and Spinor Structures

Unlike electromagnetic fields which, as we have seen (Sections 2.2, 2.3 and 2.4), fit very nicely within the framework of Minkowski spacetime (as 2- forms), gravitational fields stubbornly refuse to be modeled in ^{1,3}. The reason is quite simple (once it has been pointed out to you by Einstein). An electromagnetic field is something "external" to the structure of space-time, an additional field defined on and (apparently) not influencing the mathematical structure of ^{1,3}. Einstein realized that a gravitational field has a very special property which makes it unnatural to regard it as something external to the nature of the event world. Since Galileo it had been known that all objects with the same initial position and velocity respond to a given gravitational field in the same way (i.e., have identical worldlines) regardless of their material constitution (mass, charge, etc.). This is essentially what was verified at the Leaning Tower of Pisa and contrasts rather sharply with the behavior of electromagnetic fields. These worldlines (of particles with given initial conditions of motion) seem almost to be natural "grooves" in spacetime which anything will slide along if once placed there. But these "grooves" depend on the particular gravitational field being modeled and, in any case, ^{1,3} simply is not "grooved" (its structure does not distinguish any collection of curved worldlines). One suspects then that ^{1,3} itself is somehow lacking, that the appropriate mathematical structure for the event world may be more complex when gravitational effects are non-negligible.

To see how the structure of ^{1,3} might be generalized to accommodate the presence of gravitational fields let us begin with a structureless set X whose elements we call "events." One thing at least is clear: In regions that are distant from the source of any gravitational field, no accomodation is necessary and X must locally "look like" ^{1,3}. But a great deal more is true. In his now famous "Elevator Experiment" Einstein observed that *any* event has about it a sufficiently small region of X which "looks like" ^{1,3}. To see this we reason as follows: Imagine an elevator containing an observer and various other objects which is under the influence of some uniform external gravitational field. The cable snaps. The contents of the elevator are now in free fall. Since all of the objects inside respond to the gravitational field *in the same way* they will remain at relative rest throughout the fall. Indeed, if our observer lifts an apple from the floor and places it in mid-air next to his head it will remain there. You have witnessed these facts for yourself. While it is unlikely you have ever had the misfortune of seeing a falling elevator you have seen astronauts at play inside their space capsules while in orbit (i.e., free fall) about the earth. The objects inside the elevator (capsule) seem then to constitute an archetypical inertial frame (they satisfy Newton's First Law). By establishing spatial and temporal coordinate systems in the usual way our observer thereby becomes an inertial observer, at least within the spatial and temporal constraints imposed by his circumstances. Now picture an arbitrary event. There are any number of vantage points from which the event can be observed. One is from a freely falling elevator in the immediate spatial and temporal vicinity of the event and from this vantage point the event receives *inertial* coordinates. There is then a *local inertial frame* near any event in X.

The operative word is "local." The "spatial and temporal constraints" to which we alluded in the preceding paragraph arise from the nonuniformity of any gravitational field in the real world. For example, in an elevator which falls freely in the earth's gravitational field, all of the objects inside are pulled toward the earth's *center* so that these objects do experience some slight relative motion (toward each other). Such motion, of course, goes unnoticed if the elevator falls neither too far nor too long. Indeed, by restricting our observer to a sufficiently small region in space and time these effects become negligible and the observer is indeed inertial. But then, what is "negligible" is in the eye of the beholder. The availability of more sensitive measuring devices will require further restrictions on the size of the spacetime region which "looks like" ^{1,3}. Turn of the century mathematical terminology expressed this fact by saying that any point in X has about it an "infinitesimal neighborhood" which is identical to ^{1,3}. Today we prefer to say that X is a 4-dimensional smooth manifold, each tangent space of which has the structure of ^{1,3}.

A **spacetime** is a 4-dimensional smooth manifold X with a semi-Riemannian metric $\boldsymbol{g}$ of index 3 (called a **Lorentz metric**). Thus, for each $x \in X$ there exists a frame $p = (\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ at x such that

$$\boldsymbol{g}_x(\boldsymbol{b}_\alpha, \boldsymbol{b}_\beta) = \eta_{\alpha\beta} = \begin{cases} 1, & \alpha = \beta = 0 \\ -1, & \alpha = \beta = 1, 2, 3. \\ 0, & \alpha \neq \beta \end{cases} \tag{3.5.1}$$

For simplicity we will occasionally write $\boldsymbol{g}_x$ as $\langle\ ,\ \rangle_x$, or even $\langle\ ,\ \rangle$. A tangent vector $\boldsymbol{v}$ at x is said to be **spacelike**, **timelike**, or **null** if $\langle \boldsymbol{v}, \boldsymbol{v} \rangle_x$ is $< 0, > 0$, or $= 0$, respectively. The **null cone at** x is the set of all null tangent vectors at x. A smooth curve in X is **spacelike**, **timelike**, or **null** provided its tangent vector is spacelike, timelike, or null, respectively, at each point.

Remark: Although we have thus far thought of Minkowski spacetime ^{1,3} as a vector space, this is a luxury we have been permitted because its structure as a spacetime manifold is so simple. The underlying 4-manifold is, of course, ^{4} so that, in particular, all of its tangent spaces are canonically identified with ^{4} itself. Thus, defining a constant Lorentz metric $\boldsymbol{g}$ on ^{1,3} by specifying that its components relative to the standard global chart are

$g_{\alpha\beta}(x) = \boldsymbol{g}_x(\frac{\partial}{\partial x^\alpha}|_x, \frac{\partial}{\partial x^\beta}|_x) = \eta_{\alpha\beta}$ for each $x \in \ ^{1,3}$ is tantamount to defining the Minkowski inner product on $\ ^4$. We shall feel free to think of Minkowski spacetime in whichever of these ways is convenient at the moment.

The most serious obstacle to the general study of spacetime manifolds is their overwhelming number and diversity. Smooth 4-manifolds are, to say the least, plentiful and almost all of them admit Lorentz metrics. Indeed, a Lorentz metric can be defined on *any* noncompact 4-manifold and a compact 4-manifold admits a Lorentz metric if and only if its Euler characteristic (see Section 5.7 or Section 3.4 of [**N4**]) is zero (see [**O'N**]).

Remark: Compact spacetimes are of no real interest anyway since they always contain closed timelike curves and these do violence to our most cherished notions of causality (see [**N2**]). In effect then, anything that could possibly be a spacetime, *is* a spacetime.

There are two ways around this difficulty. One can impose additional restrictions on the structure of a spacetime in an attempt to eliminate "unphysical" behavior. A number of such restrictions are easily formulated (we will introduce one shortly) and the study of the manifolds that satisfy these conditions has led to some spectacular results (e.g., the Singularity Theorems of Penrose and Hawking discussed in [**HE**], [**Pen**], [**O'N**] and [**N2**]). Alternatively, one can restrict attention to spacetime manifolds that arise in physics by solving the field equations of general relativity for more or less realistic distributions of mass/energy. We will describe a few such examples shortly, but our real interest here (and in Chapter 6) is in an essentially global, topological question about spacetime manifolds. The question (introduced in Section 2.4) is this: On which spacetime manifolds is it possible to introduce a meaningful notion of a "spin $\frac{1}{2}$ particle"?

According to Dirac a spin $\frac{1}{2}$ particle is characterized by the fact that its wavefunction transforms under a certain (reducible) representation of the double cover $SL(2, \ )$ of $\mathcal{L}_+^\uparrow$. Now, each tangent space $T_x(X)$ to a spacetime manifold X has an inner product $\langle\ ,\ \rangle_x$ of index 3 giving it the structure of Minkowski spacetime. In particular, its orthonormal bases are related by elements of $\mathcal{L}$. We are, however, interested only in bases related by elements of $\mathcal{L}_+^\uparrow$. We must therefore choose, consistently over all of X, a family of bases for each $T_x(X)$ that are related by the $\Lambda \in \mathcal{L}$ satisfying $\det \Lambda = 1$ and $\Lambda^0{}_0 \geq 1$. This is possible only if X is orientable and has, moreover, some global notion of "time orientation" (see Exercise 3.4.2). To define such a notion we proceed in the manner familiar for smooth surfaces in $\ ^3$ (which are orientable if and only if they admit a smooth, nonzero field of normal vectors).

We say that a spacetime X is **time orientable** if one can define on it a smooth vector field $\boldsymbol{T}$ which is everywhere timelike ($\langle \boldsymbol{T}(x), \boldsymbol{T}(x)\rangle_x > 0$ for each $x \in X$). X is **time oriented** if a specific choice of such a vector field has been made. One then thinks of $\boldsymbol{T}$ as "pointing toward the future" (keeping in mind that the designation "future" is now entirely arbitrary and would be reversed if we choose $-\boldsymbol{T}$ rather than $\boldsymbol{T}$ to time orient X). Motivated by Theorem 3.4.1 we then say that a tangent vector $\boldsymbol{v} \in T_x(X)$ which is either timelike, or null and nonzero, is **future directed** if $\langle \boldsymbol{T}(x), \boldsymbol{v} \rangle_x > 0$ and **past directed** if $\langle \boldsymbol{T}(x), \boldsymbol{v} \rangle_x < 0$. The same terminology is used for smooth curves that are either timelike or null (and, in the latter case, have nonzero tangent vector at each point).

If X is an oriented, time oriented spacetime, then an **oriented**, **time oriented**, **orthonormal frame** at $x \in X$ is a frame $\{\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ in $T_x(X)$ that is consistent with the orientation of X, satisfies $\langle \boldsymbol{b}_\alpha, \boldsymbol{b}_\beta \rangle = \eta_{\alpha\beta}$ and for which the timelike vector $\boldsymbol{b}_0$ is future directed. Exercise 3.3.10 guarantees the existence of local oriented, orthonormal frame fields on any oriented semi-Riemannian manifold so, in particular, each $x_0 \in X$ has about it a connected open neighborhood U on which there exist smooth vector fields $\boldsymbol{E}_0, \boldsymbol{E}_1, \boldsymbol{E}_2$ and $\boldsymbol{E}_3$ with $\{\boldsymbol{E}_0(x), \boldsymbol{E}_1(x), \boldsymbol{E}_2(x), \boldsymbol{E}_3(x)\}$ an oriented orthonormal frame for each $x \in U$. Now, $\boldsymbol{E}_0(x)$ is timelike for every $x \in U$ and, if $\boldsymbol{T}$ is the vector field that time orients X, $\langle \boldsymbol{E}_0(x), \boldsymbol{T}(x) \rangle$ is never zero (Exercise 3.4.1) and so has the same sign everywhere on U. If this sign is positive, $\{\boldsymbol{E}_0, \boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3\}$ is a local oriented, time oriented, orthonormal frame field on U. If the sign is negative, $\{-\boldsymbol{E}_0, \boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3\}$ is a local oriented, time oriented, orthonormal frame field on U. In any case, such things always exist and the construction of $SO(k, n-k) \hookrightarrow F_+(X) \xrightarrow{\mathcal{P}_{F_+}} X$ in Exercise 3.3.17 can be repeated verbatim to yield the **oriented**, **time oriented**, **orthonormal frame bundle**

$$\mathcal{L}_+^\uparrow \hookrightarrow \mathcal{L}(X) \xrightarrow{\mathcal{P}_\mathcal{L}} X$$

with group $\mathcal{L}_+^\uparrow$. As we argued in Section 2.4 (pages 114–115) the object required to describe wavefunctions of spin$\frac{1}{2}$ particles on X is a lift of this bundle to a principal $SL(2, \ )$-bundle over X. More precisely, a **spinor structure** for X consists of a principal $SL(2, \ )$-bundle

$$SL(2, \ ) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X$$

over X and a map $\lambda : S(X) \longrightarrow \mathcal{L}(X)$ such that $\mathcal{P}_\mathcal{L}(\lambda(p)) = \mathcal{P}_S(p)$ and $\lambda(p \cdot g) = \lambda(p) \cdot \text{Spin}(g)$ for all $g \in SL(2, \ )$. In Section 6.5 we will determine a necessary and sufficient condition (on the topology of X) for the existence of such a spinor structure.

We will conclude this section with a number of examples of spacetime manifolds. We have been guided in the selection of these examples by the desire to illustrate the concepts we have introduced in a context as free of technical obfuscations as possible without wandering into the realm of physically meaningless examples contrived solely for pedegogical purposes. Some of the examples are of great physical significance, while others are primarily of historical interest, but all of them have played a role in the development of general relativity.

As we have already mentioned, the simplest example of a spacetime is ^{1,3} itself. Here the underlying 4-manifold is ^{4}. The standard coordinate functions (corresponding to the chart $id_{\ 4} : {}^{4} \longrightarrow {}^{1,3}$) are denoted x^0, x^1, x^2 and x^3. We define the Lorentz metric $\boldsymbol{g}$ on ^{1,3} by specifying its components $g_{\alpha\beta}$ relative to this global chart. For any $p \in {}^{1,3}$, $T_p({}^{1,3})$ is spanned by $\{\frac{\partial}{\partial x^0}|_p, \frac{\partial}{\partial x^1}|_p, \frac{\partial}{\partial x^2}|_p, \frac{\partial}{\partial x^3}|_p\}$ and we define

$$g_{\alpha\beta}(p) = \boldsymbol{g}_p\left(\left.\frac{\partial}{\partial x^\alpha}\right|_p, \left.\frac{\partial}{\partial x^\beta}\right|_p\right) = \eta_{\alpha\beta}, \quad \alpha, \beta = 0, 1, 2, 3,$$

where

$$\eta_{\alpha\beta} = \begin{cases} 1, & \alpha = \beta = 0 \\ -1, & \alpha = \beta = 1, 2, 3\,. \\ 0, & \alpha \neq \beta \end{cases}$$

Thus, the $g_{\alpha\beta}$ are constant and, identifying each $T_p({}^{1,3})$ with ^{4} via the canonical isomorphism, we have, in effect, just introduced the Minkowski inner product on ^{4}. Relative to the basis $\{dx^\alpha \otimes dx^\beta : \alpha, \beta = 0, 1, 2, 3\}$ for the covariant tensors of rank 2 on ^{1,3}, the metric for ^{1,3} is given by

$$\boldsymbol{g} = \eta_{\alpha\beta}\, dx^\alpha \otimes dx^\beta = dx^0 \otimes dx^0 - \sum_{i=1}^{3} dx^i \otimes dx^i.$$

To ease the typography we will often write the coordinate velocity vector fields $\frac{\partial}{\partial x^\alpha}$ as ∂_α, $\alpha = 0, 1, 2, 3$. The standard orientation for ^{1,3} is the one that assigns to each $p \in {}^{1,3}$ the orientation for $T_p({}^{1,3})$ containing $\{\partial_0(p), \partial_1(p), \partial_2(p), \partial_3(p)\}$ and we will time orient ^{1,3} with the vector field ∂_0. Thus, a tangent vector $\boldsymbol{v} = v^\alpha \partial_\alpha(p) \in T_p({}^{1,3})$ which is either timelike ($\boldsymbol{g}_p(\boldsymbol{v}, \boldsymbol{v}) = (v^0)^2 - (v^1)^2 - (v^2)^2 - (v^3)^2 > 0$) or null ($\boldsymbol{g}_p(\boldsymbol{v}, \boldsymbol{v}) = (v^0)^2 - (v^1)^2 - (v^2)^2 - (v^3)^2 = 0$) and nonzero is future directed if $\boldsymbol{g}_p(\partial_0(p), \boldsymbol{v}) = v^0$ is positive. Since any principal bundle over ^{4} is necessarily trivial, this is true, in particular, for the oriented, time oriented, orthonormal frame bundle of ^{1,3}.

Exercise 3.5.1 Introduce spherical coordinates (t, ρ, ϕ, θ) on ^{1,3} as follows: Define a map from

$$(t,\ \rho,\ \phi,\ \theta) \in (-\infty,\ \infty) \times (0,\ \infty) \times (0,\ \pi) \times (-\pi,\ \pi)$$

to ^{1,3} by

$$\begin{aligned} x^0 &= t \\ x^1 &= \rho \sin\phi \cos\theta \\ x^2 &= \rho \sin\phi \sin\theta \\ x^3 &= \rho \cos\phi. \end{aligned}$$

Show that the Jacobian of the map is $\rho^2 \sin\phi$. Use this fact to show that the map is a diffeomorphism onto an open set in ^{1,3}. What open set? The inverse of this map is therefore a chart on ^{1,3}. Denote the coordinate velocity fields for this chart $\partial_t, \partial_\rho, \partial_\phi$ and ∂_θ and compute the components of the metric in this chart (i.e., $\boldsymbol{g}(\partial_t, \partial_t)$, $\boldsymbol{g}(\partial_t, \partial_\rho)$, etc.) to show that

$$\boldsymbol{g} = dt \otimes dt - d\rho \otimes d\rho - \rho^2 (d\phi \otimes d\phi + \sin^2\phi \; d\theta \otimes d\theta).$$

Finally, alter the domain of the mapping in such a way as to obtain charts that cover as much of ^{1,3} as possible.

Exercise 3.5.2 Introduce advanced and retarded null coordinates v and w on ^{1,3} by letting $v = t + \rho$ and $w = t - \rho$ (so that $v \geq w$). Thus,

$$t = \frac{1}{2}(v + w)$$

$$\rho = \frac{1}{2}(v - w)$$

$$\phi = \phi$$

$$\theta = \theta.$$

Show that the Jacobian of the map is identically equal to $\frac{1}{2}$ so that the transformation $(v, w, \phi, \theta) \longrightarrow (t, \rho, \phi, \theta)$ is nonsingular wherever it is defined. Show that, in these coordinates, the metric is given by

$$\boldsymbol{g} = dv \otimes dw - \frac{1}{4}(v - w)^2 (d\phi \otimes d\phi + \sin^2\phi \; d\theta \otimes d\theta).$$

In particular, $\boldsymbol{g}(\partial_v,\ \partial_v) = \boldsymbol{g}(\partial_w,\ \partial_w) = 0$ so v and w are null coordinates. Convince yourself that the following describes a physical procedure for obtaining the coordinates $v(p)$ and $w(p)$ for an event $p \in {}^{1,3}$: Find a light ray that is "incoming" to the origin and experiences p and take $v(p)$ to be the arrival time t of the ray at the origin. Next, find a light ray that is "outgoing" from the origin and experiences p and take $w(p)$ to be the departure time t of the ray from the origin.

Our next example is still topologically trivial, but is geometrically much more interesting. The **Einstein-deSitter spacetime** $\mathcal{E}$ is the simplest of all the cosmological models to emerge from general relativity and is defined as follows: As a smooth manifold, $\mathcal{E}$ is just the open submanifold $(0, \infty) \times {}^3$ of ^{4}. The restriction of the standard chart on ^{4} to $(0, \infty) \times {}^3$ gives global coordinate functions on $\mathcal{E}$ that we will continue to denote (x^0, x^1, x^2, x^3). Thus, at each $p = (x^0, x^1, x^2, x^3) \in \mathcal{E}$ the tangent space $T_p(\mathcal{E})$ is spanned by $\{\partial_0(p), \partial_1(p), \partial_2(p), \partial_3(p)\}$ which, as for ^{1,3}, we take to be an oriented basis. We will define the Lorentz metric $\boldsymbol{g}$ for $\mathcal{E}$ by giving its components $g_{\alpha\beta}$ relative to this global chart. Specifically, at each $p = (x^0, x^1, x^2, x^3) \in \mathcal{E}$,

$$g_{\alpha\beta}(p) = g_{\alpha\beta}(x^0, x^1, x^2, x^3) = \begin{cases} \quad 1 \quad , & \alpha = \beta = 0 \\ -(x^0)^{4/3}, & \alpha = \beta = 1, 2, 3. \\ \quad 0 \quad , & \alpha \neq \beta \end{cases} \tag{3.5.2}$$

Thus,

$$\boldsymbol{g} = dx^0 \otimes dx^0 - (x^0)^{4/3} \sum_{i=1}^{3} dx^i \otimes dx^i \tag{3.5.3}$$

so, if $\boldsymbol{v} = v^\alpha \partial_\alpha(p)$ and $\boldsymbol{w} = w^\alpha \partial_\alpha(p)$ are in $T_p(\mathcal{E})$, we have

$$\boldsymbol{g}_p(\boldsymbol{v}, \boldsymbol{w}) = v^0 w^0 - (x^0)^{4/3}\ (v^1 w^1 + v^2 w^2 + v^3 w^3), \tag{3.5.4}$$

where x^0 is the "time coordinate" of p (the "height" of p if we picture the x^0-axis in $(0, \infty) \times \ ^3$ vertically, as we generally do). One must show, of course, that (3.5.3) does indeed define a Lorentz metric on $\mathcal{E}$. Since smoothness of the component functions (3.5.2) is clear on $x^0 > 0$ we need only produce a frame $(\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ at each $p \in \mathcal{E}$ such that (3.5.1) is satisfied. But for this one need only define

$$\begin{aligned} \boldsymbol{b}_0 &= \partial_0(p) \\ \boldsymbol{b}_i &= (x^0)^{-2/3}\ \partial_i(p), \quad i = 1, 2, 3. \end{aligned}$$

Observe that the vector field ∂_0 on $\mathcal{E}$ is everywhere timelike and so may be used to time orient $\mathcal{E}$. At each $p = (x^0, x^1, x^2, x^3) \in \mathcal{E}$ the null cone consists of all $\boldsymbol{v} = v^\alpha \partial_\alpha(p)$ satisfying

$$(v^1)^2 + (v^2)^2 + (v^3)^2 = (x^0)^{-4/3}(v^0)^2,$$

which one can interpret geometrically as saying that the null cones in $\mathcal{E}$ "get steeper" as p "gets higher," i.e., as time goes on. Since $\mathcal{E}$ is diffeomorphic to $\ ^4$ its oriented, time oriented, orthonormal frame bundle is trivial.

Remark: The physical significance of the Einstein-deSitter spacetime as a cosmological model is discussed at great length in [**SaW**]. This is not our concern here so we will content ourselves with a few observations on the proper way to view $\mathcal{E}$ without any real attempt at justification.

The vertical straight lines $\alpha(t) = (t, x_0^1, x_0^2, x_0^3)$, where $0 < t < \infty$ and $(x_0^1, x_0^2, x_0^3) \in \ ^3$ is fixed, are clearly future directed, timelike curves in $\mathcal{E}$. These are to be interpreted as the worldlines of the galaxy clusters of our universe. The "displacement vector" $\boldsymbol{v}$ between two events with the same x^0 on two different vertical worldlines is spacelike and satisfies $\|\boldsymbol{v}\| = (-\langle \boldsymbol{v}, \boldsymbol{v} \rangle)^{\frac{1}{2}} = K(x^0)^{2/3}$, where K is a positive constant (here we have in mind two "nearby" worldlines so that $\boldsymbol{v}$, which actually lies in a tangent space to $\mathcal{E}$, can be thought of as "joining" the worldlines). Observe that $\|\boldsymbol{v}\|$, which is regarded as the distance between the two galaxy clusters at the "instant" x^0 in the given global coordinate system, satisfies $\frac{d}{dx^0}\|\boldsymbol{v}\| = \frac{2}{3}K(u^0)^{-1/3} > 0$ and $\frac{d^2}{(dx^0)^2}\|\boldsymbol{v}\| = -\frac{2}{9}K(x^0)^{-4/3} < 0$ so that these clusters are receding from each other (the universe is expanding), but at a decreasing rate. Finally, we remark that one can define, on any semi-Riemannian manifold, a real-valued function S called the "scalar curvature" which is regarded as a gross numerical measure of the extent to which the manifold is "curved" at each point. In spacetime manifolds this measures the overall strength of the gravitational field at each point. In $\mathcal{E}$ the scalar curvature is given by $S = \frac{4}{3}(x^0)^{-2}$ which becomes unbounded as $x^0 \longrightarrow 0$, i.e., as one recedes into the past. Moving backward in time along the worldlines of the galaxies one approaches the "missing" 3-space $x^0 = 0$ where the curvature would have to be infinite. This is the mathematical representation in $\mathcal{E}$ of the "big bang" (notice that it is *not* a point).

Exercise 3.5.3 Show that the cubic curve $\lambda(t) = (t^{3/5}, 3t^{1/5}, 0, 0)$, $0 < t < \infty$ is a future directed null curve in $\mathcal{E}$. **Note**: This is actually what is called a null "geodesic" in $\mathcal{E}$ and, from it, one can obtain all the null geodesics of $\mathcal{E}$; see pages 133 and 162 of [**SaW**].

Our first topologically nontrivial example is called **de Sitter spacetime**, denoted $\mathcal{D}$ and defined as follows: Consider the 5-dimensional Minkowski space ^{1,4}. As we did for Minkowski spacetime we now regard ^{1,4} as a semi-Riemannian manifold with constant metric $\tilde{\boldsymbol{g}}$ given, in standard coordinates, by

$$\tilde{\boldsymbol{g}} = dx^0 \otimes dx^0 - \sum_{i=1}^{4} dx^i \otimes dx^i.$$

The map $Q : {}^{1,4} \longrightarrow \quad$ given by $Q(x) = \langle x, x\rangle = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 - (x^4)^2$ is obviously smooth and has -1 as a regular value so $Q^{-1}(-1)$ is a 4-dimensional smooth submanifold of ^{1,4}. As a manifold, $\mathcal{D}$ is just this smooth submanifold of ^{1,4}, i.e.,

$$\mathcal{D} = \Big\{ (x^0, x^1, x^2, x^3, x^4) \in {}^{1,4} : \\ (x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 - (x^0)^2 = 1 \Big\}.$$

By suppressing two spatial dimensions one can visualize $\mathcal{D}$ as a hyperboloid of one sheet (keeping in mind that the cross-sectional "circles" at constant x^0 are really S^3).

Lemma 3.5.1 *$\mathcal{D}$ is diffeomorphic to* $\quad \times S^3$.

Proof: The maps $f : \quad \times S^3 \longrightarrow \mathcal{D}$ and $g : \mathcal{D} \longrightarrow \quad \times S^3$ defined by $f(t, y^1, y^2, y^3, y^4) = \left(t, (1+t^2)^{\frac{1}{2}} y^1, (1+t^2)^{\frac{1}{2}} y^2, (1+t^2)^{\frac{1}{2}} y^3, (1+t^2)^{\frac{1}{2}} y^4\right)$ and

$$g(x^0, x^1, x^2, x^3, x^4) = \Big(x^0, (1+(x^0)^2)^{-\frac{1}{2}} x^1, (1+(x^0)^2)^{-\frac{1}{2}} x^2,$$
$$(1+(x^0)^2)^{-\frac{1}{2}} x^3,\ (1+(x^0)^2)^{-\frac{1}{2}} x^4\Big)$$

are inverse diffeomorphisms. ■

To obtain a Lorentz metric $\boldsymbol{g}$ on $\mathcal{D}$ we will identify the tangent spaces $T_p(\mathcal{D})$, $p \in \mathcal{D}$, with subspaces of $T_p(\ ^{1,4})$ and restrict $\tilde{\boldsymbol{g}}$ to these subspaces (more precisely, $\boldsymbol{g} = \iota^* \tilde{\boldsymbol{g}}$, where $\iota : \mathcal{D} \hookrightarrow \ ^{1,4}$ is the inclusion). This restriction $\boldsymbol{g}$ is clearly symmetric and bilinear at each point, but it is not obviously nondegenerate nor is it clear that it has index 3. To prove this we introduce a vector field grad Q on $\ ^{1,4}$ defined, relative to standard coordinates, by

$$\operatorname{grad} Q(p) = 2x^0 \partial_0(p) - 2x^1 \partial_1(p) - 2x^2 \partial_2(p) - 2x^3 \partial_3(p) - 2x^4 \partial_4(p)$$

for each $p = (x^0, x^1, x^2, x^3, x^4) \in \ ^{1,4}$.

Exercise 3.5.4 Show that, at each $p \in \mathcal{D}, \tilde{\boldsymbol{g}}_p(\operatorname{grad} Q(p),\ \operatorname{grad} Q(p)) = -4$ and $\tilde{\boldsymbol{g}}_p(\operatorname{grad} Q(p), \boldsymbol{v}) = 0$ for each $\boldsymbol{v} \in T_p(\mathcal{D}) \subseteq T_p(\ ^{1,4})$.

It follows from Exercise 3.5.4 that the restriction of $\tilde{\boldsymbol{g}}$ to each $T_p(\mathcal{D})$ is nondegenerate (if there were a vector in $T_p(\mathcal{D})$ orthogonal to every other vector in $T_p(\mathcal{D})$, then, being orthogonal also to grad $Q(p)$, it would be orthogonal to everything in $T_p(\ ^{1,4})$). Thus, $\boldsymbol{g}$ is a metric on $\mathcal{D}$ and we need only show that it has index 3. But $\tilde{\boldsymbol{g}}$ has index 4 and, at any $p \in \mathcal{D}$, $\boldsymbol{U}(p) = \frac{1}{4} \operatorname{grad} Q(p)$ is orthogonal to $T_p(\mathcal{D})$ and satisfies $\tilde{\boldsymbol{g}}_p(\boldsymbol{U}(p), \boldsymbol{U}(p)) = -1$ so this is clear (an orthonormal basis for $T_p(\mathcal{D})$ together with $\boldsymbol{U}(p)$ is an orthonormal basis for $T_p(\ ^{1,4})$). The existence of the globally defined unit normal field $\boldsymbol{U}$ on $\mathcal{D}$ also provides $\mathcal{D}$ with the orientation and time orientation we require for $\mathcal{D}$.

Exercise 3.5.5 Show that the following defines an orientation μ on $\mathcal{D}$: At each $p \in \mathcal{D}$, an ordered basis $\{\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ for $T_p(\mathcal{D})$ is in μ_p if and only if $\{\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \boldsymbol{U}(p)\}$ is in the standard orientation for $T_p(\ ^{1,4}) = T_p(\ ^5)$.

Exercise 3.5.6 For each $p \in \mathcal{D}$ define

$$\boldsymbol{V}(p) = \partial_0(p) + \tilde{\boldsymbol{g}}(\boldsymbol{U}(p), \partial_0(p))\, \boldsymbol{U}(p).$$

Show that $\boldsymbol{V}(p) \in T_p(\mathcal{D})$ and $\boldsymbol{g}(\boldsymbol{V}(p), \boldsymbol{V}(p)) > 0$. Conclude that $\mathcal{D}$ is time oriented by $\boldsymbol{V} = \partial_0 + \tilde{\boldsymbol{g}}(\boldsymbol{U}, \partial_0)\boldsymbol{U}$.

Exercise 3.5.7 Use the oriented, orthonormal frame field $\{\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3\}$ for S^3 constructed in Section 3.3 (see (3.3.9)) and the timelike, future directed vector field $\boldsymbol{V}$ on $\mathcal{D}$ constructed in Exercise 3.5.6 to show that the oriented, time oriented, orthonormal frame bundle of $\mathcal{D}$ is trivial.

We introduce coordinates ξ, ϕ, θ, t on $\mathcal{D}$ as follows: Define a map τ from ^{4} into ^{1,4} by

$$\begin{aligned} x^0 &= \sinh t \\ x^1 &= \sin\xi \sin\phi \cos\theta \cosh t \\ x^2 &= \sin\xi \sin\phi \sin\theta \cosh t \\ x^3 &= \sin\xi \cos\phi \cosh t \\ x^4 &= \cos\xi \cosh t. \end{aligned}$$

Then $(x^1)^2+(x^2)^2+(x^3)^2+(x^4)^2-(x^0)^2 = 1$ so the image of τ is in $\mathcal{D}$. For each fixed t in $(-\infty,\infty), x^1, x^2, x^3$ and x^4 parametrize the 3-sphere of radius $\cosh t$ and covers the entire 3-sphere for ξ, ϕ and θ restricted to $0 \le \xi \le \pi, 0 \le \phi \le \pi$ and $0 \le \theta \le 2\pi$. With these values of ξ, ϕ and θ and $-\infty < t < \infty, \tau$ maps onto $\mathcal{D}$.

Exercise 3.5.8 Compute the 5×4 Jacobian matrix of τ and show that it has rank 4 on each of the following regions in ^{4}.

$$\begin{array}{llll} 0 < \xi < \pi, & 0 < \phi < \pi, & 0 < \theta < 2\pi, & -\infty < t < \infty \\ 0 < \xi < \pi, & 0 < \phi < \pi, & -\pi < \theta < \pi, & -\infty < t < \infty \end{array}$$

Show also that τ is one-to-one when restricted to either of these regions and conclude that the inverse of each of these restrictions is a chart on $\mathcal{D}$.

Exercise 3.5.9 Show that, relative to the coordinates ξ, ϕ, θ, t on $\mathcal{D}$, the metric $\boldsymbol{g}$ takes the form

$$\boldsymbol{g} = dt \otimes dt - \cosh^2 t \left(d\xi \otimes d\xi + \sin^2 \xi (d\phi \otimes d\phi + \sin^2 \phi d\theta \otimes d\theta) \right).$$

Remarks: The deSitter spacetime can be regarded as an empty space solution to the Einstein field equations with nonzero cosmological constant (see Section 5.2 of [**HE**]). In this universe the spatial cross-sections (t = constant) are 3-spheres whose radii vary with time (at time t the radius is $\cosh t$). Our final example of a spacetime describes an analogous "static" universe. Since the details are virtually identical to those above for $\mathcal{D}$ we will leave them to the reader as exercises.

The **Einstein cylinder** spacetime $\mathcal{C}$ is the submanifold of ^{1,4} consisting of all $(x^0, x^1, x^2, x^3, x^4)$ satisfying $(x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 = 1$ with the Lorentz metric $\boldsymbol{g}$ obtained by restricting the metric $\tilde{\boldsymbol{g}}$ of ^{1,4} to this submanifold.

Exercise 3.5.10 Show that, thus defined, $\mathcal{C}$ is, indeed, a spacetime manifold.

Exercise 3.5.11 Show that $\mathcal{C}$ is diffeomorphic to $\times S^3$.

Exercise 3.5.12 Show that $\mathcal{C}$ is orientable and time orientable and that its oriented, time oriented, orthonormal frame bundle is trivial.

Exercise 3.5.13 Show that one can introduce coordinates $\xi, \phi, \theta,\ t$ on $\mathcal{C}$ by

$$\begin{aligned} x^0 &= t \\ x^1 &= \sin\xi \sin\,\phi \cos\,\theta \\ x^2 &= \sin\xi \sin\,\phi \sin\,\theta \\ x^3 &= \sin\xi \cos\,\phi \\ x^4 &= \cos\,\xi \end{aligned}$$

(same restrictions as in Exercise 3.5.8) and that, relative to these coordinates,

$$\boldsymbol{g} = dt \otimes dt - \Big(d\xi \otimes d\xi + \sin^2 \xi (d\phi \otimes d\phi + \sin^2 \phi d\theta \otimes d\theta) \Big).$$

Exercise 3.5.14 Find a few spacelike, timelike and null curves in $\mathcal{C}$.

4
Differential Forms and Integration

Introduction

Physics is expressed in the language of differential equations (e.g., Maxwell, Dirac, Yang-Mills, Einstein, etc.). Differential equations live on differentiable manifolds and differentiable manifolds have topologies that influence not only the solutions to differential equations defined on them, but even the type of equation that one can define on them. At some naive level then it is perhaps not surprising that topology and physics interact. The profound depth of this interaction in recent years, however, has made it abundantly clear that the naive level is not the appropriate one from which to view this.

It is remarkable that the deep connection between topology and the differential equations of physics can be made quite explicit. The bridge between the two subjects is the notion of an *elliptic complex* of differential operators on a manifold and its corresponding cohomology. There is, in fact, an explicit formula (the *Atiyah-Singer Index Theorem*) relating the analytic properties of the differential operators in such a complex to the topology of the underlying manifold. This is, unfortunately, quite beyond our level here, although the object will arise again in the Appendix. We will, however, in this chapter construct the simplest example of an elliptic complex (the de Rham complex) and, in the next, study its cohomology. In Chapter 6 we will use the information thus accumulated to construct the characteristic classes that have, at least informally, put in an appearance in our earlier discussions of electromagnetic and Yang-Mills fields (Chapter 2).

4.1 Multilinear Algebra

In this section we assemble the algebraic machinery required to build the objects of interest to us. Since much of this is a straightforward generalization of results we already have available for 0-, 1- and 2-forms we intend to leave many of the routine verifications to the reader in the form of exercises.

Let E denote a real vector space and $k \geq 0$ an integer (more generally, one can take E to be a module over a commutative ring with identity). A map A from $E \times \cdots \times E$ (k factors) to   is ***k*-multilinear** if, for each i with $1 \leq i \leq k$ and each $a \in$,

$$A(v_1, \ldots, v_i + v_i', \ldots, v_k) = A(v_1, \ldots, v_i, \ldots, v_k) + A(v_1, \ldots, v_i', \ldots, v_k)$$

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0_4,

and

$$A(v_1, \ldots, av_i, \ldots, v_k) = aA(v_1, \ldots, v_i, \ldots, v_k)$$

for all $v_1, \ldots, v_i, v_i', \ldots, v_k$ in E. The set $\mathcal{T}^k(E)$ of all such multilinear forms is a real vector space with pointwise operations:

$$(A + B)(v_1, \ldots, v_k) = A(v_1, \ldots, v_k) + B(v_1, \ldots, v_k)$$

$$(aA)(v_1, \ldots, v_k) = a(A(v_1, \ldots, v_k)).$$

Note that $\mathcal{T}^1(E)$ is the dual space E^*. For convenience we will take $\mathcal{T}^0(E) =$. The elements of $\mathcal{T}^k(E)$ are called **covariant tensors** of **rank** k (or simply ***k*-tensors**) on E. If $T : E_1 \longrightarrow E_2$ is a linear transformation we define the **pullback** map

$$T^* : \mathcal{T}^k(E_2) \longrightarrow \mathcal{T}^k(E_1)$$

for any k as follows: If $A \in \mathcal{T}^k(E_2)$, the $T^*A \in \mathcal{T}^k(E_1)$ is given by $(T^*A)(v_1, \ldots, v_k) = A(T(v_1), \ldots, T(v_k))$.

Exercise 4.1.1 Show that if $A \in \mathcal{T}^k(E)$, then the following three conditions are equivalent:

(a) A is zero whenever two of its arguments are equal, i.e., if $1 \leq i, j \leq k$ and $i \neq j$, then $A(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = 0$ whenever $v_i = v_j$.

(b) A changes sign whenever two of its arguments are interchanged (and the remaining arguments are left fixed), i.e., $A(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) = -A(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k)$.

(c) If $\sigma \in S_k$ is any permutation of $\{1, \ldots, k\}$ and $(-1)^\sigma$ is its sign (1 if σ is an even permutation and -1 if σ is an odd permutation), then $A(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = (-1)^\sigma A(v_1, \ldots, v_k)$.

An $A \in \mathcal{T}^k(E)$ that satisfies any (and therefore all) of the conditions in Exercise 4.1.1 is said to be **skew-symmetric** and the set $\Lambda^k(E)$ of all such is a linear subspace of $\mathcal{T}^k(E)$. Note that $\Lambda^1(E) = \mathcal{T}^1(E) = E^*$ since the conditions are satisfied vacuously. For convenience, we take $\Lambda^0(E) = \mathcal{T}^0(E) =$. The elements of $\Lambda^k(E)$ are called ***k*-forms** on E. An $A \in \mathcal{T}^k(E)$ which takes the same value whenever two of its arguments are interchanged (or, equivalently, whenever its arguments are permuted) is said to be **symmetric** and the collection of all such is likewise a linear subspace of $\mathcal{T}^k(E)$.

For any $A \in \mathcal{T}^k(E)$ and $B \in \mathcal{T}^l(E)$ we define the **tensor product** $A \otimes B \in \mathcal{T}^{k+l}(E)$ by

$$(A \otimes B)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}) = A(v_1, \ldots, v_k)\ B(v_{k+1}, \ldots, v_{k+l}).$$

Exercise 4.1.2 Show that $A \otimes B$ is, indeed, in $\mathcal{T}^{k+l}(E)$ and prove each of the following properties of the tensor product.

(a) $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B$

(b) $A \otimes (B_1 + B_2) = A \otimes B_1 + A \otimes B_2$

(c) $(aA) \otimes B = A \otimes (aB) = a(A \otimes B) \quad (a \in \quad)$

(d) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$

(e) $T^*(A \otimes B) = (T^*A) \otimes (T^*B)$.

Also, find an example to show that $B \otimes A$ is generally *not* equal to $A \otimes B$.

Because the tensor product is associative (Exercise 4.1.2 (d)) one can unambiguously define the tensor product $A_1 \otimes \cdots \otimes A_m$ of any finite number of tensors on E.

Lemma 4.1.1 *Let* $\{e_1, \ldots, e_n\}$ *be a basis for* E *and* $\{e^1, \ldots, e^n\}$ *its dual basis for* $E^*(e^i(e_j) = \delta^i{}_j)$. *Then, for any* k, $\{e^{i_1} \otimes \cdots \otimes e^{i_k} : 1 \leq i_1, \ldots, i_k \leq k\}$ *is a basis for* $\mathcal{T}^k(E)$. *Moreover, any* $A \in \mathcal{T}^k(E)$ *can be uniquely written as*

$$A = A_{i_1 \cdots i_k} \, e^{i_1} \otimes \cdots \otimes e^{i_k} \qquad \text{(summation convention)}$$

where $A_{i_1 \cdots i_k} = A(e_{i_1}, \ldots, e_{i_k})$. *In particular,* $\dim \mathcal{T}^k(E) = n^k$.

Exercise 4.1.3 Prove Lemma 4.1.1. **Hint**: The $n = 2$ case is Lemma 5.11.1, [**N4**].

In particular, any $\alpha \in \Lambda^k(E)$ can be expanded as in Lemma 4.1.1. However, the $e^{i_1} \otimes \cdots \otimes e^{i_k}$ are not skew-symmetric, in general, so these do not give a basis for $\Lambda^k(E)$. To produce such a basis we introduce a "skew-symmetrized" tensor product. Suppose $\alpha \in \Lambda^k(E)$ and $\beta \in \Lambda^l(E)$. We define their **wedge product** $\alpha \wedge \beta$ as follows: If $k = 0$ or $l = 0$ (or both), we simply let $\alpha \wedge \beta = \alpha\beta$. If $k > 0$ and $l > 0$ we define

$$\begin{aligned}(\alpha \wedge \beta)(v_1, \ldots, v_{k+l}) &= \frac{1}{k!\,l!} \sum_\sigma (-1)^\sigma \, (\alpha \otimes \beta)\left(v_{\sigma(1)}, \ldots, v_{\sigma(k+l)}\right) \\ &= \frac{1}{k!\,l!} \sum_\sigma (-1)^\sigma \, \alpha\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) \\ &\qquad \times \beta\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\right),\end{aligned}$$

where the sum is over all permutations $\sigma \in S_{k+l}$ of $\{1, \ldots, k+l\}$.

Exercise 4.1.4 Write out each of the following special cases and show that, in each case, $\alpha \wedge \beta$ is skew-symmetric.

(a) If $\alpha \in \Lambda^1(E)$ and $\beta \in \Lambda^1(E)$, then

$$\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha.$$

(b) If $\alpha \in \Lambda^1(E)$ and $\beta \in \Lambda^2(E)$, then

$$(\alpha \wedge \beta)(v_1, v_2, v_3) = \alpha(v_1)\beta(v_2, v_3) - \alpha(v_2)\beta(v_1, v_3) \\ + \alpha(v_3)\beta(v_1, v_2).$$

Showing that, in the general case, $\alpha \wedge \beta$ is skew-symmetric and deriving the basic properties of the wedge product takes a bit of work. To facilitate the arguments we introduce some notation. If $\sigma \in S_k$ is a permutation of $\{1, \ldots, k\}$ and $(v_1, \ldots, v_k)$ is any k-tuple we will write

$$\sigma \cdot (v_1, \ldots, v_k) = \Big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\Big) \tag{4.1.1}$$

for the rearranged k-tuple.

Exercise 4.1.5 Show that if $\sigma, \rho \in S_k$ are two permutations of $\{1, \ldots, k\}$, then

$$\sigma \cdot \Big(\rho \cdot (v_1, \ldots, v_k)\Big) = (\rho \circ \sigma) \cdot (v_1, \ldots, v_k)$$

for any k-tuple $(v_1, \ldots, v_k)$.

Now define, for any $A \in \mathcal{T}^k(E)$, a multilinear function $\mathrm{Alt}(A)$ on $E \times \cdots \times E$ (k factors) by

$$\mathrm{Alt}(A)\,(v_1, \ldots, v_k) = \frac{1}{k\,!} \sum_{\sigma \in S_k} (-1)^\sigma \, A\Big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\Big). \tag{4.1.2}$$

Thus, in particular, if $\alpha \in \Lambda^k(E)$ and $\beta \in \Lambda^l(E)$, then

$$\alpha \wedge \beta = \frac{(k+l)!}{k!\,l!} \mathrm{Alt}(\alpha \otimes \beta). \tag{4.1.3}$$

Lemma 4.1.2 *Let $k \geq 0$ be an integer. Then*

1. $A \in \mathcal{T}^k(E) \Rightarrow \mathrm{Alt}(A) \in \Lambda^k(E)$.
2. $\alpha \in \Lambda^k(E) \Rightarrow \mathrm{Alt}(\alpha) = \alpha$.
3. $A \in \mathcal{T}^k(E) \Rightarrow \mathrm{Alt}(\mathrm{Alt}(A)) = \mathrm{Alt}(A)$.

Proof: (3) obviously follows from (1) and (2). To prove (1) we observe first that $\mathrm{Alt}(A)$ is clearly multilinear so $\mathrm{Alt}(A) \in \mathcal{T}^k(E)$. To prove that $\mathrm{Alt}(A)$ is skew-symmetric we verify (b) in Exercise 4.1.1. Thus, we fix i and j with $1 \leq i, j \leq k$ and show that interchanging the i^{th} and j^{th} arguments changes the sign of $\mathrm{Alt}(A)$. Assume, without loss of generality, that $1 \leq i < j \leq k$ and let (ij) be the permutation of $\{1, \ldots, k\}$ that switches i and j, but leaves the others fixed. Thus, for any k-tuple $(a, b, \ldots, c)$, $(ij) \cdot (a, b, \ldots, c)$ has the i^{th} and j^{th} slots switched while the others are left fixed. Moreover, as σ varies over all the permutations in S_k, so does $\sigma' = (ij) \circ \sigma$, but $(-1)^{\sigma'} = -(-1)^\sigma$. Thus,

$$
\begin{aligned}
&\operatorname{Alt}(A)(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) \\
&\quad = \frac{1}{k\,!} \sum_{\sigma} (-1)^{\sigma} A\Big(\sigma \cdot (v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)\Big) \\
&\quad = \frac{1}{k\,!} \sum_{\sigma} (-1)^{\sigma} A\Bigg(\sigma \cdot \Big((ij) \cdot (v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k)\Big)\Bigg) \\
&\quad = \frac{1}{k\,!} \sum_{\sigma} (-1)^{\sigma} A\Big(((ij) \circ \sigma) \cdot (v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k)\Big) \\
&\quad = -\frac{1}{k\,!} \sum_{\sigma} (-1)^{\sigma'} A\Big(\sigma' \cdot (v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k)\Big) \\
&\quad = -\frac{1}{k\,!} \sum_{\sigma'} (-1)^{\sigma'} A\Big(\sigma' \cdot (v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k)\Big) \\
&\quad = -\operatorname{Alt}(A)(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k).
\end{aligned}
$$

Remark: Notice that the factor $\frac{1}{k!}$ played no role in the proof of (1). This is not the case, however, for the proof (2).

To prove (2) we begin with an $\alpha \in \Lambda^k(E)$. Observe first that, for any $\sigma \in S_k$,

$$
\alpha\Big(\sigma \cdot (v_1, \ldots, v_k)\Big) = (-1)^{\sigma}\, \alpha\,(v_1, \ldots, v_k).
$$

(Exercise 4.1.1 (c)). Thus,

$$
\begin{aligned}
\operatorname{Alt}\,(\alpha)(v_1, \ldots, v_k) &= \frac{1}{k\,!} \sum_{\sigma} (-1)^{\sigma}\, \alpha\,\Big(\sigma \cdot (v_1, \ldots, v_k)\Big) \\
&= \frac{1}{k\,!} \sum_{\sigma}\, (-1)^{\sigma}\, (-1)^{\sigma}\, \alpha(v_1, \ldots, v_k) \\
&= \frac{1}{k\,!} \Big(k!\, \alpha(v_1, \ldots, v_k)\Big) = \alpha(v_1, \ldots, v_k)
\end{aligned}
$$

as required (notice that the $\frac{1}{k!}$ is essential here). ■

From Lemma 4.1.2 (a) and (4.1.3) we conclude that, if $\alpha \in \Lambda^k(E)$ and $\beta \in \Lambda^l(E)$, then $\alpha \wedge \beta \in \Lambda^{k+l}(E)$. To establish the basic properties of the wedge product we need two more technical results.

Lemma 4.1.3 *Let* $A \in \mathcal{T}^k(E)$ *and* $B \in \mathcal{T}^l(E)$ *and suppose* $\operatorname{Alt}(A) = 0$. *Then* $\operatorname{Alt}(A \otimes B) = \operatorname{Alt}(B \otimes A) = 0$.

Proof: Begin by writing

$$
\begin{aligned}
&(k+l)!\operatorname{Alt}(A \otimes B)\,(v_1, \ldots, v_{k+l}) \\
&\quad = \sum_{\sigma \,\in\, S_{k+l}} (-1)^{\sigma}\, A\Big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\Big)\, B\Big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\Big).
\end{aligned}
$$

Now let $G = \{\sigma \in S_{k+l} : \sigma(i) = i,\ i = k+1, \ldots, k+l\}$.

Exercise 4.1.6 Show that $\sum_{\sigma \in G} (-1)^\sigma A(v_{\sigma(1)}, \ldots, v_{\sigma(k)})\, B(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}) = 0$.

Next suppose $\sigma_0 \in S_{k+l} - G$. Consider

$$\sigma_0 \circ G = \{\sigma_0 \circ \sigma : \sigma \in G\}.$$

Observe that $G \cap (\sigma_0 \circ G) = \emptyset$ since if τ were in $G \cap (\sigma_0 \circ G)$, then $\tau = \sigma_0 \circ \sigma$ for some $\sigma \in G$ and therefore $\sigma_0 = \tau \circ \sigma^{-1}$ would imply that σ_0 is in G, which it is not. Now,

$$\begin{aligned} \sum_{\tau \in \sigma_0 \circ G} & (-1)^\tau (A \otimes B)\Big(\tau \cdot (v_1, \ldots, v_{k+l})\Big) \\ &= \sum_{\sigma \in G} (-1)^{\sigma_0 \circ \sigma} (A \otimes B)\Big((\sigma_0 \circ \sigma)(v_1, \ldots, v_{k+l})\Big) \\ &= \sum_{\sigma \in G} (-1)^{\sigma_0} (-1)^\sigma (A \otimes B)\Bigg(\sigma \cdot \Big(\sigma_0 \cdot (v_1, \ldots, v_{k+l})\Big)\Bigg). \end{aligned}$$

Let $\sigma_0 \cdot (v_1, \ldots, v_{k+l}) = (w_1, \ldots, w_k, w_{k+1}, \ldots, w_{k+l})$. Then this last sum becomes

$$(-1)^{\sigma_0} \sum_{\sigma \in G} (-1)^\sigma A\left(w_{\sigma(1)}, \ldots, w_{\sigma(k)}\right) B\left(w_{\sigma(k+1)}, \ldots, w_{\sigma(k+l)}\right)$$

which is zero by Exercise 4.1.6. Thus,

$$\sum_{\tau \in \sigma_0 \circ G} (-1)^\tau (A \otimes B)\Big(\tau \cdot (v_1, \ldots, v_{k+l})\Big) = 0.$$

Next, suppose $\sigma_1 \in S_{k+l} - (G \cup \sigma_0 \circ G)$ and let $\sigma_1 \circ G = \{\sigma_1 \circ \sigma : \sigma \in G\}$. As above, $G \cap (\sigma_1 \circ G) = \emptyset$.

Exercise 4.1.7 Show that, moreover, $(\sigma_1 \circ G) \cap (\sigma_0 \circ G) = \emptyset$.

Now, exactly as above, we have

$$\sum_{\tau \in \sigma_1 \circ G} (-1)^\tau (A \otimes B)\Big(\tau \cdot (v_1, \ldots, v_{k+l})\Big) = 0.$$

Since S_{k+l} is finite we may continue in this way, splitting S_{k+l} into finitely many disjoint subsets, the sum over each being zero. Thus,

$$\sum_{\tau \in S_{k+l}} (-1)^\tau (A \otimes B)\Big(\tau \cdot (v_1, \ldots, v_{k+l})\Big) = 0$$

so $\text{Alt}(A \otimes B) = 0$. Replacing G by $G' = \{\sigma \in S_{k+l} : \sigma(i) = i,\ i = 1, \ldots, k\}$ the same argument shows that $\text{Alt}(A \otimes B) = 0$. ■

Lemma 4.1.4 *Let $\alpha \in \Lambda^k(E)$, $\beta \in \Lambda^l(E)$ and $\gamma \in \Lambda^m(E)$. Then*

$$\text{Alt}(\alpha \otimes \text{Alt}(\beta \otimes \gamma)) = \text{Alt}(\alpha \otimes \beta \otimes \gamma) = \text{Alt}(\text{Alt}(\alpha \otimes \beta) \otimes \gamma).$$

Proof: First observe that

$$\begin{aligned}\text{Alt}(\text{Alt}(\beta \otimes \gamma) - \beta \otimes \gamma) &= \text{Alt}(\text{Alt}(\beta \otimes \gamma)) - \text{Alt}(\beta \otimes \gamma)\\ &= \text{Alt}(\beta \otimes \gamma) - \text{Alt}(\beta \otimes \gamma) = 0\end{aligned}$$

by (3) of Lemma 4.1.2. Thus, by Lemma 4.1.3

$$\begin{aligned}0 &= \text{Alt}(\alpha \otimes [\text{Alt}(\beta \otimes \gamma) - \beta \otimes \gamma])\\ &= \text{Alt}(\alpha \otimes \text{Alt}(\beta \otimes \gamma) - \alpha \otimes \beta \otimes \gamma)\\ &= \text{Alt}(\alpha \otimes \text{Alt}(\beta \otimes \gamma)) - \text{Alt}(\alpha \otimes \beta \otimes \gamma)\end{aligned}$$

which proves the first equality. The second equality is proved analogously. ■

Theorem 4.1.5 *Let $\alpha, \alpha^1, \alpha^2 \in \Lambda^k(E)$, $\beta, \beta^1, \beta^2 \in \Lambda^l(E)$, $\gamma \in \Lambda^m(E)$ and $a \in$. Then*

1. $(\alpha^1 + \alpha^2) \wedge \beta = \alpha^1 \wedge \beta + \alpha^2 \wedge \beta$
2. $\alpha \wedge (\beta^1 + \beta^2) = \alpha \wedge \beta^1 + \alpha \wedge \beta^2$
3. $(a\alpha) \wedge \beta = \alpha \wedge (a\beta) = a(\alpha \wedge \beta)$
4. $\beta \wedge \alpha = (-1)^{kl}\, \alpha \wedge \beta$
5. $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma) = \frac{(k+l+m)!}{k!\,l!\,m!}\, Alt(\alpha \otimes \beta \otimes \gamma)$
6. *If $T : E_1 \longrightarrow E_2$ is a linear transformation, $\alpha \in \Lambda^k(E_2)$ and $\beta \in \Lambda^l(E_2)$, then*
$$T^*(\alpha \wedge \beta) = (T^*\alpha) \wedge (T^*\beta).$$

Proof: (1), (2) and (3) follow directly from the corresponding properties of the tensor product (Exercise 4.1.2). To prove (4) we write

$$\begin{aligned}&(\alpha \wedge \beta)\,(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l})\\ &\qquad = \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} (-1)^\sigma\, \alpha\Big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\Big)\, \beta\Big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\Big)\end{aligned}$$

and

$$\begin{aligned}
&(\beta \wedge \alpha)\,(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}) \\
&\quad = \frac{1}{l!\,k!} \sum_{\sigma \,\in\, S_{k+l}} (-1)^\sigma\, \beta\left(v_{\sigma(1)}, \ldots, v_{\sigma(l)}\right) \alpha\left(v_{\sigma(l+1)}, \ldots, v_{\sigma(l+k)}\right) \\
&\quad = \frac{1}{k!\,l!} \sum_{\sigma \,\in\, S_{k+l}} (-1)^\sigma\, \alpha\left(v_{\sigma(l+1)}, \ldots, v_{\sigma(l+k)}\right) \beta\left(v_{\sigma(1)}, \ldots, v_{\sigma(l)}\right).
\end{aligned}$$

Let ρ be the permutation of $\{1, \ldots, k+l\}$ that sends $\{1, \ldots, l, l+1, \ldots, l+k\}$ in order to $\{l+1, \ldots, l+k, 1, \ldots, l\}$. Then ρ is the product of kl transpositions (move each of the last k in order over the preceding l) so $(-1)^\rho = (-1)^{kl}$. Thus, for any $\sigma \in S_{k+l}, (-1)^{\sigma \circ \rho} = (-1)^{kl}(-1)^\sigma$. Moreover, as σ varies over S_{k+l}, so does $\sigma' = \sigma \circ \rho$. Thus,

$$\begin{aligned}
&(\beta \wedge \alpha)\,(v_1, \ldots, v_{k+l}) \\
&= \frac{1}{k!\,l!} \sum_{\sigma \,\in\, S_{k+l}} (-1)^\sigma\, (\alpha \otimes \beta)\left(v_{\sigma(l+1)}, \ldots, v_{\sigma(l+k)}, v_{\sigma(1)}, \ldots, v_{\sigma(l)}\right) \\
&= \frac{1}{k!\,l!} \sum_{\sigma \,\in\, S_{k+l}} (-1)^\sigma\, (\alpha \otimes \beta)\left(\rho \cdot \left(v_{\sigma(1)}, \ldots, v_{\sigma(l)}, v_{\sigma(l+1)}, \ldots, v_{\sigma(l+k)}\right)\right) \\
&= \frac{1}{k!\,l!} \sum_{\sigma \,\in\, S_{k+l}} (-1)^\sigma\, (\alpha \otimes \beta)\left(\rho \cdot \left(\sigma \cdot (v_1, \ldots, v_{k+l})\right)\right) \\
&= \frac{1}{k!\,l!} \sum_{\sigma \,\in\, S_{k+l}} (-1)^{kl}\, (-1)^{\sigma \circ \rho}\, (\alpha \otimes \beta)\Big((\sigma \circ \rho) \cdot (v_1, \ldots, v_{k+l})\Big) \\
&= (-1)^{kl} \frac{1}{k!\,l!} \sum_{\sigma' \in\, S_{k+l}} (-1)^{\sigma'} (\alpha \otimes \beta)\Big(\sigma' \cdot (v_1, \ldots, v_{k+l})\Big) \\
&= (-1)^{kl}\, (\alpha \wedge \beta)\,(v_1, \ldots, v_{k+l})
\end{aligned}$$

as required.

Exercise 4.1.8 Use Lemma 4.1.4 to prove (5).

Exercise 4.1.9 Prove (6). ■

A few special cases of Theorem 4.1.5 are worth pointing out explicitly. First suppose $\alpha \in \Lambda^k(E)$ and $\beta \in \Lambda^l(E)$. If either k or l is even, then

$$\beta \wedge \alpha = \alpha \wedge \beta,$$

but if both k and l are odd, then

$$\beta \wedge \alpha = -\alpha \wedge \beta.$$

Thus, if k is odd, then

$$\alpha \wedge \alpha = 0.$$

This is true, in particular, for 1-forms. Associativity of the wedge product (Theorem 4.1.5 (5)) permits the unambiguous definition of wedge products for any finite number of forms. The following special case is of particular interest.

Exercise 4.1.10 Let $\alpha^1, \ldots, \alpha^k$ be in $\Lambda^1(E)$. Show that

$$\alpha^1 \wedge \cdots \wedge \alpha^k = k!\, Alt(\alpha^1 \otimes \cdots \otimes \alpha^k) = \sum_{\sigma \in S_k} (-1)^\sigma\, \alpha^{\sigma(1)} \otimes \cdots \otimes \alpha^{\sigma(k)}.$$

Hint: Proceed by induction using (4.1.3).

Theorem 4.1.6 *Let $\{e_1, \ldots, e_n\}$ be a basis for E and $\{e^1, \ldots, e^n\}$ the corresponding dual basis for $E^* = \Lambda^1(E)$. Then, for each $k = 1, \ldots, n$, $\{e^{i_1} \wedge \cdots \wedge e^{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$ is a basis for $\Lambda^k(E)$. Moreover, any $\alpha \in \Lambda^k(E)$ can be uniquely written as*

$$\alpha = \sum_{i_1 < \cdots < i_k} \alpha_{i_1 \cdots i_k}\, e^{i_1} \wedge \cdots \wedge e^{i_k} = \frac{1}{k!}\, \alpha_{i_1 \cdots i_k}\, e^{i_1} \wedge \cdots \wedge e^{i_k},$$

where $\alpha_{i_1 \cdots i_k} = \alpha(e_{i_1}, \ldots, e_{i_k})$. In particular, $\dim \Lambda^k(E) = \binom{n}{k}$.

Proof: Since $\alpha \in \Lambda^k(E) \subseteq \mathcal{T}^k(E)$, Lemma 4.1.1 implies that $\alpha = \alpha_{i_1 \cdots i_k}\, e^{i_1} \otimes \cdots \otimes e^{i_k}$, where $\alpha_{i_1 \cdots i_k} = \alpha(e_{i_1}, \ldots, e_{i_k})$ and the sum is over *all* $1 \leq i_1, \ldots, i_k \leq n$. Exercise 4.1.10 then gives

$$\alpha = \frac{1}{k!}\, \alpha_{i_1 \cdots i_k}\, e^{i_1} \wedge \cdots \wedge e^{i_k}.$$

Now, if any term in this last sum has two of the indices $i_1, \ldots, i_k$ equal the wedge product is zero. Hence, we may sum only over those $i_1, \ldots, i_k$ that are all distinct. Each such sequence of distinct indices gives rise to $k!$ terms in the sum, each term differing from the others only in the order of the indices. But permuting the indices changes both $\alpha_{i_1 \cdots i_k}$ and $e^{i_1} \wedge \cdots \wedge e^{i_k}$ by the sign of the permutation so all of these terms are the same and equal to the unique term $\alpha_{i_1 \cdots i_k}\, e^{i_1} \wedge \cdots \wedge e^{i_k}$ with $i_1 < \cdots < i_k$. Consequently,

$$\alpha = \sum_{i_1 < \cdots < i_k} \alpha_{i_1 \cdots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k}.$$

In particular, $\{e^{i_1} \wedge \cdots \wedge e^{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$ spans $\Lambda^k(E)$.

Exercise 4.1.11 Complete the argument by proving linear independence. ∎

Note that, if $\dim E = n$ and $k > n$, then $\Lambda^k(E)$ contains only the zero function. Moreover, if $k = n$, then $\dim \Lambda^n(E) = \binom{n}{n} = 1$ so any nonzero element of $\Lambda^n(E)$ spans $\Lambda^n(E)$. We wish to investigate this 1-dimensional space of top forms a bit more closely.

Lemma 4.1.7 *Let $\alpha^1, \ldots, \alpha^k$ be elements of $\Lambda^1(E)$. Then $\{\alpha^1, \ldots, \alpha^k\}$ is linearly independent in $\Lambda^1(E)$ if and only if $\alpha^1 \wedge \cdots \wedge \alpha^k \neq 0$ in $\Lambda^k(E)$.*

Proof: First suppose $\{\alpha^1, \ldots, \alpha^k\}$ is linearly independent. Then there is a basis $\{e_1, \ldots, e_k, \ldots, e_n\}$ for E such that the dual basis $\{e^1, \ldots, e^k, \ldots, e^n\}$ satisfies $e^i = \alpha^i$ for $i = 1, \ldots, k$. Thus, $\alpha^1 \wedge \cdots \wedge \alpha^k$ is an element of a basis for $\Lambda^k(E)$ and so, in particular, is not zero.

Exercise 4.1.12 Prove the converse. ■

Theorem 4.1.8 *Let $\{e_1, \ldots, e_n\}$ be a basis for E and let $\alpha \in \Lambda^n(E)$. If $v_j = A^i{}_j e_i$, $j = 1, \ldots, n$, are any n vectors in E, then*

$$\alpha(v_1, \ldots, v_n) = \det(A^i{}_j)\ \alpha(e_1, \ldots, e_n).$$

Proof: First observe that the ordinary determinant function det can be regarded as an element of $\Lambda^n(\quad{}^n)$ as follows: For $v_1, \ldots, v_n \in \quad{}^n$ we write $v_j = (A^1{}_j, \ldots, A^n{}_j)$, $j = 1, \ldots, n$ (components relative to the standard basis for $\quad{}^n$) and define

$$\det(v_1, \ldots, v_n) = \det \begin{pmatrix} A^1{}_1 & \cdots & A^n{}_1 \\ \vdots & & \vdots \\ A^n{}_1 & \cdots & A^n{}_n \end{pmatrix}.$$

Properties of determinants ensure that this is skew-symmetric so $\det \in \Lambda^n(\quad{}^n)$. Since det is not identically zero on $\quad{}^n$ it generates $\Lambda^n(\quad{}^n)$.

Now, if $\alpha \in \Lambda^n(E)$ we define $\alpha' \in \Lambda^n(\quad{}^n)$ by

$$\alpha'\Big((A^1{}_1, \ldots, A^n{}_1), \ldots, (A^1{}_n, \ldots, A^n{}_n)\Big) = \alpha\big(A^i{}_1 e_i, \ldots, A^i{}_n e_i\big).$$

Then $\alpha' = c \det$ for some constant c. Evaluating α' at $((1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1))$ gives $\alpha'((1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)) = c \det\ (\mathrm{id})$, i.e., $\alpha(e_1, \ldots, e_n) = c$. Thus,

$$\begin{aligned} \alpha(v_1, \ldots, v_n) &= \alpha(A^i{}_1\, e_i, \ldots, A^i{}_n\, e_i) \\ &= \alpha'\Big((A^1{}_1, \ldots, A^n{}_1), \ldots, (A^1{}_n, \ldots, A^n{}_n)\Big) \\ &= \alpha(e_1, \ldots, e_n) \det(A^i{}_j). \end{aligned}$$

■

Corollary 4.1.9 *Let* $\{e_1, \ldots, e_n\}$ *be a basis for* E, $\{e^1, \ldots, e^n\}$ *its dual basis for* $\Lambda^1(E)$ *and* $v_1, \ldots, v_n$ *vectors in* E *with* $v_j = A^i{}_j\, e_i,\ j = 1, \ldots, n$. *Then*

$$(e^1 \wedge \cdots \wedge e^n)(v_1, \ldots, v_n) = \det(A^i{}_j).$$

Corollary 4.1.10 *Let* $T : E \longrightarrow E$ *be a linear transformation. Then* $T^* : \Lambda^n(E) \longrightarrow \Lambda^n(E)$ *is given by*

$$T^*\alpha = (\det T)\alpha$$

for every $\alpha \in \Lambda^n(E)$.

Proof: If $\alpha = 0$, then $T^*\alpha = 0$ so the result is trivial. Suppose then that $\alpha \neq 0$. Since $\dim \Lambda^n(E) = 1$, α generates $\Lambda^n(E)$ so $T^*\alpha = c\alpha$ for some constant c. Let $\{e_1, \ldots, e_n\}$ be a basis for E and write $Te_j = A^i{}_j e_i$, $j = 1, \ldots, n$. Then

$$\begin{aligned}(T^*\alpha)(e_1, \ldots, e_n) &= \alpha(Te_1, \ldots, Te_n)\\ &= \alpha(A^i{}_1 e_i, \ldots, A^i{}_n e_i)\\ &= \det(A^i{}_j)\,\alpha\,(e_1, \ldots, e_n)\end{aligned}$$

so

$$c\alpha\ (e_1, \ldots, e_n) = \det(A^i{}_j)\,\alpha\,(e_1, \ldots, e_n).$$

Since $\alpha \neq 0$, Theorem 4.1.8 implies that $\alpha(e_1, \ldots, e_n) \neq 0$ so $c = \det(A^i{}_j)$. ■

As another application of Theorem 4.1.8 we show that any nonzero element of $\Lambda^n(E)$ determines a unique orientation for E.

Theorem 4.1.11 *Let* E *be an* n*-dimensional real vector space and* α *a nonzero element of* $\Lambda^n(E)$. *Then there is a unique orientation* μ *for* E *such that* $[e_1, \ldots, e_n] \in \mu$ *if and only if* $\alpha(e_1, \ldots, e_n) > 0$.

Proof: Since α is nonzero, Theorem 4.1.8 implies that $\alpha(e_1, \ldots, e_n)$ is nonzero for every basis $\{e_1, \ldots, e_n\}$ for E. Thus, the *ordered* bases for E are divided into two disjoint classes according to whether $\alpha(e_1, \ldots, e_n) > 0$ or $\alpha(e_1, \ldots, e_n) < 0$. Let μ denote the set of all ordered bases for E for which $\alpha(e_1, \ldots, e_n) > 0$. We claim that μ is an orientation for E, i.e., that μ is an equivalence class of the equivalence relation $\sim$ defined on ordered bases as follows: $\{\hat{e}_1, \ldots, \hat{e}_n\} \sim \{e_1, \ldots, e_n\}$ if and only if $\det(A^i{}_j) > 0$, where $\hat{e}_j = A^i{}_j\, e_i,\ j = 1, \ldots, n$. But this is clear from Theorem 4.1.8 since $\alpha(\hat{e}_1, \ldots, \hat{e}_n) = \det(A^i{}_j)\alpha(e_1, \ldots, e_n)$. Uniqueness is also clear since E has precisely two orientations and the other one (i.e., $-\mu$) consists of the ordered bases for which $\alpha(e_1, \ldots, e_n) < 0$. ■

Exercise 4.1.13 Show that if $\{e_1, \ldots, e_n\}$ is an ordered basis for E and $\{e^1, \ldots, e^n\}$ is its dual basis for $\Lambda^1(E)$, then the orientation for E determined by $e^1 \wedge \cdots \wedge e^n$ is precisely the one that contains $\{e_1, \ldots, e_n\}$.

Thus, any nonzero n-form on E determines a unique orientation for E. Let us suppose that, conversely, one is given an orientation μ for E. Then every nonzero element of $\Lambda^n(E)$ is either positive on all the elements of μ or negative on all the elements of μ (Theorem 4.1.8). Thus, μ divides the nonzero elements of $\Lambda^n(E)$ into two equivalence classes. In general, however, μ does not single out a unique element of $\Lambda^n(E)$. However, we now show that if E is equipped with the additional structure of an inner product (nondegenerate, symmetric, bilinear form, but not necessarily positive definite), then a distinguished element of $\Lambda^n(E)$ is canonically determined.

Suppose then that E has an orientation μ and an inner product g. If $\{e_1, \ldots, e_n\}$ and $\{\hat{e}_1, \ldots, \hat{e}_n\}$ are two bases for E that are *orthonormal* with respect to g, then the matrix $(A^i{}_j)$ that relates them $(\hat{e}_j = A^i{}_j\, e_i$, $j = 1, \ldots, n)$ satisfies $\det(A^i{}_j) = \pm 1$. Thus, for any nonzero $\alpha \in \Lambda^n(E)$, $\alpha(\hat{e}_1, \ldots, \hat{e}_n) = \pm\alpha(e_1, \ldots, e_n)$.

Now let $\{e_1, \ldots, e_n\}$ be an ordered orthonormal basis in μ (i.e., an **oriented orthonormal basis**). Then there exists a nonzero element ω of $\Lambda^n(E)$ with $\omega(e_1, \ldots, e_n) = 1$ (namely, $e^1 \wedge \cdots \wedge e^n$). But observe that if $\{\hat{e}_1, \ldots, \hat{e}_n\}$ is any other oriented orthonormal basis for E, then $\omega(\hat{e}_1, \ldots, \hat{e}_n) = 1$ as well. Thus, ω takes *any* oriented orthonormal basis to 1. Moreover, it carries any orthonormal basis in $-\mu$ to -1.

Exercise 4.1.14 Show that ω is uniquely determined, i.e., that there is only one element of $\Lambda^n(E)$ that carries every oriented orthonormal basis for E to 1.

Theorem 4.1.12 *Let E be an n-dimensional real vector space with an orientation μ and an inner product g. Then there exists a unique $\omega \in \Lambda^n(E)$ such that $\omega(e_1, \ldots, e_n) = 1$ whenever $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis for E.*

For example, if $E = \ {}^n$ with its standard orientation and (positive definite) inner product, then ω is just the determinant function $\det \in \Lambda^n(\ {}^n)$ (see the first few lines in the proof of Theorem 4.1.8). In general, ω is called the (**metric**) **volume form** for E determined by μ and g (recall that, in ^{3}, $|\det(v_1, v_2, v_3)|$ is the volume of the parallelepiped spanned by v_1, v_2 and v_3). We know that if $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis for E, then $\omega = e^1 \wedge \cdots \wedge e^n$, where $\{e^1, \ldots, e^n\}$ is the dual basis. We will need to compute ω in an arbitrary oriented (but not necessarily orthonormal) basis for E.

Exercise 4.1.15 Let ω be the volume form on E determined by the orientation μ and an inner product g. Let $\{\hat{e}_1, \ldots, \hat{e}_n\}$ be an oriented basis for E and $\{\hat{e}^1, \ldots, \hat{e}^n\}$ its dual basis. For each $i, j = 1, \ldots, n$ let $\hat{g}_{ij} = g(\hat{e}_i, \hat{e}_j)$. Show that

$$\omega = |\det(\hat{g}_{ij})|^{\frac{1}{2}}\, \hat{e}^1 \wedge \cdots \wedge \hat{e}^n.$$

Hint: Let $\{e_1, \ldots, e_n\}$ be an oriented orthonormal basis for E with $\hat{e}_j = A^i{}_j\, e_i$, $j = 1, \ldots, n$ and show that $\det(A^i{}_j) = |\det(\hat{g}_{ij})|^{1/2}$.

Next we observe the following: If $\dim E = n$ and k is any integer with $0 \leq k \leq n$, then (by Theorem 4.1.6)

$$\dim \Lambda^{n-k}(E) = \left(\begin{array}{c} n \\ n-k \end{array} \right) = \left(\begin{array}{c} n \\ k \end{array} \right) = \dim \Lambda^k(E).$$

In particular, $\Lambda^{n-k}(E)$ is isomorphic to $\Lambda^k(E)$ for any vector space E. In general, however, there is no "natural" isomorphism between these two. We show next, however, that if E has an orientation μ and an inner product g, then there is a canonical isomorphism

$$^* : \Lambda^k(E) \longrightarrow \Lambda^{n-k}(E)$$

called the **Hodge star operator** (the image of a $\beta \in \Lambda^k(E)$ under this isomorphism will be denoted $^*\beta \in \Lambda^{n-k}(E)$ and called the **Hodge dual** of β). To construct the isomorphism we must first show that the inner product g on E determines an inner product on each $\Lambda^k(E)$.

Suppose then that E is an n-dimensional real vector space with an inner product g (E need not have an orientation for this part of the construction). We wish to show that g induces an inner product (also denoted g) on $\Lambda^k(E)$ for each $k = 0, 1, \ldots, n$. Since $\Lambda^0(E) =$ we will define, for $\alpha, \beta \in \Lambda^0(E)$, $g(\alpha, \beta) = \alpha\beta$. For $k > 0$ our objective is to define g in such a way that, if $\{e_1, \ldots, e_n\}$ is an orthonormal basis for E with dual basis $\{e^1, \ldots, e^n\}$, then $\{e^{i_1} \wedge \cdots \wedge e^{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$ will be an orthonormal basis for $\Lambda^k(E)$. The procedure will be to first define g in terms of an arbitrary basis for E, then show that the definition does not actually depend on the initial choice of the basis and finally prove that the definition has the desired property.

Let $\{e_1, \ldots, e_n\}$ be an arbitrary basis for E (not necessarily orthonormal). Let $g_{ij} = g(e_i, e_j)$ for $i, j = 1, \ldots, n$, and let (g^{ij}) denote the matrix inverse of (g_{ij}). Now, for $\alpha, \beta \in \Lambda^k(E)$, $k \geq 1$, we write

$$\alpha = \frac{1}{k!} \alpha_{i_1 \cdots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k}$$

and

$$\beta = \frac{1}{k!} \beta_{j_1 \cdots j_k} e^{j_1} \wedge \cdots \wedge e^{j_k}$$

as in Theorem 4.1.6. Now define

$$\alpha^{j_1 \cdots j_k} = g^{i_1 j_1} \cdots g^{i_k j_k} \alpha_{i_1 \cdots i_k} \tag{4.1.4}$$

(classically, this is called "raising the indices of α with the metric g").
Our definition of g on $\Lambda^k(E)$ is now

$$g(\alpha, \beta) = \frac{1}{k!} \alpha^{j_1 \cdots j_k} \beta_{j_1 \cdots j_k}. \tag{4.1.5}$$

To show that this definition does not depend on the choice of basis we let $\{\hat{e}_1, \ldots, \hat{e}_n\}$ be another basis with $\hat{e}_j = A^i{}_j e_i$, $j = 1, \ldots, n$. Let $(A_i{}^j)$ denote the matrix inverse of $(A^i{}_j)$. Then the dual bases $\{e^1, \ldots, e^n\}$ and $\{\hat{e}^1, \ldots, \hat{e}^n\}$ are related by $\hat{e}^j = A_i{}^j e^i,\ j = 1, \ldots, n$.

Exercise 4.1.16 Let $\hat{g}_{ij} = g(\hat{e}_i, \hat{e}_j)$ and let $(\hat{g}^{ij})$ be the matrix inverse of $(\hat{g}_{ij})$. Show that

$$\hat{g}_{ij} = A^k{}_i A^l{}_j g_{kl}$$

and

$$\hat{g}^{ij} = A_k{}^i A_l{}^j g^{kl}$$

for $i, j = 1, \ldots, n$.

Now, for $\alpha \in \Lambda^k(E)$ we write $\alpha = \dfrac{1}{k!}\alpha_{i_1\cdots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k} = \dfrac{1}{k!}\hat{\alpha}_{i_1\ldots i_k}\hat{e}^{i_1} \wedge \cdots \wedge \hat{e}^{i_k}$, where

$$\begin{aligned}\hat{\alpha}_{i_1\ldots i_k} &= \alpha(\hat{e}_{i_1}, \ldots, \hat{e}_{i_k}) = \alpha\Big(A^{j_1}{}_{i_1} e_{j_1}, \ldots, A^{j_k}{}_{i_k} e_{j_k}\Big) \\ &= A^{j_1}{}_{i_1} \cdots A^{j_k}{}_{i_k} \alpha_{j_1\cdots j_k}.\end{aligned}$$

Exercise 4.1.17 Show that $\hat{\alpha}^{j_1\cdots j_k} = A_{l_1}{}^{j_1} \ldots A_{l_k}{}^{j_k}\, \alpha^{l_1\cdots l_k}$.

Finally, we compute

$$\begin{aligned}\frac{1}{k!}\hat{\alpha}^{j_1\cdots j_k}\hat{\beta}_{j_1\cdots j_k} &= \frac{1}{k!}\Big(A_{l_1}{}^{j_1} \cdots A_{l_k}{}^{j_k}\alpha^{l_1\cdots l_k}\Big)\Big(A^{m_1}{}_{j_1} \cdots A^{m_k}{}_{j_k}\beta_{m_1\cdots m_k}\Big) \\ &= \frac{1}{k!}\Big(A_{l_1}{}^{j_1}A^{m_1}{}_{j_1}\Big) \cdots \Big(A_{l_k}{}^{j_k}A^{m_k}{}_{j_k}\Big)\alpha^{l_1\cdots l_k}\beta_{m_1\cdots m_k} \\ &= \frac{1}{k!}\delta_{l_1}{}^{m_1} \cdots \delta_{l_k}{}^{m_k}\alpha^{l_1\cdots l_k}\beta_{m_1\cdots m_k} \\ &= \frac{1}{k!}\alpha^{l_1\cdots l_k}\beta_{l_1\cdots l_k} \\ &= \frac{1}{k!}\alpha^{j_1\cdots j_k}\beta_{j_1\cdots j_k}\end{aligned}$$

as required. Thus, g is well-defined on $\Lambda^k(E)$ and it is clearly bilinear, symmetric and nondegenerate.

Theorem 4.1.13 *Let E be an n-dimensional real vector space with an inner product g and let k be an integer with $1 \leq k \leq n$. If $\{e_1, \ldots, e_n\}$ is an orthonormal basis for E with dual basis $\{e^1, \ldots, e^n\}$ for $\Lambda^1(E)$, then $\{e^{i_1} \wedge \cdots \wedge e^{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$ is an orthonomal basis for $\Lambda^k(E)$ relative to the induced metric g on $\Lambda^k(E)$ (defined by (4.1.5)).*

Proof: Let $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_k$ be fixed increasing sequences of indices and let $\alpha = e^{i_1} \wedge \cdots \wedge e^{i_k}$ and $\beta = e^{j_1} \wedge \cdots \wedge e^{j_k}$. Then $\alpha = \frac{1}{k!}\alpha_{l_1\cdots l_k}\, e^{l_1} \wedge \cdots \wedge e^{l_k}$, where $\alpha_{l_1\cdots l_k}$ is 1 if $l_1 \cdots l_k$ is an even permutation of $i_1 \cdots i_k$, -1 if $l_1 \cdots l_k$ is an odd permutation of $i_1 \cdots i_k$ and 0 otherwise. Similarly for $\beta = \frac{1}{k!}\beta_{m_1\cdots m_k}\, e^{m_1} \wedge \cdots \wedge e^{m_k}$. Since $\{e_1, \ldots, e_n\}$ is orthonormal,

(g_{ij}) is diagonal with $g_{ii} = \pm 1$ and therefore (g^{ij}) is diagonal with $g^{ii} = \pm 1$. We compute

$$g(\alpha,\beta) = \frac{1}{k!}\alpha^{m_1\cdots m_k}\beta_{m_1\cdots m_k} = \frac{1}{k!}g^{l_1 m_1}\cdots g^{l_k m_k}\alpha_{l_1\cdots l_k}\,\beta_{m_1\cdots m_k}.$$

Notice that this will be zero unless $l_1 = m_1, \ldots, l_k = m_k$ so

$$g(\alpha,\beta) = \frac{1}{k!}g^{l_1 l_1}\ldots g^{l_k l_k}\,\alpha_{l_1\cdots l_k}\,\beta_{l_1\cdots l_k}.$$

In order for $\alpha_{l_1\cdots l_k}$ to be nonzero, $\{l_1,\ldots,l_k\}$ must equal $\{i_1,\ldots,i_k\}$. Similarly, $\beta_{l_1\cdots l_k}$ will be zero unless $\{l_1,\ldots,l_k\} = \{j_1,\ldots,j_k\}$. Thus, $g(\alpha,\beta)$ will be zero unless $\{i_1,\ldots,i_k\} = \{j_1,\ldots,j_k\}$. In particular,

$$g(e^{i_1}\wedge\cdots\wedge e^{i_k}, e^{j_1}\wedge\cdots\wedge e^{j_k}) = 0 \ \text{ if } \{i_1,\ldots,i_k\} \neq \{j_1,\ldots,j_k\}.$$

Now suppose $\{i_1,\ldots,i_k\} = \{j_1,\ldots,j_k\}$. Since $1 \le i_1 < \cdots < i_k \le n$ and $1 \le j_1 < \cdots < j_k \le n$, we must then have $\alpha = \beta$ so $\beta_{l_1\cdots l_k} = \alpha_{l_1\cdots l_k}$ and therefore

$$g(\alpha,\beta) = g(\alpha,\alpha) = \frac{1}{k!}g^{l_1 l_1}\cdots g^{l_k l_k}(\alpha_{l_1\cdots l_k})^2.$$

This is a sum over all permutations of $\{i_1,\ldots,i_k\}$ and $\alpha_{l_1\cdots l_k} = \pm 1$ so all of the terms are equal and $g(\alpha,\alpha) = \pm 1$, i.e.,

$$g\ (e^{i_1}\wedge\cdots\wedge e^{i_k},\ e^{i_1}\wedge\cdots\wedge e^{i_k}) = \pm 1. \qquad \blacksquare$$

Remark: The last few lines of this proof actually show that

$$g(e^{i_1}\wedge\cdots\wedge e^{i_k}, e^{i_1}\wedge\cdots\wedge e^{i_k}) = (-1)^m,$$

where m is the number of indices among $\{i_1,\ldots,i_k\}$ for which $g_{ii} = -1$.

Theorem 4.1.14 *Let E be an n-dimensional real vector space with an orientation μ and an inner product g. Let ω be the metric volume form for E detemined by μ and g and let k be an integer with $0 \le k \le n$. Then there exists a unique isomorphism*

$${}^*: \Lambda^k(E) \longrightarrow \Lambda^{n-k}(E)$$

such that

$$\alpha \wedge {}^*\beta = g(\alpha,\beta)\omega \tag{4.1.6}$$

for all $\alpha,\beta \in \Lambda^k(E)$.

Proof: For $\gamma \in \Lambda^{n-k}(E)$ we define a map

$$\varphi_\gamma : \Lambda^k(E) \longrightarrow$$

as follows: Let $\alpha \in \Lambda^k(E)$. Then $\alpha \wedge \gamma \in \Lambda^n(E)$ so $\alpha \wedge \gamma$ is a real multiple of ω. The coefficient of ω is taken to be $\varphi_\gamma(\alpha)$ so

$$\alpha \wedge \gamma = \varphi_\gamma(\alpha)\omega.$$

φ_γ is clearly linear and is therefore an element of the dual $(\Lambda^k(E))^*$ of $\Lambda^k(E)$. Thus, $\gamma \longrightarrow \varphi_\gamma$ is a linear map of $\Lambda^{n-k}(E)$ to $(\Lambda^k(E))^*$. We claim that it is one-to-one, i.e., that if $\varphi_\gamma(\alpha) = 0$ for all $\alpha \in \Lambda^k(E)$, then $\gamma = 0$.

Exercise 4.1.18 Prove this when $k = 0$.

Now suppose $k \geq 1$ and $\varphi_\gamma(\alpha) = 0$ for all $\alpha \in \Lambda^k(E)$. Then $\alpha \wedge \gamma = 0$ for all $\alpha \in \Lambda^k(E)$. Let $\{e_1, \ldots, e_n\}$ be a basis for E with dual basis $\{e^1, \ldots, e^n\}$ and write

$$\gamma = \sum_{i_1 < \cdots < i_{n-k}} \gamma_{i_1 \cdots i_{n-k}} e^{i_1} \wedge \cdots \wedge e^{i_{n-k}}.$$

We show that each coeffiecient is zero. Fix an increasing sequence $j_1 < \cdots < j_{n-k}$. Let $l_1 < \cdots < l_k$ be the remaining indices in $\{1, \ldots, n\}$ and consider $\alpha = e^{l_1} \wedge \cdots \wedge e^{l_k} \in \Lambda^k(E)$. Then, by assumption,

$$0 = \alpha \wedge \gamma = \sum_{i_1 < \cdots < i_{n-k}} \gamma_{i_1 \cdots i_{n-k}} e^{l_1} \wedge \cdots \wedge e^{l_k} \wedge e^{i_1} \wedge \cdots \wedge e^{i_{n-k}}.$$

Evaluating at $(e_{l_1}, \ldots, e_{l_k}, e_{j_1}, \ldots, e_{j_{n-k}})$ gives $\gamma_{j_1 \cdots j_{n-k}} = 0$. Thus, $\gamma = 0$.

We conclude that $\gamma \longrightarrow \varphi_\gamma$ is one-to-one. But since $\dim \Lambda^{n-k}(E) = \dim \Lambda^k(E) = \dim\ (\Lambda^k(E))^*$, it must be an isomorphism. Now, for each $\beta \in \Lambda^k(E)$, the map $\alpha \longrightarrow g(\alpha, \beta)$ is an element of $(\Lambda^k(E))^*$ and so there is a unique element of $\Lambda^{n-k}(E)$, which we denote $^*\beta$, such that $\varphi_{^*\beta}(\alpha) = g(\alpha, \beta)$ for every $\alpha \in \Lambda^k(E)$, i.e.,

$$\alpha \wedge {}^*\beta = g(\alpha, \beta)\omega$$

for every $\alpha \in \Lambda^k(E)$. Furthermore, $^*\beta = 0 \in \Lambda^{n-k}(E)$ implies $g(\alpha, \beta) = 0$ for every α and therefore (by nondegeneracy of g), $\beta = 0$. Consequently, $\beta \longrightarrow {}^*\beta$, which is clearly linear, is an isomorphism of $\Lambda^k(E)$ onto $\Lambda^{n-k}(E)$ with the required properties. ∎

Remark: For any $\beta \in \Lambda^k(E)$ it is customary to write

$$\|\beta\|^2 = g(\beta,\ \beta).$$

Thus, when $\alpha = \beta$, (4.1.6) becomes

$$\beta \wedge {}^*\beta = \|\beta\|^2\omega. \tag{4.1.7}$$

Some caution is in order here, however, since we have not assumed that our inner product is positive definite so that $\|\beta\|^2$ need not be positive.

We will need some formulas for computing Hodge duals. These are particularly simple relative to an oriented orthonormal basis $\{e_1, \ldots, e_n\}$ so we will consider these first. First observe that if $\beta \in \Lambda^0(E) =$, then ${}^*\beta \in \Lambda^n(E)$ and so is a real multiple of $\omega = e^1 \wedge \cdots \wedge e^n$.

Exercise 4.1.19 Show that, if $\beta \in \Lambda^0(E) =$, then

$$ {}^*\beta == \beta\omega = \beta e^1 \wedge \cdots \wedge e^n, $$

where $\{e^1, \ldots, e^n\}$ is the dual of the oriented orthonormal basis $\{e_1, \ldots, e_n\}$. In particular,

$$ {}^*1 = \omega = e^1 \wedge \cdots \wedge e^n. $$

By linearity of the Hodge star isomorphism it will suffice to describe the Hodge dual of any basis element $e^{i_1} \wedge \cdots \wedge e^{i_k}$, $1 \leq i_1 < \cdots < i_k \leq n$, in $\Lambda^k(E)$. We have already seen that $\|e^{i_1} \wedge \cdots \wedge e^{i_k}\|^2 = (-1)^m$, where m is the number of indices $i_1, \ldots, i_k$ with $g_{ii} = -1$. Thus,

$$ (e^{i_1} \wedge \cdots \wedge e^{i_k}) \wedge^* (e^{i_1} \wedge \cdots \wedge e^{i_k}) = (-1)^m e^1 \wedge \cdots \wedge e^n. $$

Since ${}^*(e^{i_1} \wedge \cdots \wedge e^{i_k})$ is uniquely determined by this property it will suffice to exhibit some $\gamma \in \Lambda^{n-k}(E)$ for which $(e^{i_1} \wedge \cdots \wedge e^{i_k}) \wedge \gamma = (-1)^m\omega$. In particular, if $k = n$,

$$ {}^*(e^1 \wedge \cdots \wedge e^n) = (-1)^s 1, $$

where s is the index of g.

Exercise 4.1.20 Show that, if $k = n - 1$, then

$$ {}^*(e^{i_1} \wedge \cdots \wedge e^{i_{n-1}}) = \pm e^{i_n} $$

where $\{i_1, \ldots, i_{n-1}, i_n\} = \{1, \ldots, n\}$ and one chooses the plus (minus) sign if $i_1 \cdots i_{n-1} i_n$ is an even (odd) permutation of $1 \cdots n$.

Finally, if $k < n - 1$ one can select $l_1, \ldots, l_{n-k}$ so that $i_1 \cdots i_k l_1 \cdots l_{n-k}$ is an even permutation of $1 \cdots n$. Then

$$ (e^{i_1} \wedge \cdots \wedge e^{i_k}) \wedge ((-1)^m e^{l_1} \wedge \cdots \wedge e^{l_{n-k}}) = (-1)^m \omega $$

so

$$ {}^*(e^{i_1} \wedge \cdots \wedge e^{i_k}) = (-1)^m e^{l_1} \wedge \cdots \wedge e^{l_{n-k}}, \tag{4.1.8} $$

where, again, m is the number of indices $i_1, \ldots, i_k$ with $g_{ii} = -1$.

Exercise 4.1.21 Verify the following concrete examples:

(a) Let $E = \ ^3$ with its standard orientation and (positive definite) inner product and let $\{e_1, e_2, e_3\}$ be an oriented orthonormal basis.

Then

$$
\begin{array}{lll}
{}^*e^1 = e^2 \wedge e^3 & {}^*e^2 = -e^1 \wedge e^3 & {}^*e^3 = e^1 \wedge e^2 \\
{}^*(e^1 \wedge e^2) = e^3 & {}^*(e^2 \wedge e^3) = e^1 & {}^*(e^1 \wedge e^3) = -e^2 \\
{}^*(e^1 \wedge e^2 \wedge e^3) = 1. & &
\end{array}
$$

(b) Let $E = \ {}^4$ with its standard orientation and (positive definite) inner product and let $\{e_1, e_2, e_3, e_4\}$ be an oriented orthonormal basis. Then

$$
\begin{array}{ll}
{}^*e^1 = e^2 \wedge e^3 \wedge e^4 & {}^*(e^1 \wedge e^3) = -e^2 \wedge e^4 \\
{}^*(e^1 \wedge e^3 \wedge e^4) = e^2 & {}^*(e^1 \wedge e^2 \wedge e^4) = -e^3 \\
{}^*(e^1 \wedge e^2 \wedge e^3 \wedge e^4) = 1. &
\end{array}
$$

(c) Let $E = \ {}^{1,3}$ with its standard orientation and Minkowski inner product and let $\{e_0, e_1, e_2, e_3\}$ be an oriented orthonormal basis (assume, as usual, that $g(e_0, e_0) = 1$ and $g(e_i, e_i) = -1$ for $i = 1, 2, 3$). Then

$$
\begin{array}{ll}
{}^*e^2 = -e^0 \wedge e^1 \wedge e^3 & {}^*(e^1 \wedge e^2) = e^0 \wedge e^3 \\
{}^*(e^0 \wedge e^1 \wedge e^3) = -e^2 & {}^*(e^0 \wedge e^1 \wedge e^2 \wedge e^3) = -1.
\end{array}
$$

To describe the Hodge dual in an arbitrary oriented basis $\{\hat{e}_1, \ldots, \hat{e}_n\}$ with dual basis $\{\hat{e}^1, \ldots, \hat{e}^n\}$ we will use the Levi-Civita symbol

$$
\varepsilon_{j_1 \cdots j_n} = \begin{cases} 1, & \text{if } j_1 \cdots j_n \text{ is an even permutation of } 1 \cdots n \\ -1, & \text{if } j_1 \cdots j_n \text{ is an odd permutation of } 1 \cdots n \\ 0, & \text{otherwise} \end{cases}
$$

and borrow a result from the theory of determinants: For any $n \times n$ matrix $A = (A_i{}^j)_{i,j=1,\ldots,n}$,

$$
\varepsilon_{j_1 \cdots j_n} A_{i_1}{}^{j_1} \cdots A_{i_n}{}^{j_n} = \varepsilon_{i_1 \cdots i_n} (\det A). \tag{4.1.9}
$$

Now let $\{e_1, \ldots, e_n\}$ be an oriented orthonormal basis for E with dual basis $\{e^1, \ldots, e^n\}$. We write $\hat{e}_j = A^i{}_j e_i,\ j = 1, \ldots, n$, and denote by $(A_i{}^j)$ the inverse of the matrix $(A^i{}_j)$. Then $\hat{e}^j = A_i{}^j e^i$. In addition, we write $g_{ij} = g(e_i, e_j)$ and $\hat{g}_{ij} = g(\hat{e}_i, \hat{e}_j)$ for $i, j = 1, \ldots, n$, denote by (g^{ij}) and $(\hat{g}^{ij})$ the inverses of the matrices (g_{ij}) and $(\hat{g}_{ij})$, respectively, and recall (from Exercise 4.1.15) that

$$
\det(A^i{}_j) = |\det(\hat{g}_{ij})|^{\frac{1}{2}}. \tag{4.1.10}
$$

Now, for any $\beta \in \Lambda^k(E)$ we write

$$\beta = \frac{1}{k!}\beta_{i_1\cdots i_k} e^{i_1} \wedge \cdots \wedge e^{i_k} = \frac{1}{k!}\hat{\beta}_{i_1\cdots i_k} \hat{e}^{i_1} \wedge \cdots \wedge \hat{e}^{i_k},$$

and raise the indices as in (4.1.4) to obtain

$$\beta^{j_1\cdots j_k} = g^{i_1 j_1} \cdots g^{i_k j_k} \beta_{i_1\cdots i_k} \quad \text{and} \quad \hat{\beta}^{j_1\cdots j_k} = \hat{g}^{i_1 j_1} \cdots \hat{g}^{i_k j_k} \hat{\beta}_{i_1\cdots i_k}.$$

Exercise 4.1.22 Show that,

$$^*\beta = \frac{1}{(n-k)!} \frac{|\det(\hat{g}_{ij})|^{\frac{1}{2}}}{k!} \varepsilon_{i_1\cdots i_k\ j_1\cdots j_{n-k}}\ \hat{\beta}^{i_1\cdots i_k}\ \hat{e}^{j_1} \wedge \cdots \wedge \hat{e}^{j_{n-k}}.$$

Hint: Convert the right-hand side to unhatted coordinates by using Exercises 4.1.16 and 4.1.17 as well as (4.1.9) and (4.1.10).

Theorem 4.1.15 *Let E be an n-dimensional real vector space with an orientation μ and an inner product g of index s. Let k be an integer with $0 \leq k \leq n$. Then the corresponding Hodge star isomorphism satisfies*

$$^{**}\beta = (-1)^{k(n-k)+s}\beta$$

for every $\beta \in \Lambda^k(E)$.

Proof: Let $\{e_1, \ldots, e_n\}$ be an oriented orthonormal basis for E with dual basis $\{e^1, \ldots, e^n\}$. Then there are precisely s indices among $\{1, \ldots, n\}$ for which $g_{ii} = g(e_i, e_i) = -1$.

Exercise 4.1.23 Show that if $\beta \in \Lambda^0(E)$, then $^{**}\beta = (-1)^s\beta$.

Now suppose $1 \leq k \leq n$. By linearity it will suffice to prove our result for $\beta = e^{i_1} \wedge \cdots \wedge e^{i_k}$, where $1 \leq i_1 < \cdots < i_k \leq n$ is some fixed increasing index sequence.

Exercise 4.1.24 Prove the result for $k = n - 1$. **Hint:** Exercise 4.1.20.

Now suppose $k < n-1$ and let m be the number of indices among $\{i_1, \ldots, i_k\}$ for which $g_{ii} = -1$. Then $^*\beta = (-1)^m e^{l_1} \wedge \cdots \wedge e^{l_{n-k}}$, where $i_1 \cdots i_k\ l_1 \cdots l_{n-k}$ is an even permutation of $1 \cdots n$. Thus,

$$^{**}\beta = (-1)^{m\,*}\left(e^{l_1} \wedge \cdots \wedge e^{l_{n-k}}\right).$$

But since $\{i_1, \ldots, i_k, l_1, \ldots, l_{n-k}\} = \{1, \ldots, n\}$,

$$^*(e^{l_1} \wedge \cdots \wedge e^{l_{n-k}}) = (-1)^{m'} e^{i_1} \wedge \cdots \wedge e^{i_k}$$

for some appropriate integer m' which we now determine. Notice that $i_1 \cdots i_k$ $l_1 \cdots l_{n-k}$ is related to $l_1 \cdots l_{n-k}$ $i_1 \cdots i_k$ by a permutation of sign $(-1)^{k(n-k)}$ (move each of the k indices i to the left over all of the $n-k$ indices l in the order $i_1, \ldots, i_k$). There are $s - m$ indices among $\{l_1, \ldots, l_{n-k}\}$ for which $g_{ll} = -1$. Then

$$^*\left(e^{l_1} \wedge \cdots \wedge e^{l_{n-k}}\right) = (-1)^{s-m}(-1)^{k(n-k)} e^{i_1} \wedge \cdots \wedge e^{i_k}$$

so $m' = k(n-k) + s - m$. Thus

$$^{**}\beta = (-1)^m (-1)^{s-m} (-1)^{k(n-k)} e^{i_1} \wedge \cdots \wedge e^{i_k} = (-1)^{k(n-k)+s}\beta$$

as required. ∎

Corollary 4.1.16 *If E is an oriented vector space with a positive definite inner product g, then each Hodge star isomorphism $^*\colon \Lambda^k(E) \longrightarrow \Lambda^{n-k}(E), 0 \leq k \leq n$, is an isometry, i.e.,*

$$g\left({}^*\alpha, {}^*\beta\right) = g(\alpha, \beta)$$

for all $\alpha, \beta \in \Lambda^k(E)$.

Proof: If $\dim E = n$ and ω is the volume form on E, then

$$\begin{aligned} g\left({}^*\alpha, {}^*\beta\right)\omega &= {}^*\alpha \wedge \left({}^{**}\beta\right) = (-1)^{k(n-k)}\, {}^*\alpha \wedge \beta \\ &= (-1)^{k(n-k)}(-1)^{(n-k)k} \beta \wedge {}^*\alpha \\ &= g(\beta, \alpha)\omega \\ &= g(\alpha, \beta)\omega. \end{aligned}$$

Since $\{\omega\}$ is a basis for $\Lambda^n(E)$, $g\left({}^*\alpha, {}^*\beta\right) = g(\alpha, \beta)$. ∎

Exercise 4.1.25 Show that, if the index of g is $s = 1$, then

$$g\left({}^*\alpha, {}^*\beta\right) = -g(\alpha, \beta).$$

Exercise 4.1.26 Show that, if $\dim E = 3$ and g is positive definite, then $^{**}\beta = \beta$ for any $\beta \in \Lambda^k(E)$, $0 \leq k \leq 3$.

Exercise 4.1.27 Show that, if $\dim E = 4$ and g is positive definite, then $^{**}\beta = (-1)^k\beta$ for any $\beta \in \Lambda^k(E)$, $0 \leq k \leq 4$.

Exercise 4.1.28 Show that, if $\dim E = 4$ and g has index one, then $^{**}\beta = (-1)^{k+1}\beta$ for any $\beta \in \Lambda^k(E)$, $0 \leq k \leq 4$.

4.2 Vector-Valued Forms

Most of the forms of interest to us will take values, not in , but in some finite dimensional real vector space $\mathcal{V}$ (e.g., , , or the Lie algebra $\mathcal{G}$ of some matrix Lie group G). With the exception of wedge products, which we discuss in some detail, virtually all of the material in Section 4.1 generalizes immediately to this context by simply doing everything componentwise with respect to some basis for $\mathcal{V}$.

As before we let E denote some n-dimensional real vector space. For some purposes (to be specified as we proceed), E will be assumed to have an orientation μ and an inner product g. Our forms will be defined on E and will take values in some m-dimensional real vector space $\mathcal{V}$. When appropriate we will also assume that $\mathcal{V}$ has an inner product h (e.g., when $\mathcal{V}$ is a Lie algebra, h will generally arise from the Killing form). We will use $\{e_1, \ldots, e_n\}$ to denote a generic basis for E and $\{T_1, \ldots, T_m\}$ will be a basis for $\mathcal{V}$. A map

$$A : E \times \overset{k}{\cdots} \times E \longrightarrow \mathcal{V}$$

is **multilinear** if, for each i with $1 \leq i \leq k$ and each $a \in$, $A(v_1, \ldots, v_i + v_i', \ldots, v_k) = A(v_1, \ldots, v_i, \ldots, v_k) + A(v_1, \ldots, v_i', \ldots, v_k)$ and $A(v_1, \ldots, av_i, \ldots, v_k) = aA(v_1, \ldots, v_i, \ldots, v_k)$ for all $v_1, \ldots, v_i, v_i', \ldots, v_k \in E$. The set $\mathcal{T}^k(E, \mathcal{V})$ of all such is a real vector space with the obvious pointwise operations:

$$(A + B)(v_1, \ldots, v_k) = A(v_1, \ldots, v_k) + B(v_1, \ldots, v_k)$$
$$(aA)(v_1, \ldots, v_k) = a\Big(A(v_1, \ldots, v_k)\Big).$$

For convenience, we will take $\mathcal{T}^0(E, \mathcal{V}) = \mathcal{V}$. The elements of $\mathcal{T}^k(E, \mathcal{V})$ are called **covariant $\mathcal{V}$-valued tensors** of **rank** k (or simply **$\mathcal{V}$-valued k-tensors**) on E. If $T : E_1 \longrightarrow E_2$ is a linear transformation, then, for any k, we define the **pullback** map

$$T^* : \mathcal{T}^k(E_2, \mathcal{V}) \longrightarrow \mathcal{T}^k(E_1, \mathcal{V})$$

by

$$(T^* A)(v_1, \ldots, v_k) = A\big(T(v_1), \ldots, T(v_k)\big)$$

for any $A \in \mathcal{T}^k(E_2, \mathcal{V})$.

If $\{T_1, \ldots, T_m\}$ is any basis for $\mathcal{V}$, then any $A \in \mathcal{T}^k(E, \mathcal{V})$ can be written uniquely as

$$A = A^1 T_1 + \cdots + A^m T_m = A^i T_i,$$

where $A^i \in \mathcal{T}^k(E)$ for each $i = 1, \ldots, m$. The linear operations on $\mathcal{T}^k(E, \mathcal{V})$ as well as the pullback corresponding to any linear transformation can clearly all be performed componentwise relative to any such basis for $\mathcal{V}$.

An $A \in \mathcal{T}^k(E, \mathcal{V})$ is said to be **skew-symmetric** if it satisfies the three equivalent conditions (a), (b) and (c) stated in Exercise 4.1.1 and the set $\Lambda^k(E, \mathcal{V})$ of all such is a linear subspace of $\mathcal{T}^k(E, \mathcal{V})$ (again, we take $\Lambda^0(E, \mathcal{V}) = \mathcal{T}^0(E, \mathcal{V}) = \mathcal{V}$). The elements of $\Lambda^k(E, \mathcal{V})$ are called **$\mathcal{V}$-valued k-forms** on E. If $\alpha \in \Lambda^k(E, \mathcal{V})$ and $\alpha = \alpha^i T_i$ for some basis

$\{T_1, \ldots, T_m\}$ for $\mathcal{V}$, then each $\alpha^i \in \Lambda^k(E)$. If E has an orientation μ and an inner product g (and therefore an associated Hodge star) one can define Hodge star operators

$$^*: \Lambda^k(E,\ \mathcal{V}) \longrightarrow \Lambda^{n-k}(E, \mathcal{V})$$

componentwise:

$$^*\alpha = \ ^*(\alpha^1 T_1 + \cdots + \alpha^m T_m) = \ ^*\alpha^1 T_1 + \cdots + \ ^*\alpha^m T_m. \tag{4.2.1}$$

Exercise 4.2.1 Show that this definition does not depend on the choice of basis for $\mathcal{V}$.

If we assume that $\mathcal{V}$ also has an inner product h, then, together with g and the induced inner products on each $\Lambda^k(E)$ (also denoted g) we can define inner products, denoted (gh), on each $\Lambda^k(E, \mathcal{V})$ as follows: Let $\{T_1, \ldots, T_m\}$ be a basis for $\mathcal{V}$, write $\alpha, \beta \in \Lambda^k(E, \mathcal{V})$ as $\alpha = \alpha^i T_i$ and $\beta = \beta^j T_j$ and set $h_{ij} = h(T_i, T_j)$ for $i, j = 1, \ldots, m$. Now define

$$(gh)(\alpha, \beta) = h_{ij}\, g(\alpha^i, \beta^j). \tag{4.2.2}$$

Exercise 4.2.2 Show that this definition does not depend on the choice of basis for $\mathcal{V}$.

Exercise 4.2.3 Show that, if $\alpha, \beta \in \Lambda^0(E, \mathcal{V}) = \mathcal{V}$, then (gh) coincides with h.

Exercise 4.2.4 Show that, if h is positive definite and $\{T_1, \ldots, T_m\}$ is h-orthonormal, then $(gh)(\alpha, \beta) = g(\alpha^1, \beta^1) + \cdots + g(\alpha^m, \beta^m)$. In particular, writing $\|\alpha\|^2$ for $g(\alpha, \alpha)$,

$$\|\alpha\|^2 = \|\alpha^1\|^2 + \cdots + \|\alpha^m\|^2.$$

An $A \in \mathcal{T}^k(E, \mathcal{V})$ is said to be **symmetric** if it takes the same value whenever two of its arguments are interchanged (or, equivalently, whenever its arguments are permuted). This is clearly the case if and only if each of its components A^i relative to any basis for $\mathcal{V}$ is symmetric. Moreover, the set of all such is a linear subspace of $\mathcal{T}^k(E, \mathcal{V})$.

Tensor and wedge products for real-valued forms depend, for their definition, on the multiplicative structure of . Unless some such bilinear pairing is available in $\mathcal{V}$ one cannot define such products for $\mathcal{V}$-valued tensors. In the cases of interest to us, however, such pairings present themselves quite naturally (e.g., complex multiplication in , or the Lie bracket in a Lie algebra $\mathcal{G}$). We now show how to use such additional structures to obtain useful notions of the tensor and wedge product for vector-valued tensors.

Suppose $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ are all real vector spaces and that one is given a bilinear map $\rho : \mathcal{U} \times \mathcal{V} \longrightarrow \mathcal{W}$. Let $A \in \mathcal{T}^k(E, \mathcal{U})$ and $B \in \mathcal{T}^l(E, \mathcal{V})$. We define the **$\rho$-tensor product** $A \otimes_\rho B \in \mathcal{T}^{k+l}(E, \mathcal{W})$ by

$$\begin{aligned}\big(A \otimes_\rho B\big)\,\big(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}\big) \\ = \rho\Big(A\ (v_1, \ldots, v_k),\ B(v_{k+1}, \ldots, v_{k+l})\Big).\end{aligned} \tag{4.2.3}$$

Note that, if $k = l = 0$, then $A \in \mathcal{U},\ B \in \mathcal{V}$ and $A \otimes_\rho B = \rho(A, B)$.

Exercise 4.2.5 Verify that $A \otimes_\rho B$ is, indeed, in $\mathcal{T}^{k+l}(E, \mathcal{W})$ and that $\otimes_\rho$ has all of the properties of $\otimes$ described in Exercise 4.1.2.

If $\alpha \in \Lambda^k(E, \mathcal{U})$ and $\beta \in \Lambda^l(E, \mathcal{V})$, then their **$\rho$-wedge product** $\alpha \wedge_\rho \beta \in \Lambda^{k+l}(E, \mathcal{W})$ is defined by

$$\begin{aligned}\big(\alpha \wedge_\rho \beta\big)\big(v_1, \ldots, v_{k+l}\big) \\ = \frac{1}{k!l!}\sum_\sigma (-1)^\sigma \big(\alpha \otimes_\rho \beta\big)\Big(v_{\sigma(1)}, \ldots, v_{\sigma(k+l)}\Big),\end{aligned} \tag{4.2.4}$$

where the sum is over all permutations $\sigma \in S_{k+l}$ of $\{1, \ldots, k+l\ \}$.

Exercise 4.2.6 Let $\{U_1, \ldots, U_q\}$ be a basis for $\mathcal{U}$ with $\alpha = \alpha^i U_i$ and let $\{T_1, \ldots, T_m\}$ be a basis for $\mathcal{V}$ with $\beta = \beta^j T_j$. Show that

$$\alpha \wedge_\rho \beta = \sum_{i=1}^{q}\sum_{j=1}^{m}(\alpha^i \wedge \beta^j)\rho(U_i, T_j).$$

Remark: There is no reason to expect that the $\rho(U_i, T_j),\ i = 1, \ldots, q$, $j = 1, \ldots, m$, should constitute a basis for $\mathcal{W}$. Indeed, if $\mathcal{U} = \mathcal{V} = \mathcal{W}$, then this clearly cannot be the case (there are too many of them). Thus, the $\alpha^i \wedge \beta^j$ cannot be regarded as components of $\alpha \wedge_\rho \beta$.

Exercise 4.2.7 Let $\mathcal{U} = \mathcal{V} = \mathcal{W} =$ (thought of as a 2-dimensional real vector space) and let $\rho\colon\ \times\ \longrightarrow$ be complex multiplication. Let $\alpha \in \Lambda^k(E,\ )$ and $\beta \in \Lambda^l(E,\ )$. As a basis for (over) take $\{T_1, T_2\} = \{\mathbf{1}, \boldsymbol{i}\}$ and write $\alpha = \alpha^1\mathbf{1} + \alpha^2\boldsymbol{i} = \alpha^1 + \alpha^2\boldsymbol{i}$ and $\beta = \beta^1 + \beta^2\boldsymbol{i}$. Show that

$$\begin{aligned}\alpha \wedge_\rho \beta &= (\alpha^1 + \alpha^2\boldsymbol{i}) \wedge_\rho (\beta^1 + \beta^2\boldsymbol{i}) \\ &= (\alpha^1 \wedge \beta^1 - \alpha^2 \wedge \beta^2) + (\alpha^1 \wedge \beta^2 + \alpha^2 \wedge \beta^1)\boldsymbol{i}.\end{aligned}$$

Remark: According to Exercise 4.2.7 one can wedge complex-valued forms in the same way one multiplies complex numbers, but with the real and imaginary parts multiplied by ordinary wedge. One obtains an analogous result for quaternion-valued forms when is regarded as a 4-dimensional real vector space with basis $\{\mathbf{1}, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ and $\rho :\ \times\ \longrightarrow$ is quaternion multiplication.

Next we consider the case in which $\mathcal{U} = \mathcal{V} = \mathcal{W} = \mathcal{G}$, where $\mathcal{G}$ is the Lie algebra of some matrix Lie group G and we identify $\mathcal{G}$ also with a real vector space of (possibly complex) matrices. As the bilinear map ρ we take the Lie bracket $[\ ,\ ] : \mathcal{G} \times \mathcal{G} \longrightarrow \mathcal{G}$, i.e., the matrix commutator:

$$\rho(A, B) = [A, B] = AB - BA$$

for all $A, B \in \mathcal{G}$. With this choice of ρ it is customary to write

$$\alpha \wedge_\rho \beta = [\alpha, \beta]$$

for any $\mathcal{G}$-valued forms α and β and we will adhere to this custom. Thus, for any $\alpha \in \Lambda^k(E, \mathcal{G})$ and $\beta \in \Lambda^l(E, \mathcal{G})$ and any $v_1, \ldots, v_{k+l}$ in E we have

$$\begin{aligned} &[\alpha, \beta]\big(v_1, \ldots, v_{k+l}\big) \\ &= \frac{1}{k!l!} \sum_{\sigma \in S_{k+l}} (-1)^\sigma \Big[\alpha\Big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\Big), \beta\Big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\Big)\Big]. \end{aligned} \tag{4.2.5}$$

If $\{T_1, \ldots, T_m\}$ is any basis for $\mathcal{G}$ with structure constants C_{ij}^k (so that $[T_i, T_j] = C_{ij}^k T_k,\ i, j = 1, \ldots, m$), then, by Exercise 4.2.6,

$$\begin{aligned} [\alpha, \beta] &= [\alpha^i T_i,\ \beta^j T_j] \\ &= (\alpha^i \wedge \beta^j)[T_i,\ T_j] = C_{ij}^k (\alpha^i \wedge \beta^j) T_k. \end{aligned} \tag{4.2.6}$$

Exercise 4.2.8 Prove each of the following:

(a) $[\beta, \alpha] = (-1)^{kl+1}[\alpha, \beta]$.

(b) For any $\gamma \in \Lambda^r(E, \mathcal{G})$,

$$(-1)^{kr}\big[\,[\alpha, \beta], \gamma\big] + (-1)^{rl}\big[\,[\gamma, \alpha], \beta\big] + (-1)^{kl}\big[\,[\beta, \gamma], \alpha\big] = 0.$$

Hint: For (b) use the Jacobi identity (page 15).

For computational purposes it is convenient to introduce another wedge product on $\mathcal{G}$-valued forms based on another bilinear pairing ρ and from which $[\alpha, \beta]$ is easily obtained. Toward this end we take ρ to be simply matrix multiplication on $\mathcal{G} \times \mathcal{G}$. Note that, since a matrix Lie algebra $\mathcal{G}$ is generally not closed under matrix multiplication, ρ must be regarded as a map from $\mathcal{G} \times \mathcal{G}$ to some $\mathcal{GL}$. We will denote the corresponding wedge product simple $\alpha \wedge \beta$ so that

$$(\alpha \wedge \beta)\big(v_1, \ldots, v_{k+l}\big)$$
$$= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} (-1)^{\sigma} \alpha\Big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\Big)\, \beta\Big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\Big). \tag{4.2.7}$$

Exercise 4.2.9 Show that, if $\alpha \in \Lambda^k(E, \mathcal{G})$ and $\beta \in \Lambda^l(E, \mathcal{G})$, then

$$[\alpha, \beta] = \alpha \wedge \beta - (-1)^{kl}\, \beta \wedge \alpha. \tag{4.2.8}$$

Remark: In particular, for $\mathcal{G}$-valued 1-forms α and $\beta, [\alpha, \beta] = \alpha \wedge \beta + \beta \wedge \alpha$ so that $[\beta, \alpha] = [\alpha, \beta]$. The same is true, of course, if either α or β has any odd rank. For a single $\mathcal{G}$-valued form α of odd rank one obtains

$$[\alpha, \alpha] = 2\alpha \wedge \alpha \quad \big(\alpha \in \Lambda^{2k+1}(E, \mathcal{G})\big). \tag{4.2.9}$$

Although it is the wedge product [,] that arises most naturally in the context $\mathcal{G}$-valued forms it is, as we have seen, readily obtained from $\wedge$ and, as we now show, $\alpha \wedge \beta$ is particularly easy to compute. The idea is as follows: Any $\alpha \in \Lambda^k(E, \mathcal{G})$ can be formally identified with a matrix of k-forms that are -, -, or -valued by selecting a basis $\{T_1, \ldots, T_m\}$ for $\mathcal{G}$ and formally multiplying each entry of T^i by α^i and adding (we will illustrate with a concrete example shortly). Then evaluating α at $(v_1, \ldots, v_k)$ amounts to evaluating each entry k-form at $(v_1, \ldots, v_k)$. The result is a matrix in $\mathcal{G}$. If we express both α and β in this way as matrices of ordinary forms, then $\alpha \wedge \beta$ can be computed by simply forming the matrix product of α and β, but with the entries multiplied by the ordinary (real, complex, or quaternionic) wedge product. To see this we compute as follows: Let the matrix representations for α and β be (α^{ij}) and (β^{ij}). Then

$$\begin{pmatrix} \alpha^{11} & \alpha^{12} & \cdots \\ & & \\ \vdots & & \end{pmatrix} \begin{pmatrix} \beta^{11} & \cdots \\ \beta^{21} & \\ \vdots & \end{pmatrix} \big(v_1, \ldots, v_{k+l}\big)$$
$$= \begin{pmatrix} \alpha^{11} \wedge \beta^{11} + \alpha^{12} \wedge \beta^{21} + \cdots & \cdots \\ \vdots & \end{pmatrix} \big(v_1, \ldots, v_{k+l}\big)$$
$$= \begin{pmatrix} \sum_j \alpha^{1j} \wedge \beta^{j1} & \cdots \\ \vdots & \end{pmatrix} \big(v_1, \ldots, v_{k+l}\big)$$
$$= \begin{pmatrix} \sum_j \big(\alpha^{1j} \wedge \beta^{j1}\big)\big(v_1, \ldots, v_{k+l}\big) & \cdots \\ \vdots & \end{pmatrix}$$

$$= \begin{pmatrix} \sum_j \left\{ \frac{1}{k!\,l!} \sum_\sigma (-1)^\sigma \alpha^{1j}\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) \beta^{j1}\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\right) \right\} \cdots \\ \vdots \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{k!\,l!} \sum_\sigma (-1)^\sigma \left\{ \sum_j \alpha^{1j}\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) \beta^{j1}\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\right) \right\} \cdots \\ \vdots \end{pmatrix}$$

$$= \frac{1}{k!\,l!} \sum_\sigma (-1)^\sigma \begin{pmatrix} \sum_j \alpha^{1j}\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) \beta^{j1}\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\right) \cdots \\ \vdots \end{pmatrix}$$

$$= \frac{1}{k!\,l!} \sum_\sigma (-1)^\sigma \alpha\left(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\right) \beta\left(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\right)$$

$$= (\alpha \wedge \beta)\big(v_1, \ldots, v_{k+l}\big),$$

as promised.

We mention one concrete example that has already come up in the context of $SU(2)$-Yang-Mills-Higgs theory in Section 2.5. Let E be any n-dimensional real vector space with an orientation μ and an innerproduct g. For $\mathcal{V}$ we take the Lie algebra $su(2)$ of all 2×2 skew-Hermitian, tracefree matrices. We have defined a (positive definite) inner product h on $su(2)$ by $h(A,B) = -2$ trace(AB). The basis $\{T_1, T_2, T_3\}$ for $su(2)$ consisting of

$$T_1 = -\frac{1}{2}\,\boldsymbol{i}\,\sigma_1 \quad T_2 = -\frac{1}{2}\,\boldsymbol{i}\,\sigma_2 \quad T_3 = -\frac{1}{2}\,\boldsymbol{i}\,\sigma_3,$$

where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -\boldsymbol{i} \\ \boldsymbol{i} & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli spin matrices is h-orthonormal, i.e., $h_{ij} = h(T_i, T_j) = \delta_{ij}$ for $i, j = 1, 2, 3$. If $\alpha, \beta \in \Lambda^k(E, su(2))$, then

$$\alpha = \alpha^1\, T_1 + \alpha^2\, T_2 + \alpha^3\, T_3 = -\frac{1}{2} \begin{pmatrix} \alpha^3\, \boldsymbol{i} & \alpha^2 + \alpha^1\, \boldsymbol{i} \\ -\alpha^2 + \alpha^1\, \boldsymbol{i} & -\alpha^3\, \boldsymbol{i} \end{pmatrix} \tag{4.2.10}$$

and similarly for β. The Hodge dual ${}^*\beta$ is computed componentwise so

$${}^*\beta = {}^*\beta^1\, T_1 + {}^*\beta^2\, T_2 + {}^*\beta^3\, T_3 = -\frac{1}{2} \begin{pmatrix} {}^*\beta^3\, \boldsymbol{i} & {}^*\beta^2 + {}^*\beta^1\, \boldsymbol{i} \\ -{}^*\beta^2 + {}^*\beta^1\, \boldsymbol{i} & -{}^*\beta^3\boldsymbol{i} \end{pmatrix}.$$

Exercise 4.2.10 Compute $\alpha \wedge {}^*\beta$ and show that

$$-2\,\mathrm{trace}(\alpha \wedge {}^*\beta) = (gh)\,(\alpha, \beta)\,\omega,$$

where ω is the volume form for E determined by μ and g. In particular,

$$-2\,\mathrm{trace}(\alpha \wedge {}^*\alpha) = \|\alpha\|^2\,\omega.$$

If $\alpha = \alpha^i\,T_i$, then each α^i is a real-valued k-form on E so, if $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis for E with dual basis $\{e^1, \ldots, e^n\}$, it can be written $\alpha^i = \frac{1}{k!}\,\alpha^i_{i_1\cdots i_k}\,e^{i_1} \wedge \cdots \wedge e^{i_k}$. Thus, we can write

$$\alpha = \frac{1}{k!}\,\alpha_{i_1\cdots i_k}\,e^{i_1} \wedge \cdots \wedge e^{i_k},$$

where

$$\alpha_{i_1\cdots i_k} = \alpha^i_{i_1\cdots i_k}\,T_i$$

are $su(2)$-valued functions. Raising the indices componentwise as in (4.1.4) we obtain

$$\alpha^{j_1\cdots j_k} = g^{i_1\,j_1}\cdots g^{i_k\,j_k}\,\alpha_{i_1\cdots i_k}.$$

Similarly, writing $\beta = \beta_{j_1\cdots j_k}\,e^{j_1} \wedge \cdots \wedge e^{j_k}$, where the $\beta_{j_1\cdots j_k}$ are $su(2)$-valued functions, the summation convention gives

$$\alpha_{i_1\cdots i_k}\,\beta^{i_1\cdots i_k}$$

as a sum of matrix products, i.e., as a matrix.

Exercise 4.2.11 Show that

$$-2\,\mathrm{trace}\Big(\alpha_{i_1\cdots\,i_k}\,\beta^{i_1\cdots\,i_k}\Big) = k!\,(gh)\,(\alpha, \beta)$$

so, in particular,

$$-\frac{1}{k!}\,\mathrm{trace}\Big(\alpha_{i_1\cdots i_k}\,\alpha^{i_1\cdots i_k}\Big) = \frac{1}{2}\|\alpha\|^2.$$

Combining this with Exercise 4.2.10 gives

$$-\mathrm{trace}(\alpha \wedge {}^*\alpha) = \frac{1}{2}\|\alpha\|^2\omega = -\frac{1}{k!}\mathrm{trace}\Big(\alpha_{i_1\cdots i_k}\;\alpha^{i_1\cdots i_k}\Big)\omega. \qquad (4.2.11)$$

4.3 Differential Forms

Roughly, a "differential k-form" $\boldsymbol{\alpha}$ on a manifold X is a "smooth assignment" to each $x \in X$ of a k-form $\boldsymbol{\alpha}(x) \in \Lambda^k(T_x(X))$. We have seen that, in the $k = 1$ case, there are a number of equivalent ways in which this notion of a "smooth assignment" can be made precise. A map $\boldsymbol{\Theta}$ that assigns to each $x \in X$ a covector (1-form) $\boldsymbol{\Theta}(x) \in T^*_x(X) = \Lambda^1(T_x(X))$ is smooth if its components $\Theta_i(x) = \boldsymbol{\Theta}(x)(\frac{\partial}{\partial x^i}|_x)$ are C^∞ for any chart (U, φ). Equivalently,

$\boldsymbol{\Theta}$ can be regarded as a $C^\infty(X)$-module homomorphism from $\mathcal{X}(X)$ to $C^\infty(X)$ so "smoothness" means that $\boldsymbol{\Theta}$ operates on C^∞ vector fields to give C^∞ real-valued functions. Yet again, a smooth 1-form can be identified with a smooth cross-section of the cotangent bundle $T^*(X)$ (or, if you prefer, a smooth, equivariant map on the frame bundle). All of these generalize quite nicely. We will formulate all of the appropriate versions of the definition and then leave it to the reader to prove their equivalence.

Remark: Notice that each of the following equivalent definitions has an obvious modification in which "smooth" is replaced by some less stringent regularity condition (e.g., continuous, continuous almost everywhere, etc.). For the integration theory we construct in Section 4.6 we will want to consider "measurable" and "integrable" k-forms on a manifold.

1. A **differential *k*-form $\boldsymbol{\alpha}$** on a manifold X is an assignment to each $x \in X$ of a k-form $\boldsymbol{\alpha}(x) = \boldsymbol{\alpha}_x \in \Lambda^k(T_x(X))$ on the tangent space at x which is smooth in the following sense: Let (U, φ) be a chart for X with coordinate functions $x^1, \ldots, x^n$. Then at each $x \in U$, $\boldsymbol{\alpha}(x) = \frac{1}{k!}\, \alpha_{i_1 \cdots i_k}(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, where $\alpha_{i_1 \cdots i_k}(x) = \boldsymbol{\alpha}(x)\left(\frac{\partial}{\partial x^{i_1}}|_x, \ldots, \frac{\partial}{\partial x^{i_k}}|_x\right)$. We say that $\boldsymbol{\alpha}$ is smooth if these component functions $\alpha_{i_1 \cdots i_k}(x)$ are C^∞ on U for all charts in some atlas for X. The set of all such is denoted $\Lambda^k(X)$ (with $\Lambda^0(X) = C^\infty(X)$) and has the obvious pointwise structure of a $C^\infty(X)$-module.

2. Every $\boldsymbol{\alpha} \in \Lambda^k(X)$ gives rise to a map $\boldsymbol{\alpha} : \mathcal{X}(X) \times \overset{k}{\cdots} \times \mathcal{X}(X) \longrightarrow C^\infty(X)$ defined by $\boldsymbol{\alpha}(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_k)(x) = \boldsymbol{\alpha}(x)(\boldsymbol{V}_1(x), \ldots, \boldsymbol{V}_k(x))$ that is skew-symmetric and $C^\infty(X)$-multilinear. Conversely, any skew-symmetric, $C^\infty(X)$-multilinear map $A : \mathcal{X}(X) \times \overset{k}{\cdots} \times \mathcal{X}(X) \longrightarrow C^\infty(X)$ determines a unique $\boldsymbol{\alpha} \in \Lambda^k(X)$ with $\boldsymbol{\alpha}(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_k) = A(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_k)$ and this one-to-one correspondence is a $C^\infty(X)$-module isomorphism. Thus, a **differential *k*-form $\boldsymbol{\alpha}$** on X can be identified with a skew-symmetric, $C^\infty(X)$-multilinear map $\boldsymbol{\alpha} : \mathcal{X}(X) \times \cdots \times \mathcal{X}(X) \longrightarrow C^\infty(X)$.

3. Let $GL(n, \ ) \hookrightarrow L(X) \longrightarrow X$ be the linear frame bundle of X. Let $\mathcal{V}$ be the real vector space $\Lambda^k(\ ^n)$ of skew-symmetric, k-multilinear forms on $\ ^n$ and define a representation $\rho : GL(n, \ ) \longrightarrow GL(\Lambda^k(\ ^n))$ by $(\rho(g)(\alpha))(v_1, \ldots, v_k) = \alpha(g^T v_1, \ldots, g^T v_k)$, where each $v_i \in \ ^n$ is written as a column matrix and $g^T v_i$: is the matrix product. The corresponding associated vector bundle $L(X) \times_\rho \Lambda^k(\ ^n)$ is called the **exterior *k*-bundle** of X and a **differential *k*-form $\boldsymbol{\alpha}$** on X can be identified with a smooth cross-section $\boldsymbol{\alpha} : X \longrightarrow L(X) \times_\rho \Lambda^k(\ ^n)$, $\mathcal{P}_\rho \circ \boldsymbol{\alpha} = id_X$, of $L(X) \times_\rho \Lambda^k(\ ^n)$. Equivalently (Section 6.8, **[N4]**) a **differential *k*-form $\boldsymbol{\alpha}$** on X can be identified with a smooth map $\boldsymbol{\alpha} : L(X) \longrightarrow \Lambda^k(\ ^n)$ that is ρ-equivariant, i.e., satisfies $\boldsymbol{\alpha}(p \cdot g) = (\rho(g^{-1}))(\boldsymbol{\alpha}(p))$ for each $g \in GL(n, \ )$ and each $p \in L(X)$.

Exercise 4.3.1 Show that these three definitions are equivalent in the following sense. Let $\Lambda_1^k(X)$, $\Lambda_2^k(X)$ and $\Lambda_3^k(X)$ denote the C^∞-modules specified by definitions #1, 2 and 3, respectively, and find explicit isomorphisms between $\Lambda_1^k(X)$ and $\Lambda_2^k(X)$ and between $\Lambda_1^k(X)$ and $\Lambda_3^k(X)$.

Hint: The equivalence of the first two definitions is established for 1-forms on pages 265–266 of [**N4**]. The argument hinges on Lemma 5.7.1 of [**N4**] which will be needed here as well. For the equivalence of the first and third definitions proceed as we did for 1-forms in Section 3.3.

Each of these three views of a k-form on a manifold has its uses (unless some particular emphasis is required we will generally omit the adjective "differential"). For example, the purely algebraic operations on forms such as the wedge product extend immediately to the manifold setting via the first definition by simply doing everything pointwise: If $\boldsymbol{\alpha} \in \Lambda^k(X)$ and $\boldsymbol{\beta} \in \Lambda^l(X)$, then $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ is the element of $\Lambda^{k+l}(X)$ defined, at each $x \in X$, by

$$\big((\boldsymbol{\alpha} \wedge \boldsymbol{\beta})(x)\big)\ (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+l}) = (\boldsymbol{\alpha}(x) \wedge \boldsymbol{\beta}(x))\ (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+l})$$

for all $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+l} \in T_x(X)$. Similarly, if $F : X \longrightarrow Y$ is a smooth map and $\boldsymbol{\alpha} \in \Lambda^k(Y)$, then, for each $x \in X$, $F_{*x} : T_x(X) \longrightarrow T_{f(x)}(Y)$ is a linear map and we can define $F^*\boldsymbol{\alpha} \in \Lambda^k(X)$ by pulling $\boldsymbol{\alpha}(F(x))$ back to x by F_{*x}, i.e.,

$$\big((F^*\boldsymbol{\alpha})(x)\big)\ (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \big(\boldsymbol{\alpha}(F(x))\big)\ \big(F_{*x}(\boldsymbol{v}_1), \ldots, F_{*x}(\boldsymbol{v}_k)\big)$$

for all $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in T_x(X)$. All of the purely algebraic properties of these operations (e.g., those to be found in Theorem 4.1.5) are also satisfied by differential forms (or continuous forms, etc.).

Exercise 4.3.2 Show that the wedge product is $C^\infty(X)$-bilinear, i.e., that if $\boldsymbol{\alpha} \in \Lambda^k(X)$, and $\boldsymbol{\beta} \in \Lambda^l(X)$ and $f \in C^\infty(X)$, then

$$(f\boldsymbol{\alpha}) \wedge \boldsymbol{\beta} = \boldsymbol{\alpha} \wedge (f\boldsymbol{\beta}) = f(\boldsymbol{\alpha} \wedge \boldsymbol{\beta})\,.$$

Exercise 4.3.3 Let X and Y be n-dimensional manifolds, (U, φ) a chart on X with coordinate functions $x^1, \ldots, x^n$ and (V, ψ) a chart on Y with coordinate functions $y^1, \ldots, y^n$. Let $F : X \longrightarrow Y$ be a smooth map for which $F(U) \subseteq V$. Show that, for any $k \in C^\infty(U)$,

$$F^*(k\, dy^1 \wedge \cdots \wedge dy^n) = (k \circ F) \det\left(\frac{\partial(y^i \circ F)}{\partial x^j}\right) dx^1 \wedge \cdots \wedge dx^n\,.$$

Hint: $F^*(k\, dy^1 \wedge \cdots \wedge dy^n) = (k \circ F) F^*(dy^1 \wedge \cdots \wedge dy^n)$ and, at each point, $F^*(dy^1 \wedge \cdots \wedge dy^n)$ is a scalar multiple of $dx^1 \wedge \cdots \wedge dx^n$. Now use Corollary 4.1.9.

Exercise 4.3.4 Let X be an n-dimensional manifold, (U, φ) a chart on X with coordinate functions $x^1, \ldots, x^n$ and (V, ψ) a chart on X with coordinate functions $y^1, \ldots, y^n$ and with $U \cap V \neq \emptyset$. Show that if

$$k\, dy^1 \wedge \cdots \wedge dy^n = h\, dx^1 \wedge \cdots \wedge dx^n$$

on $U \cap V$, then

$$h = k \det\left(\frac{\partial y^i}{\partial x^j}\right).$$

Exercise 4.3.4 and Theorem 4.1.11 provide a link between differential forms and the orientability of a smooth manifold.

Theorem 4.3.1 *An n-dimensional manifold X is orientable if and only if X admits a nowhere zero smooth n-form $\boldsymbol{\omega}$.*

Proof: Suppose first that $\boldsymbol{\omega}$ is an n-form on X which is nonzero at every point of X. If (U, φ) is any chart for which U is connected and $x^1, \ldots, x^n$ are the coordinate functions, then $\boldsymbol{\omega}(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n})$ is either everywhere positive or everywhere negative on U. By renumbering, or changing the sign of a coordinate, if necessary, we may assume that $\boldsymbol{\omega}(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}) > 0$ on U. Obtain an atlas of such charts by selecting one at each point of X. Let (U, φ) and (V, ψ) be two charts in this atlas with coordinate functions $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$, respectively, and with $U \cap V \neq \emptyset$. On $U \cap V$ we write $\boldsymbol{\omega} = h\, dx^1 \wedge \cdots \wedge\ dx^n = k\, dy^1 \wedge \cdots \wedge\ dy^n$. Since $\boldsymbol{\omega}(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}) = n!h$ and $\boldsymbol{\omega}(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}) = n!k$ we conclude that $h > 0$ and $k > 0$ on $U \cap V$. From Exercise 4.3.4 it follows that $\det(\frac{\partial y^i}{\partial x^j})$ must be positive on $U \cap V$ so the atlas we have constructed is an oriented atlas and X is orientable.

For the converse we assume X is orientable and let $\{(U_i, \varphi_i)\}$ be an oriented atlas for X. On each U_i define $\boldsymbol{\omega}_i = dx^1 \wedge \cdots \wedge dx^n$, where $x^1, \ldots, x^n$ are the coordinate functions of φ_i. Each $\boldsymbol{\omega}_i$ is a nonzero n-form on U_i. If $U_i \cap U_j \neq \emptyset$, then, on this intersection, $\boldsymbol{\omega}_j = h\, \boldsymbol{\omega}_i$, where, by assumption, $h > 0$. Select for $\{U_i\}$ a family $\{\phi_i\}$ of smooth real-valued functions on X with $\mathrm{supp}(\phi_i) \in U_i$ and define an n-form $\boldsymbol{\omega}$ on all of X by

$$\boldsymbol{\omega} = \sum_i \phi_i\, \boldsymbol{\omega}_i\,.$$

Exercise 4.3.5 Complete the proof by showing that $\boldsymbol{\omega}$ is nowhere zero on X. ∎

Thus, a nonzero n-form on an n-dimensional manifold X determines a unique orientation for X. Just as was the case for vector spaces, however, a given orientation μ for X does not single out a unique element of $\Lambda^n(X)$, but only two equivalence classes (consisting of those elements of $\Lambda^n(X)$ that determine μ and those that determine $-\mu$). With the additional structure of a semi-Riemannian metric $\boldsymbol{g}$, however, a unique nonzero n-form on X is determined by the requirement that it send any oriented, orthonormal basis for any tangent space to 1.

Theorem 4.3.2 *Let X be an n-dimensional manifold with orientation μ and semi-Riemannian metric $\boldsymbol{g}$. Then there exists a unique $\boldsymbol{\omega} \in \Lambda^n(X)$ such that, for any $x \in X$ and any oriented, orthonormal basis $\{e_1, \ldots, e_n\}$ for $T_x(X)$, $\boldsymbol{\omega}(e_1, \ldots, e_n) = 1$.*

Exercise 4.3.6 Prove Theorem 4.3.2. **Hint**: Each point of X is contained in a connected open set U on which is defined a local oriented orthonormal frame field $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ (Exercise 3.3.9). On each such U consider the n-form $\boldsymbol{e}^1 \wedge \cdots \wedge \boldsymbol{e}^n$, where $\{\boldsymbol{e}^1, \ldots, \boldsymbol{e}^n\}$ are the 1-forms dual to $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$. Show that any two such n-forms agree on the intersection of their domains.

The n-form $\boldsymbol{\omega}$ described in Theorem 4.3.2 is called the (**metric**) **volume form** for X determined by μ and $\boldsymbol{g}$. Since we will have need of it somewhat later we intend to explicitly calculate the volume form for the n-sphere $S^n, n \geq 1$. For this we will regard S^n as a submanifold of ${}^{n+1}$ and, as usual, identify each $T_p(S^n)$ with a subspace of $T_p({}^{n+1})$ which, in turn, we identify with ${}^{n+1}$ itself via the canonical isomorphism (Example #3, page 4). ${}^{n+1}$ has its standard orientation and Riemannian metric and we will denote by $x^1, \ldots, x^{n+1}$ the standard coordinate functions on ${}^{n+1}$. Thus, the volume form on ${}^{n+1}$ is just

$$dx^1 \wedge \cdots \wedge dx^{n+1}.$$

Notice that, if $v_0 = (v_0^1, \ldots, v_0^{n+1}), \ldots, v_n = (v_n^1, \ldots, v_n^{n+1})$ are tangent vectors at some point of ${}^{n+1}$, then Corollary 4.1.9 gives

$$(dx^1 \wedge \cdots \wedge dx^{n+1})\ (v_0, v_1, \ldots, v_n) = \det \begin{pmatrix} v_0^1 & \cdots & v_0^{n+1} \\ v_1^1 & \cdots & v_1^{n+1} \\ \vdots & & \vdots \\ v_n^1 & \cdots & v_n^{n+1} \end{pmatrix}.$$

Now, the metric for S^n is just the restriction to S^n of the standard metric on ${}^{n+1}$. The standard orientation for S^n (Section 5.10, [**N4**]) is the one for which the stereographic projection map $\varphi_S : U_S \longrightarrow {}^n$ (Example #4, page 4) is an orientation preserving diffeomorphism (and $\varphi_N : U_N \longrightarrow {}^n$ is orientation reversing). A basis for the tangent space at some point p of S^n is in this orientation if and only if one obtains an oriented basis for ${}^{n+1}$ by adjoining to it (at the beginning) an "outward pointing" normal vector to S^n at p.

Exercise 4.3.7 Let p be a point in S^n and $\{e_1, \ldots, e_n\}$ a basis for $T_p(S^n) \subseteq T_p({}^{n+1}) = {}^{n+1}$. Show that $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis for $T_p(S^n)$ if and only if $\{p, e_1, \ldots, e_n\}$ is an oriented orthonormal basis for $T_p({}^{n+1}) = {}^{n+1}$.

Exercise 4.3.7 makes it clear how to define the volume form $\boldsymbol{\omega}$ on S^n. At each $p \in S^n$ and for any $v_1, \ldots, v_n \in T_p(S^n)$ we set

$$\begin{aligned}\boldsymbol{\omega}_p\left(v_1, \ldots, v_n\right) &= \left(dx^1 \wedge \cdots \wedge dx^{n+1}\right)\left(p, v_1, \ldots, v_n\right) \\ &= \det\begin{pmatrix} p^1 & \ldots & p^{n+1} \\ v_1^1 & \ldots & v_1^{n+1} \\ \vdots & & \vdots \\ v_n^1 & \ldots & v_n^{n+1} \end{pmatrix}. \end{aligned} \tag{4.3.1}$$

This clearly defines a smooth n-form on S^n (indeed, it is the restriction to S^n of a smooth n-form on ${}^{n+1}$ that we will write out shortly). Moreover, if $\{e_1, \ldots, e_n\}$ is an oriented orthonormal basis for $T_p(S^n)$, then $\{p, e_1, \ldots, e_n\}$ is an oriented orthonormal basis for $T_p({}^{n+1})$ so $\boldsymbol{\omega}_p(e_1, \ldots, e_n) = 1$ and $\boldsymbol{\omega}$ must be the volume form for S^n (Theorem 4.3.2). To obtain an explicit representation for $\boldsymbol{\omega}$ we will expand the determinant in (4.3.1) by the minors of the first row.

$$\begin{aligned}\boldsymbol{\omega}_p(v_1, \ldots, v_n) &= p^1 \begin{vmatrix} v_1^2 & v_1^3 & \cdots & v_1^{n+1} \\ \vdots & \vdots & & \vdots \\ v_n^2 & v_n^3 & \cdots & v_n^{n+1} \end{vmatrix} - p^2 \begin{vmatrix} v_1^1 & v_1^3 & \cdots & v_1^{n+1} \\ \vdots & \vdots & & \vdots \\ v_n^1 & v_n^3 & \cdots & v_n^{n+1} \end{vmatrix} + \cdots \\ &\quad + (-1)^{n-1} p^{n+1} \begin{vmatrix} v_1^1 & v_1^2 & \cdots & v_1^n \\ \vdots & \vdots & & \vdots \\ v_n^1 & v_n^2 & \cdots & v_n^n \end{vmatrix} \\ &= p^1 dx^2 \wedge dx^3 \wedge \cdots \wedge dx^{n+1}\left(v_1, \ldots, v_n\right) \\ &\quad - p^2 dx^1 \wedge dx^3 \wedge \cdots \wedge dx^{n+1}\left(v_1, \ldots, v_n\right) + \cdots \\ &\quad + (-1)^{n-1} p^{n+1} dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n \left(v_1, \ldots, v_n\right) \\ &= \sum_{i=1}^{n+1} (-1)^{i-1} p^i \, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1}\left(v_1, \ldots, v_n\right) \\ &= \left(\sum_{i=1}^{n+1} (-1)^{i-1} x^i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1} \right)_p \left(v_1, \ldots, v_n\right). \end{aligned}$$

Thus, $\boldsymbol{\omega}$ is the restriction to S^n of the n-form

$$\tilde{\boldsymbol{\omega}} = \sum_{i=1}^{n+1} (-1)^{i-1} x^i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^{n+1} \tag{4.3.2}$$

on ${}^{n+1}$ ($\boldsymbol{\omega} = \iota^* \tilde{\boldsymbol{\omega}}$, where $\iota : S^n \hookrightarrow {}^{n+1}$). For future reference we will write this out for the first few values of n.

$$\tilde{\boldsymbol{\omega}} = x^1 dx^2 - x^2 dx^1 \quad (n = 1) \tag{4.3.3}$$

$$\tilde{\boldsymbol{\omega}} = x^1 dx^2 \wedge dx^3 - x^2 dx^1 \wedge dx^3 + x^3 dx^1 \wedge dx^2 \quad (n = 2) \tag{4.3.4}$$

$$\begin{aligned} \tilde{\boldsymbol{\omega}} = {} & x^1 dx^2 \wedge dx^3 \wedge dx^4 - x^2 dx^1 \wedge dx^3 \wedge dx^4 \\ & + x^3 dx^1 \wedge dx^2 \wedge dx^4 - x^4 dx^1 \wedge dx^2 \wedge dx^3 \quad (n = 3) \end{aligned} \tag{4.3.5}$$

An orientation and semi-Riemannian metric on an n-manifold X orients and supplies an inner product to each tangent space and, from this, one obtains, for each $k = 0, \ldots, n$, a Hodge star operator

$$^* : \Lambda^k(T_x(X)) \longrightarrow \Lambda^{n-k}(T_x(X))$$

at each point. Thus, one can define the Hodge dual of a k-form on X pointwise:

$$({}^*\boldsymbol{\beta})_x(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}) \;=\; {}^*(\boldsymbol{\beta}_x)(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}).$$

Exercise 4.3.8 Show that, thus defined, the Hodge dual of a smooth k-form is a smooth $(n-k)$-form and that the resulting operator

$$^* : \Lambda^k(X) \longrightarrow \Lambda^{n-k}(X)$$

is $C^\infty(X)$-linear.

Hint: Prove smoothness locally by using local oriented orthonormal frame fields.

Because our definition is pointwise all of the algebraic material on Hodge duals in Section 4.1 carries over at once to the setting of differential forms. In particular, if $\boldsymbol{\omega}$ is the metric volume form on X and $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \Lambda^k(X)$, then

$$\boldsymbol{\alpha} \wedge {}^*\boldsymbol{\beta} = \boldsymbol{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\omega},$$

where $\boldsymbol{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the inner product of the forms $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ defined pointwise by (4.1.4) and (4.1.5) (and therefore a C^∞ function on X).

We conclude this section by observing that, if $\mathcal{V}$ is a finite dimensional real vector space, then a **$\boldsymbol{\mathcal{V}}$-valued differential form $\boldsymbol{\alpha}$** on a manifold X is defined pointwise in the obvious way and, for any choice of a basis $\{T_1, \ldots, T_m\}$ for $\mathcal{V}$, can be written $\boldsymbol{\alpha} = \boldsymbol{\alpha}^1 T_1 + \cdots + \boldsymbol{\alpha}^m T_m$, where each $\boldsymbol{\alpha}^i$ is a real-valued differential form on X. All of the algebraic material in Section 4.2 extends at once to this context by simply doing everything pointwise (the reader who is skeptical and/or scrupulously honest is encouraged to check all of this out).

4.4 The de Rham Complex

The algebraic operations on differential forms introduced in the last section are most conveniently described pointwise. For the exterior differentation operator d to which we now turn, however, it is more efficient to regard forms as operators on vector fields. Recall that if $f \in \Lambda^0(X) = C^\infty(X)$, then the exterior derivative (or differential) of f is the 1-form $df \in \Lambda^1(X)$ given by $df(\boldsymbol{V}) = \boldsymbol{V}f$. In local coordinates, $df = \frac{\partial f}{\partial x^i} dx^i$. Recall also that if $\boldsymbol{\Theta} \in \Lambda^1(X)$ is a 1-form (thought of as a $C^\infty(X)$-module homomorphism from $\mathcal{X}(X)$ to $C^\infty(X)$), then its exterior derivative $d\boldsymbol{\Theta}$ is the map $d\boldsymbol{\Theta} : \mathcal{X}(X) \times \mathcal{X}(X) \longrightarrow C^\infty(X)$ defined by

$$d\boldsymbol{\Theta}(\boldsymbol{V}_1, \boldsymbol{V}_2) = \boldsymbol{V}_1(\boldsymbol{\Theta}\,\boldsymbol{V}_2) - \boldsymbol{V}_2(\boldsymbol{\Theta}\,\boldsymbol{V}_1) - \boldsymbol{\Theta}([\boldsymbol{V}_1, \boldsymbol{V}_2]).$$

$d\boldsymbol{\Theta}$ is $C^\infty(X)$-bilinear and skew-symmetric (page 320, [**N4**]) and so is an element of $\Lambda^2(X)$. In local coordinates

$$d\boldsymbol{\Theta} = d(\Theta_j\ dx^j) = d\Theta_j \wedge dx^j = \frac{\partial \Theta_j}{\partial x^i} dx^i \wedge dx^j\,.$$

Before proceeding with the general definition we ask the reader to take this one step further.

Exercise 4.4.1 Let $\boldsymbol{\Omega} \in \Lambda^2(X)$ be a 2-form on X (thought of as a bilinear operator on vector fields) and define $d\boldsymbol{\Omega} : \mathcal{X}(X) \times \mathcal{X}(X) \times \mathcal{X}(X) \longrightarrow C^\infty(X)$ by

$$\begin{aligned} d\boldsymbol{\Omega}(\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3) = \ & \boldsymbol{V}_1(\boldsymbol{\Omega}(\boldsymbol{V}_2, \boldsymbol{V}_3)) - \boldsymbol{V}_2(\boldsymbol{\Omega}(\boldsymbol{V}_1, \boldsymbol{V}_3)) \\ & + \boldsymbol{V}_3(\boldsymbol{\Omega}(\boldsymbol{V}_1, V_2)) - \boldsymbol{\Omega}([\boldsymbol{V}_1, \boldsymbol{V}_2], \boldsymbol{V}_3) \\ & + \boldsymbol{\Omega}([\boldsymbol{V}_1, \boldsymbol{V}_3], \boldsymbol{V}_2) - \boldsymbol{\Omega}([\boldsymbol{V}_2, \boldsymbol{V}_3], \boldsymbol{V}_1). \end{aligned}$$

Show that $d\boldsymbol{\Omega} \in \Lambda^3(X)$ and that, in local coordinates,

$$d\boldsymbol{\Omega} = d\left(\frac{1}{2}\,\Omega_{jk}\, dx^j \wedge dx^k\right) = \frac{1}{2}\, d\Omega_{jk} \wedge dx^j \wedge dx^k = \frac{1}{2}\frac{\partial \Omega_{jk}}{\partial x^i}\, dx^i \wedge dx^j \wedge dx^k.$$

Hint: To prove that $d\boldsymbol{\Omega} \in \Lambda^3(X)$ follow the proof for $d\boldsymbol{\Theta}$, $\boldsymbol{\Theta} \in \Lambda^1(X)$, on page 320, [**N4**]. You will need the identities $\boldsymbol{V}(fg) = \boldsymbol{V}(f)g + f\,\boldsymbol{V}(g)$ and $[f\,\boldsymbol{V}, g\,\boldsymbol{W}] = fg[\boldsymbol{V}, \boldsymbol{W}] + f(\boldsymbol{V}g)\,\boldsymbol{W} - g(\boldsymbol{W}f)\,\boldsymbol{V}$, which are proved in Section 5.7 of [**N4**]. For the coordinate expression, keep in mind that Lie brackets of coordinate vector fields must vanish.

We record a few properties of the exterior differentiation operator d thus defined on k-forms for $k = 0, 1, 2$. If $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are two such k-forms and $a_1, a_2 \in$, it is clear that

$$d(a_1\boldsymbol{\omega}_1 + a_2\boldsymbol{\omega}_2) = a_1 d\boldsymbol{\omega}_1 + a_2 d\boldsymbol{\omega}_2\,. \tag{4.4.1}$$

We know already that, if $f \in \Lambda^0(X)$ and $\boldsymbol{\omega} \in \Lambda^1(X)$, then $d(f\boldsymbol{\omega}) = f d\boldsymbol{\omega} + df \wedge \boldsymbol{\omega}$ and, since wedge product by 0-forms coincides with ordinary multiplication, this may be written

$$d(f \wedge \boldsymbol{\omega}) = df \wedge \boldsymbol{\omega} + f \wedge d\boldsymbol{\omega}\,.$$

From this and the anti-commutativity of wedge products for 1-forms we obtain

$$d(\boldsymbol{\omega} \wedge f) = d\boldsymbol{\omega} \wedge f - \boldsymbol{\omega} \wedge df\,.$$

Exercise 4.4.2 Let $\boldsymbol{\omega}_1 \in \Lambda^k(X)$ and $\boldsymbol{\omega}_2 \in \Lambda^l(X)$ with $k + l \leq 2$. Show that

$$d(\boldsymbol{\omega}_1 \wedge \boldsymbol{\omega}_2) = d\boldsymbol{\omega}_1 \wedge \boldsymbol{\omega}_2 + (-1)^k \boldsymbol{\omega}_1 \wedge d\boldsymbol{\omega}_2. \tag{4.4.2}$$

We know also that $d(df) = 0$ for any $f \in \Lambda^0(X)$.

Exercise 4.4.3 Show that, for any $\boldsymbol{\omega} \in \Lambda^1(X)$,

$$d(d\boldsymbol{\omega}) = 0\,. \tag{4.4.3}$$

We show now that there is exactly one way to generalize all of this to higher degree forms and retain properties (4.4.1), (4.4.2) and (4.4.3).

Theorem 4.4.1 *Let X be a smooth n-dimensional manifold. Then there exists a unique family of operators*

$$d^k : \Lambda^k(X) \longrightarrow \Lambda^{k+1}(X), \quad k = 0, \ldots, n$$

(all written simply d unless particular emphasis is required) which satisfy

1. $d(a_1\,\boldsymbol{\omega}_1 + a_2\,\boldsymbol{\omega}_2) = a_1\,d\boldsymbol{\omega}_1 + a_2\,d\boldsymbol{\omega}_2$ *for all* $a_1, a_2 \in$ *and* $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \in \Lambda^k(X)$.
2. $d(\boldsymbol{\omega}_1 \wedge \boldsymbol{\omega}_2) = d\boldsymbol{\omega}_1 \wedge \boldsymbol{\omega}_2 + (-1)^k \boldsymbol{\omega}_1 \wedge d\boldsymbol{\omega}_2$ *for all* $\boldsymbol{\omega}_1 \in \Lambda^k(X)$ *and* $\boldsymbol{\omega}_2 \in \Lambda^l(X)$.
3. $d(d\boldsymbol{\omega}) = 0$ *for all* $\boldsymbol{\omega} \in \Lambda^k(X)$.
4. df *is given by* $df(\boldsymbol{V}) = \boldsymbol{V}f$ *for all* $f \in \Lambda^0(X)$.

Remark: Each d^k is called an **exterior differentiation** operator and $d\boldsymbol{\omega}$ is the **exterior derivative** of $\boldsymbol{\omega}$.

Proof: First we show that if a family of operators satisfying (1)–(4) exists, then it is unique. This we do by showing that if $\boldsymbol{\omega} \in \Lambda^k(X)$, then (1)–(4) imply that $d\boldsymbol{\omega}$ must be given in any local coordinate system by

$$\begin{aligned} d\boldsymbol{\omega} &= d\left(\frac{1}{k!}\omega_{i_1 \cdots i_k}\; dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) \\ &= \frac{1}{k!}\left(d\omega_{i_1 \cdots i_k}\right) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \end{aligned} \tag{4.4.4}$$

This follows from

$$\begin{aligned} d\boldsymbol{\omega} &= d\left(\frac{1}{k!}\omega_{i_1\cdots i_k}\ dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right) \\ &= \frac{1}{k!}d\Big(\omega_{i_1\cdots i_k}\ dx^{i_1}\wedge\cdots\wedge dx^{i_k}\Big) && \text{(by (1))} \\ &= \frac{1}{k!}d\Big(\omega_{i_1\cdots i_k}\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\Big) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{k!}\Big[\Big(d\omega_{i_1\cdots i_k}\Big)\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k} \\ &\qquad + (-1)^0\omega_{i_1\cdots i_k}\wedge d(dx^{i_1}\wedge\cdots\wedge dx^{i_k})\Big] && \text{(by (1) and (2))} \\ &= \frac{1}{k!}\Big(d\omega_{i_1\cdots i_k}\Big)\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k} \end{aligned}$$

since each $d(dx^{i_1}\wedge\cdots\wedge dx^{i_k}) = 0$ by (2), (3) and induction.

Next we write out an explicit, coordinate independent formula for an operator on $\Lambda^k(X)$ which yields elements of $\Lambda^{k+1}(X)$ and satisfies (1)–(4) (and so must, therefore, be given in local coordinates by (4.4.4)). Specifically, we claim that, for each $\boldsymbol{\omega}\in\Lambda^k(X)$, $d\boldsymbol{\omega}$ must be given by

$$\begin{aligned} d\boldsymbol{\omega}(\boldsymbol{V}_1,\ldots,\boldsymbol{V}_{k+1}) &= \sum_{i=1}^{k+1}(-1)^{i+1}\,\boldsymbol{V}_i\left(\boldsymbol{\omega}\left(\boldsymbol{V}_1,\ldots,\widehat{\boldsymbol{V}}_i,\ldots,\boldsymbol{V}_{k+1}\right)\right) \\ &+ \sum_{1\le i<j\le n}(-1)^{i+j}\boldsymbol{\omega}\left([\boldsymbol{V}_i,\,\boldsymbol{V}_j],\,\boldsymbol{V}_1,\ldots,\widehat{\boldsymbol{V}}_i,\ldots,\widehat{\boldsymbol{V}}_j,\ldots,\boldsymbol{V}_{k+1}\right). \end{aligned} \tag{4.4.5}$$

Exercise 4.4.4 Show that $d\boldsymbol{\omega}\in\Lambda^{k+1}(X)$. **Hint**: See Exercise 4.4.1 and its Hint.

Exercise 4.4.5 Show that when $\boldsymbol{\omega}$ is $f\in\Lambda^0(X)$, $\boldsymbol{\Theta}\in\Lambda^1(X)$, or $\boldsymbol{\Omega}\in\Lambda^2(X)$, (4.4.5) reduces to our earlier definitions of $df, d\boldsymbol{\Theta}$ and $d\boldsymbol{\Omega}$.

Exercise 4.4.6 Show directly from (4.4.5) that, in local coordinates, $d\boldsymbol{\omega}$ is given by (4.4.4). **Hint**: See Exercise 4.4.1 and its Hint.

All that remains is to prove that, when $d\boldsymbol{\omega}$ is defined by (4.4.5), conditions (1)–(4) are all satisfied. Now, (1) is obvious from (4.4.5) and (4) is the first part of Exercise 4.4.5. On the other hand, (2) and (3) can be proved at each fixed point of X and therefore locally, in coordinates. Furthermore, linearity (1) of d and bilinearity of the wedge product imply that we may restrict our attention to k-forms such as $\boldsymbol{\omega} = f dx^{i_1}\wedge\cdots\wedge dx^{i_k}$, where $f\in C^\infty(X)$. Appealing to Exercise 4.4.6 we compute

$$
\begin{aligned}
d\boldsymbol{\omega} &= d(f\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}) = df\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\\
&= \left(\sum_{i=1}^{n}\frac{\partial f}{\partial x^i}dx^i\right)\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k} = \sum_{i=1}^{n}\frac{\partial f}{\partial x^i}dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}
\end{aligned}
$$

and

$$
\begin{aligned}
d(d\boldsymbol{\omega}) &= d\left(\sum_{i=1}^{n}\frac{\partial f}{\partial x^i}\;dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\\
&= \sum_{i=1}^{n} d\left(\frac{\partial f}{\partial x^i}\;dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\\
&= \sum_{i=1}^{n} d\left(\frac{\partial f}{\partial x^i}\right)\wedge dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\\
&= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x^j\partial x^i}dx^j\right)\wedge dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x^j\partial x^i}dx^j\wedge dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}.
\end{aligned}
$$

Now, every term in this sum with $i=j$ is zero since $dx^i\wedge dx^i=0$ and, when $i\neq j$, the terms

$$\frac{\partial^2 f}{\partial x^j\partial x^i}dx^j\wedge dx^i\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}$$

and

$$\frac{\partial^2 f}{\partial x^i\partial x^j}\;dx^i\wedge dx^j\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}$$

cancel in pairs. Thus, $d(d\boldsymbol{\omega})=0$. Finally, we let $\boldsymbol{\omega}_1 = f\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}$ and $\boldsymbol{\omega}_2 = g\,dx^{j_1}\wedge\cdots\wedge dx^{j_l}$ and compute

$$
\begin{aligned}
d\big(\boldsymbol{\omega}_1\wedge\boldsymbol{\omega}_2\big) &= d((fg)dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\big)\\
&= d(fg)\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\\
&= (g\,df + f\,dg)\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\\
&= g\,df\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\\
&\qquad + f\,dg\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\\
&= \big(df\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\big)\wedge\big(g\,dx^{j_1}\wedge\cdots\wedge dx^{j_l}\big)\\
&\qquad + (-1)^k\big(f\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}\big)\wedge\big(dg\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\big)\\
&= d\omega_1\wedge\omega_2 + (-1)^k\,\omega_1\wedge d\omega_2.
\end{aligned}
$$

■

Exercise 4.4.7 Show that d commutes with pullback. More precisely, let $F : X \longrightarrow Y$ be a smooth map and $\boldsymbol{\omega} \in \Lambda^k(Y)$. Show that

$$d(F^*\boldsymbol{\omega}) = F^*(d\boldsymbol{\omega}). \tag{4.4.6}$$

Hint: This is already known when $k = 0$ (page 10) and $k = 1$ (page 12). Assume the result for $(k-1)$-forms and prove (4.4.6) for $\boldsymbol{\omega} = f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}$.

Exercise 4.4.8 Let $X = \ ^3$ with its standard Riemannian metric and orientation and let x, y and z be the standard coordinate functions on $\ ^3$ (so that $\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\}$ is a global oriented orthonormal frame field and $\{dx, dy, dz\}$ is the dual oriented orthonormal coframe field). Let * denote the corresponding Hodge star operator. Finally, consider the natural isomorphism between $\Lambda^1(\ ^3)$ and $\mathcal{X}(\ ^3)$ under which a 1-form $\boldsymbol{\alpha} = f\,dx + g\,dy + h\,dz$ corresponds to the vector field $\boldsymbol{V} = f\frac{\partial}{\partial x} + g\frac{\partial}{\partial y} + h\frac{\partial}{\partial z}$.

(a) Show that, for any $f \in \Lambda^0(\ ^3)$, $df \in \Lambda^1(\ ^3)$ corresponds to $\operatorname{grad} f \in \mathcal{X}(\ ^3)$.

(b) Show that if $\boldsymbol{\alpha} \in \Lambda^1(\ ^3)$ corresponds to $\boldsymbol{V} \in \mathcal{X}(\ ^3)$, then $^*d\boldsymbol{\alpha} \in \Lambda^1(\ ^3)$ corresponds to $\operatorname{curl} \boldsymbol{V} \in \mathcal{X}(\ ^3)$.

(c) Show that if $\boldsymbol{\alpha} \in \Lambda^1(\ ^3)$ corresponds to $\boldsymbol{V} \in \mathcal{X}(\ ^3)$, then $^*d^*\boldsymbol{\alpha} \in \Lambda^0(\ ^3)$ is $\operatorname{div} \boldsymbol{V}$.

(d) Show that, for any $f \in \Lambda^0(\ ^3)$, $^*d\,^*df$ is the Laplacian of f (i.e., $^*d\,^*df = \nabla^2 f = \operatorname{div}\operatorname{grad} f$).

(e) Show that if $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \Lambda^1(\ ^3)$ correspond to $\boldsymbol{V}, \boldsymbol{W} \in \mathcal{X}(\ ^3)$, then $^*(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \in \Lambda^1(\ ^3)$ corresponds to $\boldsymbol{V} \times \boldsymbol{W} \in \mathcal{X}(\ ^3)$.

(f) Use (3) of Theorem 4.4.1 to prove that $\operatorname{curl}(\operatorname{grad} f) = \boldsymbol{0}$ and $\operatorname{div}(\operatorname{curl} \boldsymbol{V}) = 0$.

(g) Use (2) of Theorem 4.4.1 to prove that $\operatorname{curl}(f\boldsymbol{V}) = \operatorname{grad} f \times \boldsymbol{V} + f \operatorname{curl} \boldsymbol{V}$.

Remark: The upshot of Exercise 4.4.8 is that the calculus of forms on $\ ^3$ is just a disguised version of classical vector analysis. One reason for preferring forms (aside from the elegant integration theory we construct in Section 4.6) is that they make sense on higher dimensional manifolds whereas much of vector calculus (e.g., the cross product and curl) do not. Thus, for example, once they are written in terms of forms, Maxwell's equations are defined on any spacetime.

The exterior differentiation operators are linear transformations on vector spaces of smooth forms and when all of these are collected into the sequence

$$\Lambda^0(X) \xrightarrow{d^0} \Lambda^1(X) \xrightarrow{d^1} \cdots \xrightarrow{d^{n-2}} \Lambda^{n-1}(X) \xrightarrow{d^{n-1}} \Lambda^n(X) \xrightarrow{d^n} 0$$

one obtains what is called the **de Rham complex** of the n-dimensional manifold X. Theorem 4.4.1 (3) asserts that the composition of any two consecutive maps in this sequence is identically zero, i.e., the image of any d^{k-1} is contained in the kernel of d^k. A differential form $\boldsymbol{\omega}$ on X is said to be **closed** if $d\boldsymbol{\omega} = 0$ and **exact** if $\boldsymbol{\omega} = d\boldsymbol{\eta}$ for some form $\boldsymbol{\eta}$ of degree one less. Thus, one may rephrase Theorem 4.4.1 (3) by saying that *any exact form is closed.* The converse is generally not true and we will construct an explicit example shortly. Indeed, we will devote what remains of this section and all of Chapter 5 to the issue of when closed implies exact and, when it does not, the extent to which it does not. Although not apparent at the moment, these are questions about the topology of X.

We begin our discussion in $\mathbb{R}^2$. If $\boldsymbol{\omega}$ is a 1-form on $\mathbb{R}^2$ and standard coordinates are x^1 and x^2, then we can write $\boldsymbol{\omega} = \omega_1\, dx^1 + \omega_2\, dx^2$ and compute $d\boldsymbol{\omega} = \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2}\right) dx^1 \wedge dx^2$. Thus, if $\boldsymbol{\omega}$ is closed we must have

$$\frac{\partial \omega_2}{\partial x^1} = \frac{\partial \omega_1}{\partial x^2}.$$

Now define $\eta \in \Lambda^0(\mathbb{R}^2)$ by

$$\eta(x^1, x^2) = x^1 \int_0^1 \omega_1(tx^1, tx^2)dt + x^2 \int_0^1 \omega_2(tx^1, tx^2)dt\,.$$

Observe that

$$\begin{aligned}
\frac{\partial \eta}{\partial x^1} &= x^1 \int_0^1 \frac{\partial \omega_1}{\partial x^1}\,(tx^1, tx^2)t\,dt + \int_0^1 \omega_1(tx^1, tx^2)dt \\
&\qquad + x^2 \int_0^1 \frac{\partial \omega_2}{\partial x^1}(tx^1, tx^2)t\,dt \\
&= \int_0^1 \left[\omega_1(tx^1, tx^2) + t\left(x^1 \frac{\partial \omega_1}{\partial x^1}\,(tx^1, tx^2)\right)\right. \\
&\qquad \left. + t\left(x^2 \frac{\partial \omega_1}{\partial x^2}\,(tx^1, tx^2)\right)\right] dt \\
&= \int_0^1 \frac{d}{dt}\left[t\,\omega_1(tx^1, tx^2)\right] dt = t\omega_1(tx^1, tx^2)\Big|_0^1 \\
&= \omega_1(x^1, x^2).
\end{aligned}$$

Similarly,

$$\frac{\partial \eta}{\partial x^2} = \omega_2(x^1, x^2)$$

so

$$\boldsymbol{\omega} = d\eta\,.$$

Thus, on $\mathbb{R}^2$, every closed 1-form is, indeed, exact. We will generalize this simple result quite substantially when we prove the Poincaré Lemma, but for the moment we simply wish to contrast it with the situation on $\mathbb{R}^2 - \{(0,0)\}$. Here we will write down an explicit 1-form that is closed, but not exact.

We construct our 1-form from two polar angular coordinate functions θ_1 and θ_2. First let $L_1 = \{\, p \in {}^2 : x^1(p) \geq 0,\ x^2(p) = 0 \,\}$ and $U_1 = {}^2 - L_1$. Define $\varphi_1 : U_1 \longrightarrow {}^2$ by $\varphi_1(x^1, x^2) = (r(x^1, x^2), \theta_1(x^1, x^2))$, where $r(x^1, x^2) = \sqrt{(x^1)^2 + (x^2)^2}$ and $\theta_1(x^1, x^2))$ is the unique angle in $(0, 2\pi)$ such that $\tan\theta_1(x^1, x^2)) = x^2/x^1$. More precisely,

$$\theta_1(x^1, x^2) = \begin{cases} \arctan(x^2/x^1) & , \quad x^1 > 0,\ x^2 > 0 \\ \dfrac{\pi}{2} & , \quad x^1 = 0,\ x^2 > 0 \\ \pi + \arctan(x^2/x^1) & , \quad x^1 < 0 \\ \dfrac{3\pi}{2} & , \quad x^1 = 0,\ x^2 < 0 \\ 2\,\pi + \arctan(x^2/x^1), & \quad x^1 > 0,\ x^2 < 0 \end{cases}$$

Then $\varphi_1(U_1) = (0, \infty) \times (0, 2\pi)$ and $\varphi_1^{-1} : \varphi_1(U_1) \longrightarrow U_1$ is given by $\varphi_1^{-1}(r, \theta_1) = (r\cos\theta_1, r\sin\theta_1)$. Thus, $\varphi_1 : U_1 \longrightarrow \varphi_1(U_1)$ is a diffeomorphism and, in particular, θ_1 is smooth on U_1, i.e., θ_1 is a 0-form on U_1. A simple calculation shows that

$$d\theta_1 = \frac{-x^2}{(x^1)^2 + (x^2)^2} dx^1 + \frac{x^1}{(x^1)^2 + (x^2)^2} dx^2 \quad \text{on } U_1.$$

Now let $L_2 = \{\, p \in {}^2 : x^1(p) \leq 0,\ x^2(p) = 0 \,\}$ and $U_2 = {}^2 - L_2$. Just as above we define a polar coordinate chart $\varphi_2 : U_2 \longrightarrow (0, \infty) \times (-\pi, \pi)$, $\varphi_2(x^1, x^2) = (r(x^1, x^2), \theta_2(x^1, x^2))$, where $\theta_2(x^1, x^2))$ is now the unique angle in $(-\pi, \pi)$ with $\tan\theta_2(x^1, x^2)) = x^2/x^1$. Thus, θ_2 is a 0-form on U_2 and the same calculation as for θ_1 shows that

$$d\theta_2 = \frac{-x^2}{(x^1)^2 + (x^2)^2}\, dx^1 + \frac{x^1}{(x^1)^2 + (x^2)^2}\, dx^2 \quad \text{on } U_2.$$

In particular, $d\theta_1$ and $d\theta_2$ agree on the intersection $U_1 \cap U_2$ and so together define a 1-form $\boldsymbol{\omega}$ on $U_1 \cup U_2 = {}^2 - \{(0,0)\}$. Moreover, $\boldsymbol{\omega}$ is obviously closed on ${}^2 - \{(0,0)\}$ since, on U_1, $d\boldsymbol{\omega} = d(d\theta_1) = 0$ and, on U_2, $d\boldsymbol{\omega} = d(d\theta_2) = 0$. We claim, however, that $\boldsymbol{\omega}$ is not exact on ${}^2 - \{(0,0)\}$.

Remark: Once we have learned how to integrate 1-forms we will be able to give the traditional proof of this that one finds in calculus books, i.e., we will show that the integral of $\boldsymbol{\omega}$ over the unit circle is nonzero.

Suppose to the contrary that there exists an $f \in C^\infty({}^2 - \{(0,0)\})$ such that $df = \boldsymbol{\omega}$ on ${}^2 - \{(0,0)\}$. Then, in particular, on ${}^2 - L_1$, $df = d\theta_1$ so $d(f - \theta_1) = 0$. Thus $\frac{\partial f}{\partial x^1} = \frac{\partial \theta_1}{\partial x^1}$ and $\frac{\partial f}{\partial x^2} = \frac{\partial \theta_1}{\partial x^2}$ on ${}^2 - L_1$. Since ${}^2 - L_1$ is connected it follows that there is a constant c_1 such that $f = \theta_1 + c_1$ on

$^2 - L_1$. Similarly, there is a constant c_2 such that $f = \theta_2 + c_2$ on $^2 - L_2$. But then $\theta_2 = \theta_1 + (c_1 - c_2)$ on $^2 - (L_1 \cup L_2)$ and this is not the case since $\theta_2 = \theta_1 + 0$ when $x^1 > 0$ and $\theta_2 = \theta_1 - 2\pi$ when $x^2 < 0$.

Remark: This last argument actually shows that $\boldsymbol{\omega}$ fails to be exact on every neighborhood of $(0,0)$. On the other hand, $\boldsymbol{\omega}$ clearly is exact on a neighborhood of any point other than the origin.

A subset U of ^{n} is said to **star-shaped** with respect to $p \in U$ if it contains the entire line segment from p to anything else in U, i.e., if, for each $x \in U$, $tx + (1-t)p \in U$ for each t in $[0,1]$. In particular, U is star-shaped with respect to the origin if $tx \in U$ whenever $x \in U$ and $0 \leq t \leq 1$.

Theorem 4.4.2 (The Poincaré Lemma): *Let U be an open, star-shaped subset of ^{n}. Then every closed form on U is exact.*

Proof: We claim that we may assume U is star-shaped with respect to the origin. Indeed, this is just a very special case of the following result.

Exercise 4.4.9 Let X be a smooth n-dimensional manifold and suppose that, for some integer k with $1 \leq k \leq n$, every closed k-form on X is exact. Let Y be a smooth manifold diffeomorphic to X. Show that every closed k-form on Y is exact. **Hint**: d commutes with pullback (Exercise 4.4.7).

Thus, we will assume that $tx \in U$ whenever $x \in U$ and $0 \leq t \leq 1$.

The idea behind the proof of Theorem 4.4.2 invariably seems rather strange the first time one encounters it, but it has turned out to be such a good idea that it has acquired a name and an honored place in algebraic topology (and will be familiar to those who have read Chapter 3 of [**N4**]). We will construct a family of linear maps

$$h^k : \Lambda^k(U) \longrightarrow \Lambda^{k-1}(U), \quad k \geq 1,$$

such that

$$d^{k-1} \circ h^k + h^{k+1} \circ d^k = id_{\Lambda^k(U)} \tag{4.4.7}$$

(such a family of maps is called an *algebraic homotopy* or *cochain homotopy* and we will investigate them more formally in Section 5.4). Observe that if we manage to build such maps the proof of the Poincaré Lemma will be trivial since, if $d^k\boldsymbol{\omega} = 0$, then $h^{k+1}(d^k\boldsymbol{\omega}) = 0$ so (4.4.7) gives $d^{k-1}(h^k\boldsymbol{\omega}) = \boldsymbol{\omega}$ so $h^k\boldsymbol{\omega}$ is a $(k-1)$-form $\boldsymbol{\eta}$ with $d\boldsymbol{\eta} = \boldsymbol{\omega}$.

Our definition of the maps h^k will be in terms of standard coordinates $x^1, \ldots, x^n$ on ^{n}. We define h^k on elements of $\Lambda^k(U)$ of the form

$$\boldsymbol{\omega} = f\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

(no assumption about the ordering of the indices) and extend by linearity. Since U is star-shaped with respect to the origin we may define $h^k\boldsymbol{\omega}$ for such an $\boldsymbol{\omega}$ by

$$(h^k\boldsymbol{\omega})(x) = \sum_{a=1}^{k} (-1)^{a-1} \left(\int_0^1 t^{k-1} f(tx) dt \right) x^{i_a} dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k}. \tag{4.4.8}$$

Exercise 4.4.10 Show that $h^k\boldsymbol{\omega}$ is well-defined, i.e., that the right-hand side is the same if $\boldsymbol{\omega}$ is written $(-1)^\sigma f dx^{\sigma(i_1)} \wedge \cdots \wedge dx^{\sigma(i_k)}$ for some permutation σ of $\{1, \ldots, k\}$.

Now we compute $d^{k-1}(h^k\boldsymbol{\omega})$ and $h^{k+1}(d^k\boldsymbol{\omega})$.

$$\begin{aligned} d^{k-1}(h^k\boldsymbol{\omega}) &= \sum_{a=1}^{k} (-1)^{a-1} d \left\{ \left(\int_0^1 t^{k-1} f(tx) dt \right) x^{i_a} dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k} \right\} \\ &= \sum_{a=1}^{k} (-1)^{a-1} d \left(\left(\int_0^1 t^{k-1} f(tx) dt \right) x^{i_a} \right) \wedge dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k} \end{aligned}$$

Exercise 4.4.11 Show that

$$d \left(\left(\int_0^1 t^{k-1} f(tx) dt \right) x^{i_a} \right) = \left(\int_0^1 t^{k-1} f(tx) dt \right) dx^{i_a} + \sum_{j=1}^{n} \left(\int_0^1 t^k \frac{\partial f}{\partial x^j}(tx) dt \right) x^{i_a} dx^j.$$

Thus,

$$\begin{aligned} d^{k-1}(h^k\boldsymbol{\omega}) &= \sum_{a=1}^{k} (-1)^{a-1} \left(\int_0^1 t^{k-1} f(tx) dt \right) dx^{i_a} \wedge dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k} \\ &\quad + \sum_{a=1}^{k} \sum_{j=1}^{n} (-1)^{a-1} \left(\int_0^1 t^k \frac{\partial f}{\partial x^j}(tx) dt \right) x^{i_a} dx^j \wedge dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k} \\ &= k \left(\int_0^1 t^{k-1} f(tx) dt \right) dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &\quad + \sum_{a=1}^{k} \sum_{j=1}^{n} (-1)^{a-1} \left(\int_0^1 t^k \frac{\partial f}{\partial x^j}(tx) dt \right) x^{i_a} dx^j \wedge dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k}. \end{aligned}$$

Now,

$$d^k\boldsymbol{\omega} = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

so, by linearity of h^{k+1},

$$\begin{aligned} h^{k+1}(d^k\boldsymbol{\omega}) = &\sum_{j=1}^{n} \left(\int_0^1 t^k \frac{\partial f}{\partial x^j}(tx)dt \right) x^j dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &+ \sum_{j=1}^{n}\sum_{a=1}^{k} (-1)^a \left(\int_0^1 t^k \frac{\partial f}{\partial x^j}(tx)dt \right) x^{i_a} dx^j \wedge dx^{i_1} \wedge \cdots \\ &\quad \wedge \widehat{dx^{i_a}} \wedge \cdots \wedge dx^{i_k}. \end{aligned}$$

Thus, computing $d^{k-1}(h^k\boldsymbol{\omega}) + h^{k+1}(d^k\boldsymbol{\omega})$, the two double sums cancel and we obtain

$$\begin{aligned} d^{k-1}(h^k\boldsymbol{\omega}) + h^{k+1}(d^k\boldsymbol{\omega}) &= k\left(\int_0^1 t^{k-1} f\,(tx)dt \right) dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &\quad + \sum_{j=1}^{n} \left(\int_0^1 t^k \frac{\partial f}{\partial x^j}(tx)dt \right) x^j dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &= \left\{ \int_0^1 \left[kt^{k-1} f\,(tx) \right.\right. \\ &\quad \left.\left. + \sum_{j=1}^{n} t^k \frac{\partial f}{\partial x^j}(tx)x^j \right] dt \right\} dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &= \left\{ \int_0^1 \frac{d}{dt}[t^k f(tx)]dt \right\} dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\ &= f(x)dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \boldsymbol{\omega} \end{aligned}$$

■

Exercise 4.4.12 Let X be a smooth n-dimensional manifold and $\boldsymbol{\omega} \in \Lambda^k(X)$, $1 \leq k \leq n$, with $d\boldsymbol{\omega} = 0$. Show that $\boldsymbol{\omega}$ is locally exact, i.e., that for each $x \in X$ there is an open neighborhood U of x in X and a $(k-1)$-form $\boldsymbol{\eta}$ on U such that $d\boldsymbol{\eta} = \boldsymbol{\omega}$ on U. **Hint**: Exercise 4.4.9.

Exercise 4.4.13 Let $\mathcal{V}$ be a finite dimensional real vector space and $\boldsymbol{\omega}$ a $\mathcal{V}$-valued k-form on the manifold X. Let $\{\,T_1, \ldots, T_m\,\}$ be a basis for $\mathcal{V}$ and write $\boldsymbol{\omega} = \boldsymbol{\omega}^1 T_1 + \cdots + \boldsymbol{\omega}^m T_m$, where each $\boldsymbol{\omega}^i$ is in $\Lambda^k(X)$. Define $d\boldsymbol{\omega}$ by

$$d\boldsymbol{\omega} = d\boldsymbol{\omega}^1 T_1 + \cdots + d\boldsymbol{\omega}^m T_m$$

and show that this definition does not depend on the choice of $\{T_1, \ldots, T_m\}$.

Exercise 4.4.14 Let $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ be real vector spaces and $\rho : \mathcal{U} \times \mathcal{V} \longrightarrow \mathcal{W}$ a bilinear map. Let $\boldsymbol{\omega}_1$ be a $\mathcal{U}$-valued k-form and $\boldsymbol{\omega}_2$ a $\mathcal{V}$-valued l-form on X. Then $\boldsymbol{\omega}_1 \wedge_\rho \boldsymbol{\omega}_2$ a $\mathcal{W}$-valued $(k+l)$-form on X. Show that

$$d(\boldsymbol{\omega}_1 \wedge_\rho \boldsymbol{\omega}_2) = d\boldsymbol{\omega}_1 \wedge_\rho \boldsymbol{\omega}_2 + (-1)^k \boldsymbol{\omega}_1 \wedge_\rho d\boldsymbol{\omega}_2 .$$

We denote the set of all smooth $\mathcal{V}$-valued k-forms on X by $\Lambda^k(X, \mathcal{V})$ and provide it with the obvious pointwise structure of a $C^\infty(X)$-module.

4.5 Tensorial Forms

We consider now a smooth principal G-bundle $G \hookrightarrow P \stackrel{\mathcal{P}}{\longrightarrow} X$, where G is a matrix Lie group, and with right action $\sigma : P \times G \longrightarrow P$ written $\sigma(p, g) = p \cdot g$. $\mathcal{V}$ will denote a finite dimensional real vector space and $\rho : G \longrightarrow GL(\mathcal{V})$ a representation of G on $\mathcal{V}$. We will write $(\rho(g))(v) = g \cdot v$ for each $g \in G$ and $v \in \mathcal{V}$. Recall that a $\mathcal{V}$-valued 0-form $\phi : P \longrightarrow \mathcal{V}$ on P is said to be equivariant if $\phi(p \cdot g) = g^{-1} \cdot \phi(p)$ and notice that this can be written $\sigma_g^* \phi = g^{-1} \cdot \phi$, where $\sigma_g : P \longrightarrow P$ is the diffeomorphism $\sigma_g(p) = p \cdot g$. Recall also that the definition of a connection form $\boldsymbol{\omega}$ on the bundle includes the analogous condition $\sigma_g^* \boldsymbol{\omega} = g^{-1} \cdot \boldsymbol{\omega}$, where $\mathcal{V}$ is now the Lie algebra $\mathcal{G}$ of G and ρ is the adjoint representation. In this case, the curvature $\boldsymbol{\Omega}$ of the connection $\boldsymbol{\omega}$ also satisfies $\sigma_g^* \boldsymbol{\Omega} = g^{-1} \cdot \boldsymbol{\Omega}$ (Lemma 6.2.2, [**N4**]). In general, a $\mathcal{V}$-valued k-form $\boldsymbol{\varphi}$ on P is said to be **pseudotensorial** of **type** ρ if it satisfies $\sigma_g^* \boldsymbol{\varphi} = g^{-1} \cdot \boldsymbol{\varphi}$ for each $g \in G$, i.e., if

$$\boldsymbol{\varphi}_{p \cdot g}\big((\sigma_g)_{*p}(\boldsymbol{v}_1), \ldots, (\sigma_g)_{*p}(\boldsymbol{v}_k)\big) = g^{-1} \cdot \boldsymbol{\varphi}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$$

for all $g \in G$, $p \in P$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in T_p(P)$. $\boldsymbol{\varphi}$ is said to be **tensorial** of **type** ρ if it is pseudotensorial of type ρ and **horizontal** in the sense that $\boldsymbol{\varphi}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = 0 \in \mathcal{V}$ if any one of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in T_p(P)$ is vertical (despite the terminology this notion does not require the existence of a connection on the bundle). A 0-form is taken to be vacuously horizontal so, for these, tensorial and pseudotensorial are the same. They are not the same in general, however, since a connection form $\boldsymbol{\omega}$ is pseudotensorial of type ad, but not tensorial (indeed, ker $\boldsymbol{\omega}_p = \mathrm{Hor}_p(P)$). The curvature $\boldsymbol{\Omega}$ of any connection is tensorial, however. We shall denote the set of all $\mathcal{V}$-valued k-forms on P that are tensorial of type ρ by $\Lambda_\rho^k(P, \mathcal{V})$ and provide it with its obvious real vector space structure.

Remark: The elements of $\Lambda_\rho^0(P, \mathcal{V})$ are in one-to-one correspondence with the cross-sections of the associated bundle $P \times_\rho \mathcal{V}$ (Section 6.8, [**N4**]). These cross-sections are often more convenient from the point of view of physics since they are defined on the base manifold X (e.g., spacetime) rather than the bundle space P. One can do something similar for the elements of any $\Lambda_\rho^k(P, \mathcal{V})$, but, for $k \geq 1$, the objects defined on X that correspond to tensorial

k-forms on P are k-forms taking values in the bundle $P \times_\rho \mathcal{V}$. Thus, for example, the curvature form $\boldsymbol{\Omega}$ of some connection $\boldsymbol{\omega}$ corresponds to a $P \times_{ad} \mathcal{G} = ad\, P$ valued k-form $F_{\boldsymbol{\omega}}$ defined on all of X (see Appendix B of [**N4**] for a few more details).

Before proceeding with a general study of tensorial forms we mention one other special case that will be of particular interest to us in Chapter 6. If the representation ρ of G on $\mathcal{V}$ is the trivial one (i.e., $g \cdot v = v$ for all $g \in G$ and $v \in \mathcal{V}$), then the condition $\sigma_g^* \boldsymbol{\varphi} = g^{-1} \cdot \boldsymbol{\varphi}$ reduces to $\sigma_g^* \boldsymbol{\varphi} = \boldsymbol{\varphi}$. Tensorial forms of this type are characterized by the fact that they project (uniquely) to forms on X, i.e., there is a unique form $\bar{\boldsymbol{\varphi}}$ on X with $\boldsymbol{\varphi} = \mathcal{P}^* \bar{\boldsymbol{\varphi}}$. It will clearly suffice to establish this for real-valued forms.

Lemma 4.5.1 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal G-bundle and $\boldsymbol{\varphi}$ an element of $\Lambda^k(P)$ for some $k \geq 0$ that satisfies*

1. $\boldsymbol{\varphi}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = 0$ *whenever at least one of* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in T_p(P)$ *is vertical.*

2. $\sigma_g^* \boldsymbol{\varphi} = \boldsymbol{\varphi}$ *for every $g \in G$, i.e., for each $p \in P$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in T_p(P)$,*

$$\boldsymbol{\varphi}_{p \cdot g}\Big((\sigma_g)_{*p}(\boldsymbol{v}_1), \ldots, (\sigma_g)_{*p}(\boldsymbol{v}_k) \Big) = \boldsymbol{\varphi}_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$$

   for each $g \in G$.

Then there exists a (necessarily unique) k-form $\bar{\boldsymbol{\varphi}} \in \Lambda^k(X)$ such that $\mathcal{P}^* \bar{\boldsymbol{\varphi}} = \boldsymbol{\varphi}$. Conversely, if $\boldsymbol{\varphi} \in \Lambda^k(P)$ and if there exists a (necessarily unique) $\bar{\boldsymbol{\varphi}} \in \Lambda^k(X)$ with $\mathcal{P}^* \bar{\boldsymbol{\varphi}} = \boldsymbol{\varphi}$, then $\boldsymbol{\varphi}$ must satisfy #1 and #2.

Proof: We ask the reader to prove uniqueness and the necessity of the two conditions.

Exercise 4.5.1 Show that projections, when they exist, must be unique (i.e., that $\mathcal{P}^* \bar{\boldsymbol{\varphi}}_1 = \mathcal{P}^* \bar{\boldsymbol{\varphi}}_2$ implies $\bar{\boldsymbol{\varphi}}_1 = \bar{\boldsymbol{\varphi}}_2$) and that any $\boldsymbol{\varphi}$ which does project to X must satisfy #1 and #2. **Note**: When $k = 0$, #1 is taken to be satisfied vacuously and #2 simply says that $\boldsymbol{\varphi}(p \cdot g) = \boldsymbol{\varphi}(p)$. In this case the unique projection $\bar{\boldsymbol{\varphi}}$ is obviously given by $\bar{\boldsymbol{\varphi}}(x) = \boldsymbol{\varphi}(p)$ for any $p \in \mathcal{P}^{-1}(x)$. Henceforth, we assume $k \geq 1$.

Now, suppose $\boldsymbol{\varphi}$ is a k-form on P satisfying #1 and #2. We define $\bar{\boldsymbol{\varphi}}$ on X as follows: Let $x \in X$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k \in T_x(X)$. Select some $p \in \mathcal{P}^{-1}(x)$ and then select $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in T_p(P)$ with $\mathcal{P}_{*p}(\boldsymbol{v}_i) = \boldsymbol{w}_i$ for each $i = 1, \ldots, k$. Now define

$$\bar{\boldsymbol{\varphi}}_x(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k) = \boldsymbol{\varphi}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k). \tag{4.5.1}$$

Remark: We will prove next that the definition (4.5.1) does not depend on the choices of p and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$. First, however, we point out the most convenient way to actually make these choices in practice. Choose a local cross-section $s : V \longrightarrow \mathcal{P}^{-1}(V)$ with $x \in V$. Let $p = s(x)$ and $\boldsymbol{v}_i = s_{*x}(\boldsymbol{w}_i)$ for $i = 1, \ldots, k$. Then $\mathcal{P}(p) = \mathcal{P}(s(x)) = x$ and $\mathcal{P}_{*p}(\boldsymbol{v}_i) = \mathcal{P}_{*s(x)}(s_{*x}(\boldsymbol{w}_i)) = (\mathcal{P} \circ s)_{*x}(\boldsymbol{w}_i) = (\,\mathrm{id}_V)_{*x}(\boldsymbol{w}_i) = \boldsymbol{w}_i$ as required. Notice that (4.5.1) now gives

$$\bar{\boldsymbol{\varphi}}_x(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k) = \boldsymbol{\varphi}_{s(x)}(s_{*x}(\boldsymbol{w}_1), \ldots, s_{*x}(\boldsymbol{w}_k)) = (s^*\boldsymbol{\varphi})_x(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k)$$

so

$$\bar{\boldsymbol{\varphi}} = s^*\boldsymbol{\varphi} \tag{4.5.2}$$

on V.

To see that our definition is independent of the choices suppose that $p' \in \mathcal{P}^{-1}(x)$ and $\boldsymbol{v}_1', \ldots, \boldsymbol{v}_k' \in T_{p'}(P)$ are such that $\mathcal{P}_{*p'}(\boldsymbol{v}_i') = \boldsymbol{w}_i$ for $i = 1, \ldots, k$. We must show that

$$\boldsymbol{\varphi}_{p'}(\boldsymbol{v}_1', \ldots, \boldsymbol{v}_k') = \boldsymbol{\varphi}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k). \tag{4.5.3}$$

There exists a $g \in G$ such that $p' = p \cdot g$ (Lemma 4.1.1, [**N4**]). Thus,

$$\boldsymbol{\varphi}_{p'}(\boldsymbol{v}_1', \ldots, \boldsymbol{v}_k') = \boldsymbol{\varphi}_{p \cdot g}(\boldsymbol{v}_1', \ldots, \boldsymbol{v}_k') = \boldsymbol{\varphi}_p(\boldsymbol{v}_1'', \ldots, \boldsymbol{v}_k''), \tag{4.5.4}$$

where

$$\boldsymbol{v}_i'' = (\sigma_g)_{*p}^{-1}(\boldsymbol{v}_i') = (\sigma_{g^{-1}})_{*p \cdot g}(\boldsymbol{v}_i'), \quad i = 1, \ldots, k$$

(by assumption #2).

Exercise 4.5.2 Show that $\mathcal{P}_{*p}(\boldsymbol{v}_i'') = \boldsymbol{w}_i$ for $i = 1, \ldots, k$.

Thus, $\mathcal{P}_{*p}(\boldsymbol{v}_i - \boldsymbol{v}_i'') = 0$ so $\boldsymbol{v}_i - \boldsymbol{v}_i''$ is vertical for $i = 1, \ldots, k$. By assumption #1, $\boldsymbol{\varphi}_p(\boldsymbol{v}_1 - \boldsymbol{v}_1'', \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k) = 0$ so

$$\boldsymbol{\varphi}_p(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k) = \boldsymbol{\varphi}_p\left(\boldsymbol{v}_1'', \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\right).$$

Similarly, $\boldsymbol{\varphi}_p(\boldsymbol{v}_1'', \boldsymbol{v}_2 - \boldsymbol{v}_2'', \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k) = 0$ so $\boldsymbol{\varphi}_p\left(\boldsymbol{v}_1'', \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k\right) = \boldsymbol{\varphi}_p\left(\boldsymbol{v}_1'', \boldsymbol{v}_2'', \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k\right)$ and therefore

$$\boldsymbol{\varphi}_p(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k) = \boldsymbol{\varphi}_p\left(\boldsymbol{v}_1'', \boldsymbol{v}_2'', \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k\right).$$

Continuing in this way gives

$$\boldsymbol{\varphi}_p(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k) = \boldsymbol{\varphi}_p\left(\boldsymbol{v}_1'', \boldsymbol{v}_2'', \boldsymbol{v}_3'', \ldots, \boldsymbol{v}_k''\right)$$

which, by (4.5.4), proves (4.5.3).

We conclude that $\bar{\boldsymbol{\varphi}}$ is well-defined. It satisfies $\mathcal{P}^*\bar{\boldsymbol{\varphi}} = \boldsymbol{\varphi}$ by definition and is clearly a k-form on X so we need only verify smoothness. But this can be proved locally and so follows from (4.5.2). ∎

We will have quite a bit more to say about projectable forms on bundles, but first we derive some general results on derivatives of tensorial forms. First we show that the exterior derivative of a pseudotensorial form is itself pseudotensorial.

Lemma 4.5.2 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle, $\mathcal{V}$ a finite dimensional real vector space, $\rho : G \longrightarrow GL(\mathcal{V})$ a representation of G on $\mathcal{V}$ and $\boldsymbol{\varphi}$ a $\mathcal{V}$-valued k-form on P that is pseudotensorial of type ρ. Then $d\boldsymbol{\varphi}$ is pseudotensorial of type ρ.*

Proof: $\boldsymbol{\varphi}$ satisfies $\sigma_g^* \boldsymbol{\varphi} = g^{-1} \cdot \boldsymbol{\varphi}$ for every $g \in G$ and we must show that $\sigma_g^*(d\boldsymbol{\varphi}) = g^{-1} \cdot d\boldsymbol{\varphi}$. But d commutes with pullback and the action of G on $\mathcal{V}$ is linear so

$$\sigma_g^*(d\boldsymbol{\varphi}) = d(\sigma_g^* \boldsymbol{\varphi}) = d(g^{-1} \cdot \boldsymbol{\varphi}) = g^{-1} \cdot d\boldsymbol{\varphi}. \qquad \blacksquare$$

The exterior derivative of a tensorial form is therefore pseudotensorial, but need not be horizontal. To obtain a differentiation operator that carries tensorial forms to tensorial forms requires the existence of a connection on $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$. With such a connection any pseudotensorial form $\boldsymbol{\tau}$ gives rise to a tensorial form by allowing $\boldsymbol{\tau}$ to operate only on horizontal parts.

Lemma 4.5.3 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with a connection $\boldsymbol{\omega}, \mathcal{V}$ a finite dimensional real vector space, $\rho : G \longrightarrow GL(\mathcal{V})$ a representation of G on $\mathcal{V}$ and $\boldsymbol{\tau}$ a $\mathcal{V}$-valued k-form on P that is pseudotensorial of type ρ. Then the $\mathcal{V}$-valued k-form $\boldsymbol{\tau}^H$ on P defined by*

$$\boldsymbol{\tau}^H(p)(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \boldsymbol{\tau}(p)\big(\boldsymbol{v}_1^H, \ldots, \boldsymbol{v}_k^H\big)$$

is tensorial of type ρ ($\boldsymbol{v}^H$ is the horizontal part of $\boldsymbol{v} \in T_p(P)$; see page 36).

Exercise 4.5.3 Prove Lemma 4.5.3. **Hint**: By (6.8.3) of **[N4]**, $((\sigma_g)_{*p}(\boldsymbol{v}))^H = (\sigma_g)_{*p}(\boldsymbol{v}^H)$. $\blacksquare$

Theorem 4.5.4 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with a connection $\boldsymbol{\omega}, \mathcal{V}$ a finite dimensional real vector space, $\rho : G \longrightarrow GL(\mathcal{V})$ a representation of G on $\mathcal{V}$ and $\boldsymbol{\varphi}$ a $\mathcal{V}$-valued k-form on P that is pseu-dotensorial of type ρ. Then the covariant exterior derivative $d^{\boldsymbol{\omega}}\boldsymbol{\varphi}$ of $\boldsymbol{\varphi}$, defined by*

$$\begin{aligned}(d^{\boldsymbol{\omega}}\boldsymbol{\varphi})(p)(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}) &= (d\boldsymbol{\varphi})^H(p)\big(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}\big) \\ &= d\boldsymbol{\varphi}(p)\Big(\boldsymbol{v}_1^H, \ldots, \boldsymbol{v}_{k+1}^H\Big)\end{aligned}$$

is a tensorial $(k+1)$-form of type ρ. In particular,

$$d^{\boldsymbol{\omega}} : \Lambda_\rho^k(P, \mathcal{V}) \longrightarrow \Lambda_\rho^{k+1}(P, \mathcal{V}).$$

Remark: The curvature $\boldsymbol{\Omega}$ of a connection $\boldsymbol{\omega}$ is, of course, just its covariant exterior derivative ($\boldsymbol{\Omega} = d^{\boldsymbol{\omega}}\boldsymbol{\omega}$). Another special case that we have had occasion to consider previously is the covariant exterior derivative of a matter field (element of $\Lambda^0_\rho(X, \mathcal{V})$). The Cartan Structure Equation provides a convenient computational formula for $d^{\boldsymbol{\omega}}\boldsymbol{\omega}$ and (6.8.4) of [**N4**] gives an analogous formula for matter fields. Shortly we will generalize both of these, but first we observe that, for forms that project to the base manifold $X, d^{\boldsymbol{\omega}}\boldsymbol{\varphi}$ and $d\boldsymbol{\varphi}$ coincide.

Lemma 4.5.5 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with a connection $\boldsymbol{\omega}$ and $\boldsymbol{\varphi}$ a k-form on P that projects to a k-form $\bar{\boldsymbol{\varphi}}$ on X (i.e., $\boldsymbol{\varphi} = \mathcal{P}^*\bar{\boldsymbol{\varphi}}$). Then*

$$d^{\boldsymbol{\omega}}\boldsymbol{\varphi} = d\boldsymbol{\varphi}.$$

Proof: We simply compute

$$\begin{aligned}
(d\boldsymbol{\varphi})_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}) &= (d(\mathcal{P}^*\bar{\boldsymbol{\varphi}}))_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}) \\
&= (\mathcal{P}^*(d\bar{\boldsymbol{\varphi}}))_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}) \\
&= (d\bar{\boldsymbol{\varphi}})_{\mathcal{P}(p)}\big(\mathcal{P}_{*p}(\boldsymbol{v}_1), \ldots, \mathcal{P}_{*p}(\boldsymbol{v}_{k+1})\big) \\
&= (d\bar{\boldsymbol{\varphi}})_{\mathcal{P}(p)}\,\big(\mathcal{P}_{*p}(\boldsymbol{v}_1^H), \ldots, \mathcal{P}_{*p}(\boldsymbol{v}_{k+1}^H)\big) \\
&= (\mathcal{P}^*(d\bar{\boldsymbol{\varphi}}))_p\big(\boldsymbol{v}_1^H, \ldots, \boldsymbol{v}_{k+1}^H\big) \\
&= (d(P^*\bar{\boldsymbol{\varphi}}))_p\big(\boldsymbol{v}_1^H, \ldots, \boldsymbol{v}_{k+1}^H\big) \\
&= (d\boldsymbol{\varphi})_p\big(\boldsymbol{v}_1^H, \ldots, \boldsymbol{v}_{k+1}^H\big) \\
&= (d^{\boldsymbol{\omega}}\boldsymbol{\varphi})_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}).
\end{aligned}$$

■

Now, to find the promised computational formula for $d^{\boldsymbol{\omega}}\boldsymbol{\varphi}$ we begin as in Section 6.8 of [**N4**] for the special case of matter fields by introducing yet another wedge product for $\mathcal{V}$-valued forms. We consider the bilinear map from $\mathcal{G} \times \mathcal{V}$ to $\mathcal{V}$ which sends $(A, v) \in \mathcal{G} \times \mathcal{V}$ to $A \cdot v \in \mathcal{V}$ defined by

$$A \cdot v = \frac{d}{dt}\,(\exp(tA) \cdot v)|_{t=0} = \frac{d}{dt}\,(\rho(\exp(tA))(v))|_{t=0},$$

where $\rho : G \longrightarrow GL(\mathcal{V})$ is the representation relative to which our forms are tensorial. Notice that if ρ is just the natural representation of $G \subseteq GL(k, \ )$ on ^{k} (matrix multiplication), then $A \cdot v = Av$ is also matrix multiplication for each $A \in \mathcal{G}$ (Exercise 6.8.6, [**N4**]). A special case of more immediate concern to us here is the following: Suppose $\rho = ad : G \longrightarrow GL(\mathcal{G})$ is the adjoint representation. Then, for any $A, B \in \mathcal{G}$,

$$\begin{aligned} A \cdot B &= \frac{d}{dt}\Big(ad_{\exp(tA)}(B)\Big)\Big|_{t=0} \\ &= \lim_{t\longrightarrow 0}\frac{1}{t}\Big(ad_{\exp(tA)}(B) - B\Big) \\ &= [A,B], \end{aligned}$$

where this last equality follows from the definition of the Lie bracket and is proved on pages 288–289 of [**N4**].

Now, the bilinear map $(A,v) \longrightarrow A \cdot v$ of $\mathcal{G} \times \mathcal{V}$ to $\mathcal{V}$ determines a wedge product for $\mathcal{G}$-valued forms and $\mathcal{V}$-valued forms. Specifically, if $\boldsymbol{\alpha} \in \Lambda^k(P,\mathcal{G})$ and $\boldsymbol{\beta} \in \Lambda^l(P,\mathcal{V})$, then we define $\boldsymbol{\alpha}\dot{\wedge}\boldsymbol{\beta} \in \Lambda^{k+l}(P,\mathcal{V})$ at each point of P by

$$\begin{aligned} (\boldsymbol{\alpha}\dot{\wedge}\boldsymbol{\beta})\big(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+l}\big) = &\frac{1}{k!\,l!}\sum_{\sigma}(-1)^{\sigma}\boldsymbol{\alpha}\Big(\boldsymbol{v}_{\sigma(1)},\ldots,\boldsymbol{v}_{\sigma(k)}\Big) \\ &\cdot\boldsymbol{\beta}\Big(\boldsymbol{v}_{\sigma(k+1)},\ldots,\boldsymbol{v}_{\sigma(k+l)}\Big), \end{aligned}$$

where the sum is over all permutations $\sigma \in S_{k+l}$ of $\{1,\ldots,k+l\}$. Notice, in particular, that when ρ is the adjoint representation of G on $\mathcal{G}$, then $\boldsymbol{\alpha}\dot{\wedge}\boldsymbol{\beta}$ is just the bracket wedge product $[\boldsymbol{\alpha},\boldsymbol{\beta}]$ described in Section 4.2. We are now in a position to prove our major result.

Theorem 4.5.6 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with a connection $\boldsymbol{\omega}$, $\mathcal{V}$ a finite dimensional real vector space, $\rho : G \longrightarrow GL(\mathcal{V})$ a representation of G on $\mathcal{V}$ and $\boldsymbol{\varphi} \in \Lambda^k_\rho(P,\mathcal{V})$ a $\mathcal{V}$-valued, tensorial k-form of type ρ. Then*

$$d^{\boldsymbol{\omega}}\boldsymbol{\varphi} = d\boldsymbol{\varphi} + \boldsymbol{\omega}\dot{\wedge}\boldsymbol{\varphi}.$$

Proof: We must show that, for each $p \in P$ and all $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1} \in T_p(P)$,

$$\begin{aligned} (d\boldsymbol{\varphi})_p\Big(\boldsymbol{v}_1^H,\ldots,\boldsymbol{v}_{k+1}^H\Big) = &(d\boldsymbol{\varphi})_p(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1}) \\ &+\frac{1}{k!}\sum_{\sigma\in S_{k+1}}(-1)^{\sigma}\boldsymbol{\omega}_p\Big(\boldsymbol{v}_{\sigma(1)}\Big) \\ &\cdot\boldsymbol{\varphi}_p\Big(\boldsymbol{v}_{\sigma(2)},\ldots,\boldsymbol{v}_{\sigma(k+1)}\Big). \end{aligned} \tag{4.5.5}$$

We consider three cases separately.

I. Each of $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1}$ is horizontal.

In this case $\boldsymbol{v}_i^H = \boldsymbol{v}_i$ for $i = 1,\ldots,k+1$ and $\boldsymbol{\omega}_p(\boldsymbol{v}_i) = 0$ for $i = 1,\ldots,k+1$ so both sides of (4.5.5) are just $(d\boldsymbol{\varphi})_p(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1})$.

II. Two or more of $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1}$ are vertical.

The left-hand side of (4.5.5) is now zero since $\boldsymbol{v}_i^H = 0$ for at least two values of i. Moreover, the sum on the right-hand side is zero since $\boldsymbol{\varphi}_p$ is horizontal and at least one of $\boldsymbol{v}_{\sigma(2)}, \ldots, \boldsymbol{v}_{\sigma(k+1)}$ is vertical for each σ. Thus, we need only show that $(d\,\boldsymbol{\varphi})_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1})$ is zero as well. Each of the vertical vectors among $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}$ can be extended to fundamental vector fields on P (Corollary 5.8.9, [**N4**]). Extend the remaining $\boldsymbol{v}_i$ to vector fields on P in any manner (e.g., constant components on some coordinate neighborhood and then use Exercise 5.7.2, [**N4**]). Denote these vector fields $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{k+1}$. Now, $(d\,\boldsymbol{\varphi})_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1})$ is the value at p of

$$d\,\boldsymbol{\varphi}\,(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{k+1}) = \sum_{i=1}^{k+1} (-1)^{i+1}\; \boldsymbol{V}_i\left(\boldsymbol{\varphi}\left(\boldsymbol{V}_1, \ldots, \widehat{\boldsymbol{V}}_i, \ldots, \boldsymbol{V}_{k+1}\right)\right) + \sum_{1 \le i < j \le n} (-1)^{i+j}\, \boldsymbol{\varphi}\left([\boldsymbol{V}_i, \boldsymbol{V}_j], \boldsymbol{V}_1, \ldots, \widehat{\boldsymbol{V}}_i, \ldots, \widehat{\boldsymbol{V}}_j, \ldots, \boldsymbol{V}_{k+1}\right)$$

(componentwise in $\mathcal{V}$). The first sum vanishes since at least one of $\boldsymbol{V}_1, \ldots, \widehat{\boldsymbol{V}}_i, \ldots, \boldsymbol{V}_{k+1}$ is vertical and $\boldsymbol{\varphi}$ is horizontal. The same is true of the second sum since the Lie bracket of two fundamental vector fields is another fundamental vector field (Theorem 5.8.8, [**N4**]) and so is vertical.

III. Precisely one of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1}$ is vertical and the rest are horizontal.

Remark: Establishing (4.5.5) in this case will actually complete the proof since both sides are multilinear and any $\boldsymbol{v}$ can be written as $\boldsymbol{v} = \boldsymbol{v}^H + \boldsymbol{v}^V$.

We may assume that $\boldsymbol{v}_1$ is vertical and $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k+1}$ are horizontal. Extend $\boldsymbol{v}_1$ to a fundamental vector field $\boldsymbol{V}_1$. We wish to extend $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k+1}$ to horizontal, $(\sigma_g)_*$-invariant vector fields on P and for this we need a lemma on horizontal lifts of vector fields.

Lemma 4.5.7 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with a connection $\boldsymbol{\omega}$ and let $\boldsymbol{W}$ be a smooth vector field on X. Then there is a unique smooth vector field $\tilde{\boldsymbol{W}}$ on P that satisfies*

1. *$\tilde{\boldsymbol{W}}(p)$ is horizontal for each $p \in P$.*
2. *$\mathcal{P}_{*p}\,(\tilde{\boldsymbol{W}}(p)) = \boldsymbol{W}(\mathcal{P}(p))$ for each $p \in P$.*
3. *$(\sigma_g)_{*p}\,(\tilde{\boldsymbol{W}}(p)) = \tilde{\boldsymbol{W}}(p \cdot g)$ for each $p \in P$ and $g \in G$.*

Proof: We claim first that if we establish the existence of a unique $\tilde{\boldsymbol{W}}$ satisfying (1) and (2), then (3) necessarily follows. Indeed, since σ_g is a diffeomorphism and $(\sigma_g)_{*p}\,(\mathrm{Hor}_p(P)) = \mathrm{Hor}_{p\cdot g}(P)$ ((6.1.3), [**N4**]), $(\sigma_g)_*\,\tilde{\boldsymbol{W}}$ (defined by $((\sigma_g)_*\,\tilde{\boldsymbol{W}})_p = (\sigma_g)_{*p\cdot g^{-1}}\,(\tilde{\boldsymbol{W}}_{p\cdot g^{-1}}))$ is a smooth horizontal vector field and

$$\begin{aligned}\mathcal{P}_{*p}\left(\left((\sigma_g)_*\tilde{\boldsymbol{W}}\right)_p\right) &= \mathcal{P}_{*p}\left((\sigma_g)_{*p\cdot g^{-1}}\left(\tilde{\boldsymbol{W}}_{p\cdot g^{-1}}\right)\right)\\ &= (\mathcal{P}\circ\sigma_g)_{*p\cdot g^{-1}}\left(\tilde{\boldsymbol{W}}_{p\cdot g^{-1}}\right)\\ &= \mathcal{P}_{*p\cdot g^{-1}}\left(\tilde{\boldsymbol{W}}_{p\cdot g^{-1}}\right) = \boldsymbol{W}\left(\mathcal{P}(p\cdot g^{-1})\right)\\ &= \boldsymbol{W}\left(\mathcal{P}(p)\right)\end{aligned}$$

so uniqueness implies $(\sigma_g)_*\tilde{\boldsymbol{W}} = \tilde{\boldsymbol{W}}$ so $((\sigma_g)_*\tilde{\boldsymbol{W}})_{p\cdot g} = \tilde{\boldsymbol{W}}(p\cdot g)$, i.e., $(\sigma_g)_{*p}(\tilde{\boldsymbol{W}}(p)) = \tilde{\boldsymbol{W}}(p\cdot g)$ as required.

Now, since $\mathcal{P}_{*p} : \text{Hor}_p(P) \longrightarrow T_{\mathcal{P}(p)}(X)$ is an isomorphism (Exercise 6.1.7, [**N4**]), there is only one possible definition of $\tilde{\boldsymbol{W}}$: For each $p \in P, \mathcal{P}(p) \in X$ and there exists a unique $\tilde{\boldsymbol{W}}(p) \in \text{Hor}_p(P)$ such that $\mathcal{P}_{*p}(\tilde{\boldsymbol{W}}(p)) = \boldsymbol{W}(\mathcal{P}(p))$. Thus, $\tilde{\boldsymbol{W}}$ is obviously a vector field on P satisfying (1) and (2), but is not obviously smooth. We prove smoothness locally. Thus, we select a trivialization $\Psi : \mathcal{P}^{-1}(V) \longrightarrow V\times G$. There is obviously a smooth vector field on $V\times G$ which lifts the restriction $\boldsymbol{W}|V$ of $\boldsymbol{W}$ to V. Transferring this to $\mathcal{P}^{-1}(V)$ via Ψ_*^{-1} we obtain a smooth vector field on $\mathcal{P}^{-1}(V)$ which lifts $\boldsymbol{W}|V$. The horizontal part of this vector field is smooth on $\mathcal{P}^{-1}(V)$ (Exercise 6.2.2, [**N4**]), and lifts $\boldsymbol{W}|V$ so it must agree with $\tilde{\boldsymbol{W}}$ on $\mathcal{P}^{-1}(V)$. Thus, $\tilde{\boldsymbol{W}}$ is smooth on $\mathcal{P}^{-1}(V)$. Since the trivialization was arbitrary, $\tilde{\boldsymbol{W}}$ is smooth. ■

Now we return to the proof of Theorem 4.5.6 in Case III. The vectors $\mathcal{P}_{*p}(\boldsymbol{v}_2),\ldots,\mathcal{P}_{*p}(\boldsymbol{v}_{k+1})$ extend to vector fields on X and these, by Lemma 4.5.7, have unique, smooth, horizontal, $(\sigma_g)_*$-invariant lifts $\boldsymbol{V}_2,\ldots,\boldsymbol{V}_{k+1}$ to P.

Exercise 4.5.4 Show that $\boldsymbol{V}_i(p) = \boldsymbol{v}_i$ for $i = 2,\ldots,k+1$.

Since $\boldsymbol{V}_1$ is a fundamental vector field and each $\boldsymbol{V}_2,\ldots,\boldsymbol{V}_{k+1}$ is horizontal, all of the Lie brackets $[\boldsymbol{V}_1, \boldsymbol{V}_i]$, $i = 2,\ldots,k+1$, are horizontal (page 349, [**N4**]).

Exercise 4.5.5 Show that, in fact, $[\boldsymbol{V}_1, \boldsymbol{V}_2] = \cdots = [\boldsymbol{V}_1, \boldsymbol{V}_{k+1}| = 0$.

From Exercise 4.5.5 and the fact that $\boldsymbol{V}_1$ is vertical and $\boldsymbol{\varphi}$ is horizontal we conclude that

$$\begin{aligned}d\boldsymbol{\varphi}(\boldsymbol{V}_1, \boldsymbol{V}_2,\ldots,\boldsymbol{V}_{k+1}) &= (-1)^{1+1}\boldsymbol{V}_1\Big(\boldsymbol{\varphi}(\boldsymbol{V}_2,\ldots,\boldsymbol{V}_{k+1})\Big) + 0\\ &= \boldsymbol{V}_1\Big(\boldsymbol{\varphi}(\boldsymbol{V}_2,\ldots,\boldsymbol{V}_{k+1})\Big).\end{aligned}$$

Since the left-hand side of (4.5.5) is obviously zero in Case III, the proof now reduces to showing that

$$0 = \boldsymbol{V}_1\Big(\boldsymbol{\varphi}\,(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1})\Big) + \frac{1}{k!}\sum_{\sigma\,\in\, S_{k+1}} (-1)^{\sigma}\,\boldsymbol{\omega}\Big(\boldsymbol{V}_{\sigma(1)}\Big)\cdot\boldsymbol{\varphi}\Big(\boldsymbol{V}_{\sigma(2)}, \ldots, \boldsymbol{V}_{\sigma(k+1)}\Big).$$

But since $\boldsymbol{\omega}(\boldsymbol{V}_{\sigma(1)})$ will be zero whenever $\sigma(1) \neq 1$ we may rewrite this as

$$0 = \boldsymbol{V}_1\Big(\boldsymbol{\varphi}\,(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1})\Big) + \frac{1}{k!}\sum_{\substack{\sigma\in S_{k+1}\\ \sigma(1)=1}} (-1)^{\sigma}\,\boldsymbol{\omega}(\boldsymbol{V}_1)\cdot\boldsymbol{\varphi}\Big(\boldsymbol{V}_{\sigma(2)}, \ldots, \boldsymbol{V}_{\sigma(k+1)}\Big)$$

$$0 = \boldsymbol{V}_1\Big(\boldsymbol{\varphi}\,(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1})\Big) + \boldsymbol{\omega}(\boldsymbol{V}_1)\cdot\left\{\frac{1}{k!}\sum_{\substack{\sigma\in S_{k+1}\\ \sigma(1)=1}} (-1)^{\sigma}\boldsymbol{\varphi}\Big(\boldsymbol{V}_{\sigma(2)}, \ldots, \boldsymbol{V}_{\sigma(k+1)}\Big)\right\}$$

$$0 = \boldsymbol{V}_1\Big(\boldsymbol{\varphi}\,(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1})\Big) + \boldsymbol{\omega}(\boldsymbol{V}_1)\cdot\boldsymbol{\varphi}(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1}). \tag{4.5.6}$$

To prove (4.5.6) we proceed as follows: Since $\boldsymbol{V}_1$ is a fundamental vector field we may write $\boldsymbol{V}_1 = A^{\#}$ for some $A \in \mathcal{G}$. Also let $\alpha(t) = \exp(tA)$ for $t \in$. Then, for each $p \in P$,

$$\boldsymbol{V}_1(p) = A^{\#}(p) = (\sigma_p \circ \alpha)'\,(0),$$

where $\sigma_p : G \longrightarrow P$ is given by $\sigma_p(g) = p \cdot g$ ((5.8.8), [**N4**]). Thus, for any $f \in C^{\infty}(P)$,

$$\boldsymbol{V}_1(p)\,(f) = (\sigma_p \circ \alpha)'\,(0)\,(f) = (f \circ \sigma_p \circ \alpha)'\,(0).$$

To compute $\boldsymbol{V}_1\,(p)(\boldsymbol{\varphi}(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1}))$ we let $f = \boldsymbol{\varphi}(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1})$. Then $f \circ \sigma_p \circ \alpha$ is given by

$$\begin{aligned}
\boldsymbol{\varphi}\big(\boldsymbol{V}_2, \ldots, \boldsymbol{V}_{k+1}\big)\big(\sigma_p(\alpha(t))\big) &= \boldsymbol{\varphi}_{p\cdot\alpha(t)}\big(\boldsymbol{V}_2(p\cdot\alpha(t)), \ldots, \boldsymbol{V}_{k+1}(p\cdot\alpha(t))\big)\\
&= \boldsymbol{\varphi}_{p\cdot\alpha(t)}\Big(\Big(\sigma_{\alpha(t)}\Big)_{*p}(\boldsymbol{V}_2(p)), \ldots, \Big(\sigma_{\alpha(t)}\Big)_{*p}\Big(\boldsymbol{V}_{k+1}(p)\Big)\Big)\\
&\qquad\qquad \text{(by Lemma 4.5.7 (3))}\\
&= \alpha(t)^{-1}\cdot\boldsymbol{\varphi}_p\big(\boldsymbol{V}_2(p), \ldots, \boldsymbol{V}_{k+1}(p)\big)\\
&\qquad\qquad \big(\text{since } \boldsymbol{\varphi} \in \Lambda^k_{\rho}\,(P, \mathcal{V})\big)\\
&= \exp(-tA)\cdot\boldsymbol{\varphi}_p\big(\boldsymbol{V}_2(p), \ldots, \boldsymbol{V}_{k+1}(p)\big).
\end{aligned}$$

Thus,

$$\begin{aligned}\boldsymbol{V}_1(p)\Big(\boldsymbol{\varphi}\,(\boldsymbol{V}_2,\ldots,\boldsymbol{V}_{k+1})\Big) &= \frac{d}{dt}\Big(\exp\,(t(-A))\cdot\boldsymbol{\varphi}_p\,(\boldsymbol{V}_2(p),\ldots,\boldsymbol{V}_{k+1}(p))\Big)\Big|_{t=0}\\ &= (-A)\cdot\boldsymbol{\varphi}_p\,\big(\boldsymbol{V}_2(p),\ldots,\boldsymbol{V}_{k+1}(p)\big)\\ &= -\boldsymbol{\omega}_p(A^{\#})\cdot\boldsymbol{\varphi}_p\big(\boldsymbol{V}_2(p),\ldots,\boldsymbol{V}_{k+1}(p)\big)\\ &= -\boldsymbol{\omega}_p(\boldsymbol{V}_1(p))\cdot\boldsymbol{\varphi}_p\big(\boldsymbol{V}_2(p),\ldots,\boldsymbol{V}_{k+1}(p)\big)\end{aligned}$$

which gives (4.5.6) at the arbitrary point $p \in P$ and therefore completes the proof. ■

Corollary 4.5.8 *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with a connection $\boldsymbol{\omega}$ and $ad : G \longrightarrow GL(\mathcal{G})$ the adjoint representation of G on $\mathcal{G}$. Then, for each $\boldsymbol{\varphi} \in \Lambda^k_{ad}(P,\mathcal{G})$,*

$$d^{\boldsymbol{\omega}}\,\boldsymbol{\varphi} = d\,\boldsymbol{\varphi} + [\omega,\boldsymbol{\varphi}].$$

Notice that the Corollary does not apply to $\boldsymbol{\omega}$ itself, which is pseudotensorial of type ad, but not horizontal. Indeed, the Cartan Structure Equation gives $d^{\boldsymbol{\omega}}\,\boldsymbol{\omega} = \boldsymbol{\Omega} = d\,\boldsymbol{\omega} + \frac{1}{2}[\boldsymbol{\omega},\boldsymbol{\omega}]$. However, the Corollary does apply to the curvature $\boldsymbol{\Omega}$ and gives $d^{\boldsymbol{\omega}}\,\boldsymbol{\Omega} = d\,\boldsymbol{\Omega} + [\boldsymbol{\omega},\boldsymbol{\Omega}]$ which we now show is identically zero.

Theorem 4.5.9 (Bianchi Identity) *Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a smooth principal bundle with connection $\boldsymbol{\omega}$ and curvature $\boldsymbol{\Omega} = d^{\boldsymbol{\omega}}\boldsymbol{\omega}$. Then*

$$d^{\boldsymbol{\omega}}\,\boldsymbol{\Omega} = 0. \tag{4.5.7}$$

Proof: We simply compute, from Corollary 4.5.8,

$$\begin{aligned}d^{\boldsymbol{\omega}}\,\boldsymbol{\Omega} &= d\,\boldsymbol{\Omega} + [\boldsymbol{\omega},\boldsymbol{\Omega}]\\ &= d\left(d\,\boldsymbol{\omega} + \frac{1}{2}[\boldsymbol{\omega},\boldsymbol{\omega}]\right) + \left[\boldsymbol{\omega}, d\,\boldsymbol{\omega} + \frac{1}{2}[\boldsymbol{\omega},\boldsymbol{\omega}]\right]\\ &= d\,(d\,\boldsymbol{\omega}) + \frac{1}{2}\,d\,([\boldsymbol{\omega},\boldsymbol{\omega}]) + [\boldsymbol{\omega}, d\,\boldsymbol{\omega}] + \frac{1}{2}[\boldsymbol{\omega},[\boldsymbol{\omega},\boldsymbol{\omega}]]\\ &= \frac{1}{2}\,d\,([\boldsymbol{\omega},\boldsymbol{\omega}]) + [\boldsymbol{\omega}, d\,\boldsymbol{\omega}]\\ &= \frac{1}{2}\,([d\,\boldsymbol{\omega},\boldsymbol{\omega}] - [\boldsymbol{\omega}, d\,\boldsymbol{\omega}]) + [\boldsymbol{\omega}, d\,\boldsymbol{\omega}]\\ &= \frac{1}{2}\,(-[\boldsymbol{\omega}, d\,\boldsymbol{\omega}] - [\boldsymbol{\omega}, d\,\boldsymbol{\omega}]) + [\boldsymbol{\omega}, d\,\boldsymbol{\omega}] = 0.\end{aligned}$$

■

Since $[\boldsymbol{\omega},\boldsymbol{\Omega}] = -[\boldsymbol{\Omega},\boldsymbol{\omega}]$ one can write (4.5.7) as

$$d\,\boldsymbol{\Omega} = [\boldsymbol{\Omega},\boldsymbol{\omega}]. \tag{4.5.8}$$

4.6 Integration on Manifolds

We begin by reviewing a little calculus. Consider a function $f : \quad \longrightarrow$ which, for simplicity, we will temporarily assume to be continuous. If $[a,b]$ is an interval in , then the integral

$$\int_{[a,b]} f$$

of f over $[a,b]$ is defined (either as a Riemann or Lebesgue integral) as a limit of certain sums. It is invariably computed, however, by appealing to the Fundamental Theorem of Calculus which asserts that there exists a differentiable function $F : \quad \longrightarrow \quad$ with $F' = f$ and that, for any such F,

$$\int_{[a,b]} f = F \,|_{\,a}^{\,b} = F(b) - F(a).$$

The task of finding such an antiderivative is often facilitated by the Change of Variables Formula, the content of which is as follows: Let g be a diffeomorphism (actually, one-to-one and continuously differentiable will suffice) on some open set containing an interval $[\alpha,\beta]$ with $g([\alpha,\beta]) = [a,b]$. Then

$$\int_{[a,b]} f = \int_{[\alpha,\beta]} (f \circ g)\,|g'|.$$

In particular, if g is orientation preserving (increasing), then

$$\int_{[a,b]} f = \int_{[\alpha,\beta]} (f \circ g)\,g'. \tag{4.6.1}$$

Calculus students are provided with ample opportunity to persuade themselves of the efficacy of this formula. Suppose, for example, that the problem is to integrate over $[0,1]$ the function f whose standard coordinate representation is $f(x) = \frac{1}{\sqrt{1+x^2}}$. Define $g : (-\frac{\pi}{2}, \frac{\pi}{2}) \longrightarrow \quad$ by $g = \tan$. Then $g([0,\frac{\pi}{4}]) = [0,1]$ and, on $[0,\frac{\pi}{4}]$, $f \circ g = \cos$ and $g' = \sec^2$ so

$$\int_{[0,1]} f = \int_{[0,\frac{\pi}{4}]} \sec = \ln|\sec + \tan|\,\Big|_0^{\frac{\pi}{4}} = \ln\left(1 + \sqrt{2}\right).$$

Of course, calculus students would generally not phrase their calculations in just these terms. Perhaps something more along the following lines:

$$\int_0^1 \frac{1}{\sqrt{1+x^2}}\,dx = \int_{\tan^{-1}0}^{\tan^{-1}1} \frac{1}{\sqrt{1+\tan^2 u}}\,\sec^2 u\,du$$

$$\begin{aligned} x &= \tan u \\ dx &= \sec^2 u\,du \end{aligned} \qquad = \int_0^{\frac{\pi}{4}} \sec u\,du$$

$$= \ln|\sec u + \tan u|\,\Big|_0^{\frac{\pi}{4}}$$

$$= \ln\left(1+\sqrt{2}\right).$$

There is much to be said for this. The student's calculations, which, on the surface, appear more formal and less rigorous, are, in fact, entirely rigorous once one has faced the fact that it is not functions, but 1-forms that should be integrated over intervals in . Specifically, any $f :\ \longrightarrow\ $ gives rise to a unique 1-form $\boldsymbol{\omega} = f\,dx$ on and, if g is an orientation preserving diffeomorphism with $g([\alpha,\beta]) = [a,b]$, then

$$g^*\,\boldsymbol{\omega} = g^*\,(f\,dx) = (f\circ g)\,g^*\,(dx) = (f\circ g)\,g'\,dx.$$

If we define the integral of any 1-form on to be the integral of its standard coordinate function, then the Change of Variables Formula becomes

$$\int_{[a,b]} \boldsymbol{\omega} = \int_{[\alpha,\beta]} g^*\,\boldsymbol{\omega}. \tag{4.6.2}$$

Our calculus student has computed (unwittingly, perhaps) the pullback of the 1-form $f(x) = \frac{1}{\sqrt{1+x^2}}\,dx$ by g and found the interval corresponding to $[0,1]$ under g. The Change of Variables Formula is built into the properties of 1-forms.

This is amusing enough, but scarcely sufficient justification for our claim that "it is not functions, but 1-forms that should be integrated over intervals in ." Much more significant is the fact that the integral of a 1-form, which we defined in terms of its standard coordinate representation, is actually *independent of coordinates* and so will generalize immediately to arbitrary 1-manifolds (where there is no "standard coordinate"). This too is a consequence of (4.6.1), i.e., of the Change of Variables Formula. For the proof we regard as an abstract 1-manifold and consider two oriented charts (U,φ) and (V,ψ) on with coordinate functions x and y, respectively. To simplify the argument we will temporarily assume that $[a,b] \subseteq U \cap V$. Write $\boldsymbol{\omega} = h\,dx$ on U and $\boldsymbol{\omega} = k\,dy$ on V. The coordinate expressions for h and k are therefore $h\circ\varphi^{-1} : \varphi(U) \longrightarrow\ $ and $k\circ\psi^{-1} : \psi(V) \longrightarrow\ $ (if (U,φ) is the standard chart on , then $h\circ\varphi^{-1}$ is the f in our earlier discussion). If we suppose that

$\varphi^{-1}([\alpha_1, \beta_1]) = [a, b]$ and $\psi^{-1}([\alpha_2, \beta_2]) = [a, b]$, then our claim is that

$$\int_{[\alpha_1,\beta_1]} h \circ \varphi^{-1} = \int_{[\alpha_2,\beta_2]} k \circ \psi^{-1}$$

so that $\int_{[a,b]} \boldsymbol{\omega}$ could be defined as the integral of any of its coordinate expressions. To prove this we define

$$g = \psi \circ \varphi^{-1} : \varphi\,(U \cap V) \longrightarrow \psi\,(U \cap V).$$

This is a diffeomorphism between open sets in and

$$(k \circ \psi^{-1}) \circ g = k \circ \varphi^{-1}.$$

Furthermore, $g([\alpha_1, \beta_1]) = [\alpha_2, \beta_2]$. Now, on $U \cap V$,

$$h = \boldsymbol{\omega}\left(\frac{\partial}{\partial x}\right) = \boldsymbol{\omega}\left(g'\,\frac{\partial}{\partial y}\right) = g'\,\boldsymbol{\omega}\left(\frac{\partial}{\partial y}\right) = g'\,k$$

so

$$h \circ \varphi^{-1} = g'\left(k \circ \varphi^{-1}\right) = g'\left(k \circ \psi^{-1}\right) \circ g = \left(\left(k \circ \psi^{-1}\right) \circ g\right) g'$$

and the Change of Variables Formula (4.6.1) gives

$$\int_{[\alpha_2,\beta_2]} k \circ \psi^{-1} = \int_{[\alpha_1,\beta_1]} \left(\left(k \circ \psi^{-1}\right) \circ g\right) g' = \int_{[\alpha_1,\beta_1]} h \circ \varphi^{-1}$$

as required.

This we regard as rather persuasive evidence in favor of the view that it is not functions, but 1-forms that one should integrate on 1-manifolds. There is no standard coordinate system on a general 1-manifold and therefore no standard coordinate expression for a function on a 1-manifold and, even in , it matters very much which coordinate expression one chooses to integrate over a given subset. Coordinate expressions for 1-forms, on the other hand, have just what it takes (according to (4.6.1)) to possess invariant integrals. That all of this works out just as nicely for n-forms on n-manifolds will become clear from a glance at Exercise 4.3.4 and the Change of Variables Formula for ^{n} (*sans* a few hypotheses that we will add a bit later): Suppose U and V are open in ^{n} and $g : U \longrightarrow V$ is a diffeomorphism. If $M \subseteq U$ and $f : V \longrightarrow$, then

$$\int_{g(M)} f = \int_M (f \circ g)\left|\det(g')\right|,$$

where g' is the Jacobian of g, i.e., the matrix of g_* relative to standard coordinates.

With this motivation behind us we set about building a theory of integration for n-forms on n-dimensional manifolds. We will presume a familiarity with the basics of Lebesgue integration on ^{n}.

Remark: This decision to build upon the Lebesgue rather than the Riemann integral smooths the theoretical development in the early stages, but has essentially no practical impact on the explicit calculations we must perform. The reader not versed in the Lebesgue theory has a number of options. Perhaps the most sensible is to simply ignore all references to the subject and be assured that, in the end, the integrals we must actually compute are accessible to the tools available in the theory of the Riemann integral. Alternatively, one might actually learn something about this indispensible part of modern mathematics. We will provide a brief synopsis, but everything we need (and more) is covered very nicely and in fewer than fifty pages (specifically, pages 201–247) in [**Brow**].

Throughout the remainder of this section X will denote a smooth, oriented, n-dimensional manifold with $n \geq 1$ and all charts will be assumed consistent with the given orientation.

Recall that a subset A of $\ ^n$ is said to have (**Lebesgue**) **measure zero** in $\ ^n$ if it can be covered by countable families of open rectangles with arbitrarily small total volume. In more detail, an open rectangle in $\ ^n$ is a set of the form $R = (a^1, b^1) \times \cdots \times (a^n, b^n)$, where $a^i < b^i$ for $i = 1, \ldots, n$, and its volume is $\text{vol}(R) = (b^1 - a^1) \cdots (b^n - a^n)$. Then $A \subseteq \ ^n$ has measure zero if, for every $\epsilon > 0$, there exists a family $R_1, R_2, \ldots$ of open rectangles with $A \subseteq \bigcup_{i=1}^{\infty} R_i$ and $\sum_{i=1}^{\infty} \text{vol}(R_i) < \epsilon$. Subsets of sets of measure zero obviously also have measure zero and countable unions of sets of measure zero also have measure zero (cover the k^{th} set with a family of rectangles of total volume less than $\epsilon/2^k$). It is also true, but not at all obvious, that smooth images of sets of measure zero have measure zero, i.e., if $U \subseteq \ ^n$ is open, $f : U \longrightarrow \ ^n$ is smooth and $A \subseteq U$ has measure zero, then $f(A)$ has measure zero (Proposition 5–17 of [**N1**], or Lemma 1.1, Chapter 3, of [**Hir**]).

A subset A of a smooth manifold X is said to have **measure zero** in X if, for every chart (U, φ) for X, the set $\varphi(U \cap A)$ has measure zero in $\ ^n$.

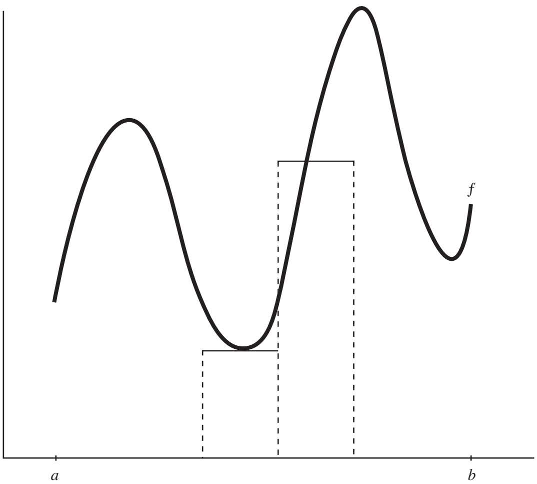
Exercise 4.6.1 Show that A has measure zero in X if and only if, for each $a \in A$, there exists a chart (U, φ) at a for which $\varphi(U \cap A)$ has measure zero in $\ ^n$. **Hint**: Keep in mind that our manifolds are assumed second countable.

If S is any set, then a collection $\mathcal{A}$ of subsets of S is called a σ-algebra if it contains the empty set and is closed under the formation of complements and countable unions, i.e., if (1) $\emptyset \in \mathcal{A}$, (2) $S - A \in \mathcal{A}$ whenever $A \in \mathcal{A}$, and (3) $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ whenever $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots$. Any collection $\mathcal{C}$ of subsets of S is contained in various σ-algebras of subsets of S (e.g., the collection of *all* subsets of S) and the intersection of all of these σ-algebras containing $\mathcal{C}$ is itself a σ-algebra. This intersection is called the σ-algebra generated by $\mathcal{C}$. In any topological space S the σ-algebra generated by the collection $\mathcal{C}$ of closed sets is called the Borel σ-algebra and its elements are called **Borel sets** in S. Closed sets, open sets, countable unions of countable intersections of open sets, etc., are all Borel sets.

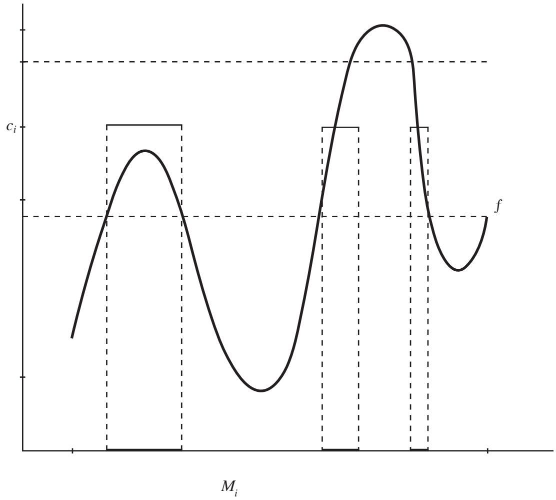
The collection of Borel sets in ^{n} is huge, but we need to enlarge the collection still further. A subset M of ^{n} is said to be (**Lebesgue**) **measurable** if it can be written as $M = A \cup B$, where $A \cap B = \emptyset$, A has measure zero in ^{n} and B is a Borel set in ^{n}. Thus, measurable sets are those which differ from a Borel set by a set of measure zero. Since the empty set has measure zero, every Borel set is measurable. Since the empty set is a Borel set, every set of measure zero is measurable.

Remark: Those of our readers who are "in the know" realize that this is not the usual definition of Lebesgue measurable. That it is equivalent to the usual definition follows from Theorem 9.29 of [**Brow**].

The collection $\mathcal{M}$ of Lebesgue measurable sets is itself a σ-algebra. As the name suggests, the Lebesgue theory assigns to every $M \in \mathcal{M}$ a "measure" $m(M)$ of its size. For example, $m(R) = \text{vol}(R)$ for every rectangle R and $m(A) = 0$ for any set A of measure zero. The ability to measure a large collection of sets provided Lebesgue with the means to integrate a large collection of functions. Roughly, his idea was as follows: Riemann integrated a function $f : [a, b] \longrightarrow$ by partitioning $[a, b]$ into subintervals, approximating f by a step function that was constant on each subinterval, defining the integral of the step function in the only reasonable way and taking the limit as the partition became finer and finer.



This limit will not exist unless f is relatively nice, e.g., a bounded, real-valued function on $[a, b]$ must be continuous except perhaps on a set of measure zero (Theorem 10.23 of [**Brow**]). This is unfortunate since, for example, limits of even very nice functions are often not at all nice so the Riemann integral does not react well to taking limits. Lebesgue's idea was to partition, not the domain, but the range, approximate f by a step function that is constant on the inverse image of each set in the partition, define the integral of the step function by $\sum c_i \; m(M_i)$, where c_i is the value on the inverse image M_i and $m(M_i)$ is its measure and then take a limit as the partition becomes finer and finer.



For the record we will now briefly sketch the construction of the Lebesgue integral in somewhat more detail and formulate those definitions and results from the theory that we will need. A real-valued function f on ^{n} is **measurable** if the inverse image of every closed interval in is Lebesgue measurable in ^{n} (if f is an extended real-valued function, i.e., maps to $\cup \{\pm\infty\}$, then it is also required that $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ be measurable). A (**measurable**) **step function** is a function s that is zero except on a finite number of disjoint measurable sets $M_1, \ldots, M_k$ with $m(M_i) < \infty$ for each $i = 1, \ldots, k$ and which takes a finite constant value c_i on M_i for $i = 1, \ldots, k$. The integral of such a step function s over ^{n} is defined by

$$\int_{\;\,n} s \; dm = \sum_{i=1}^{k} c_i \, m(M_i).$$

One can show that if f is any non-negative measurable function on ^{n}, then there exists a sequence $0 \leq s_1 \leq s_2 \leq \cdots$ of measurable step functions which converges pointwise to f. The **Lebesgue integral of f over** ^{n} is defined by

$$\int_{\;\,n} f \, dm = \sup \int_{\;\,n} s \, dm,$$

where the supremum is over all measurable step functions s with $0 \leq s \leq f$. This integral might well be infinite; if it is finite we say that f is **Lebesgue integrable** (or **summable**) **on** ^{n}. If f is measurable, but not necessarily non-negative, we write $f = f^+ - f^-$, where f^+ and f^- are measurable and non-negative (e.g., $f^+ = \frac{1}{2}(|f| + f)$ and $f^- = \frac{1}{2}(|f| - f)$) and say that f is **Lebesgue integrable on** ^{n} if and only if f^+ and f^- are both integrable. In this case we define

$$\int_{\ ^n} f\,dm = \int_{\ ^n} f^+ dm - \int_{\ ^n} f^- dm.$$

It follows that f is Lebesgue integrable on ^{n} if and only if $|f|$ is Lebesgue integrable on ^{n}. Finally, suppose that $M \subseteq\ ^n$ is measurable. Then its characteristic function χ_M (1 on M and 0 on $^n - M$) is a measurable function. If $\chi_M f$ is measurable (as it will be whenever f is measurable), then we define

$$\int_M f\,dm = \int_{\ ^n} \chi_M f\,dm.$$

If this is finite (as it will be if f is integrable on ^{n}), then we say that f is **Lebesgue integrable on** M.

The Lebesgue integral has all of the properties one would expect of an integral. The set of integrable functions is a linear space and the integral is a linear functional on it:

$$\int_M (af + bg)dm = a\int_M f\,dm + b\int_M g\,dm.$$

Sets of measure zero "don't count," i.e., if f is integrable and g agrees with f except perhaps on a set of measure zero (we say that g equals f **almost everywhere** and write $g = f$ a.e.), then g is integrable and

$$\int_M g\,dm = \int_M f\,dm \qquad (g = f \text{ a.e.}).$$

It follows that if M' differs from M by a set of measure zero and f is integrable on M, then f is integrable on M' and

$$\int_{M'} f\,dm = \int_M f\,dm.$$

Similarly, if M and N are measurable and $M \cap N$ is a set of measure zero, then

$$\int_{M \cup N} f\,dm = \int_M f\,dm + \int_N f\,dm \qquad (m(M \cap N) = 0).$$

The Change of Variables Formula takes exactly the same form as it does for the Riemann integral: Let U and V be open sets in ^{n}, $g : U \longrightarrow V$ a diffeomorphism, $M \subseteq U$ measurable and $f : V \longrightarrow$ Lebesgue integrable. Then $g(M)$ is measurable, $(f \circ g)|\det(g')|$ is integrable on M and

$$\int_{g(M)} f\, dm = \int_M (f \circ g)|\det(g')| dm \tag{4.6.3}$$

(Theorem 10.46 of [**Brow**]). A very useful result for the purposes of calculation is that if f is continuous on a closed rectangle $\bar{R} = [a^1, b^1] \times \cdots \times [a^n, b^n]$, then it is Lebesgue integrable on $\bar{R}$ and $\int_{\bar{R}} f\, dm$ agrees with the ordinary Riemann integral of f over $\bar{R}$. Consequently, the value of the integral can be calculated as an iterated integral (Fubini's Theorem) from the Fundamental Theorem of Calculus:

$$\int_{\bar{R}} f\, dm = \int_{a^1}^{b^1} \left(\int_{a^2}^{b^2} \left(\cdots \int_{a^n}^{b^n} f(x^1, x^2, \ldots, x^n)\, dx^n \cdots \right) dx^2 \right) dx^1.$$

Remark: Fubini's Theorem becomes trickier when the continuity assumption is weakened (see Sections 10.8 and 10.9 of [**Brow**]).

The most important properties of the Lebesgue integral, however, are those related to limits. The so-called *Monotone Convergence Theorem* (Theorem 10.15 of [**Brow**]) asserts that if $0 \leq f_1 \leq f_2 \leq \cdots$ is a sequence of measurable functions that converges pointwise (except perhaps on a set of measure zero) to f, then f is measurable and

$$\lim_{k \longrightarrow \infty} \int_{\ ^n} f_k\, dm = \int_{\ ^n} \lim_{k \longrightarrow \infty} f_k\, dm = \int_{\ ^n} f\, dm. \tag{4.6.4}$$

The *Lebesgue Dominated Convergence Theorem* (Theorem 10.15 of [**Brow**]) asserts that if $\{f_k\}$ is a sequence of measurable functions that converges pointwise (almost everywhere) to f and $|f_k| \leq g$ for all k, where g is integrable, then f is integrable and its integral is given by (4.6.4).

With this detour into Lebesgue integration behind us we are prepared to build an integration theory on any smooth, oriented, n-dimensional manifold X. We have already transferred the notion of measure zero to X (Exercise 4.6.1). Although there is no reasonable way to move the Lebesgue measure itself to X we can easily transfer the notion of measurability. Indeed, X is a topological space and so has a σ-algebra of Borel sets so we may say that a subset M of X is **measurable** if it can be written as $M = A \cup B$, where $A \cap B = \emptyset$, A has measure zero in X and B is a Borel set in X.

Exercise 4.6.2 Show that $M \subseteq X$ is measurable if and only if, for each $m \in M$, there exists a chart (U, φ) at m for which $\varphi(U \cap M)$ is Lebesgue measurable in ^{n}. **Hint**: Use Exercise 4.6.1 and the fact that diffeomorphisms on open sets in ^{n} preserve measure zero, Borel sets and disjoint unions.

Now we consider an n-form $\boldsymbol{\omega}$ on X (not necessarily smooth). We will say that $\boldsymbol{\omega}$ is **measurable** if, for every chart (U, φ) on X with coordinate functions $x^1, \ldots, x^n$ and $\boldsymbol{\omega} = h\, dx^1 \wedge \cdots \wedge dx^n$ on U, the coordinate expression $h \circ \varphi^{-1}$ is Lebesgue measurable on $\varphi(U)$. We show now that it is enough to check this condition for all charts in some atlas for X. Thus, we suppose that $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in \mathcal{A}}$ is an atlas and that our condition has been verified for every chart in this atlas. Let (V, ψ) be some other chart for X with coordinate functions $y^1, \ldots, y^n$ and with $\boldsymbol{\omega} = k\, dy^1 \wedge \cdots \wedge dy^n$ on V. We must show that $k \circ \psi^{-1}$ is Lebesgue measurable on $\psi(V)$.

Exercise 4.6.3 Show that it will suffice to prove that $k \circ \psi^{-1}$ is Lebesgue measurable on some neighborhood of each point of $\psi(V)$.

Any point in $\psi(V)$ is also in $\varphi_\alpha(U_\alpha)$ for some $\alpha \in \mathcal{A}$. If we let $g = \varphi_\alpha \circ \psi^{-1} : \psi(U_\alpha \cap V) \longrightarrow \varphi_\alpha(U_\alpha \cap V)$, then g is a diffeomorphism between open sets in ^{n}. If $x^1, \ldots, x^n$ denote the coordinate functions for $(U_\alpha, \varphi_\alpha)$ and we write $\boldsymbol{\omega} = h_\alpha\, dx^1 \wedge \cdots \wedge dx^n$ on U_α, then Exercise 4.3.4 gives, for every $p \in U_\alpha \cap V$,

$$k(p) = h_\alpha(p) \det\left(\frac{\partial x^i}{\partial y^j}(p)\right) = h_\alpha(p) \det\big(D_j(x^i \circ \psi^{-1})(\psi(p))\big).$$

Writing $p = \psi^{-1}(y)$ for $y \in \psi(U_\alpha \cap V)$ this becomes

$$\begin{aligned} k \circ \psi^{-1}(y) &= \big(h_\alpha \circ \psi^{-1}\big)(y) \det\big(D_j(x^i \circ \psi^{-1})(y)\big) \\ &= \big((h_\alpha \circ \varphi_\alpha^{-1}) \circ g\big)(y) \det\big(D_j g^i(y)\big) \end{aligned}$$

so

$$k \circ \psi^{-1} = \big((h_\alpha \circ \varphi_\alpha^{-1}) \circ g\big) \det(g') \tag{4.6.5}$$

on $\psi(U_\alpha \cap V)$. Now, on open sets in ^{n}, smooth maps are measurable and compositions and products of measurable functions are measurable. Since we have assumed that $h_\alpha \circ \varphi_\alpha^{-1}$ is measurable it follows that $k \circ \psi^{-1}$ is also measurable.

If $\boldsymbol{\omega}$ is any n-form on X we define its **support** $\operatorname{supp} \boldsymbol{\omega}$ to be the closure in X of the set of points where $\boldsymbol{\omega}$ is not the zero form:

$$\operatorname{supp} \boldsymbol{\omega} = \overline{\{x \in X : \boldsymbol{\omega}_x \neq 0\}}.$$

We define the integral of $\boldsymbol{\omega}$ over X first in the case in which $\operatorname{supp} \boldsymbol{\omega}$ is compact and contained in some coordinate neighborhood and then, in the general case, use a partition of unity to define the integral as a sum of integrals of the first type.

Suppose then that $\boldsymbol{\omega}$ is a measurable n-form on X with compact support $\operatorname{supp} \boldsymbol{\omega} \subseteq U$, where (U, φ) is a chart on X (keep in mind that we consider only charts consistent with the given orientation of X). If $x^1, \ldots, x^n$ are the coordinate functions of φ and $\boldsymbol{\omega} = h\, dx^1 \wedge \cdots \wedge dx^n$ on U, then the coordinate expression $h \circ \varphi^{-1}$ is measurable on the open set $\varphi(U)$ in ^{n}. We say that

$\boldsymbol{\omega}$ is **integrable on** U (and also on X) if $h \circ \varphi^{-1}$ is integrable on $\varphi(U)$ (and therefore on $\;^n$) and, in this case, we define

$$\int_X \boldsymbol{\omega} = \int_U \boldsymbol{\omega} = \int_{\;^n} h \circ \varphi^{-1}\, dm = \int_{\varphi(U)} h \ o\ \varphi^{-1}\, dm. \tag{4.6.6}$$

Remark: Integrability will be assured, for example, if $\boldsymbol{\omega}$ is continuous with compact support $\operatorname{supp} \boldsymbol{\omega} \subseteq U$.

Of course, we must prove that our definition does not depend on the choice of the chart (U, φ) with $\operatorname{supp} \boldsymbol{\omega} \subseteq U$. Thus we let (V, ψ) be another (oriented) chart on X with $\operatorname{supp} \boldsymbol{\omega} \subseteq V$. Let $y^1, \ldots, y^n$ be the coordinate functions for ψ, and write $\boldsymbol{\omega} = k\ dy^1 \wedge \cdots \wedge dy^n$ on V. We must show that

$$\int_{\varphi(U)} h \circ \varphi^{-1}\, dm = \int_{\psi(V)} k \circ \psi^{-1}\, dm.$$

Since $\operatorname{supp} \boldsymbol{\omega} \subseteq U \cap V,\ h \circ \varphi^{-1}$ is zero outside of $\varphi(U \cap V)$ and $k \circ \psi^{-1}$ is zero outside of $\psi(U \cap V)$ and we need only show that

$$\int_{\varphi(U\cap V)} h \circ \varphi^{-1} dm = \int_{\psi(U\cap V)} k \circ \psi^{-1} dm.$$

Let $g = \varphi \circ \psi^{-1} : \psi(U \cap V) \longrightarrow \varphi(U \cap V)$. This is a diffeomorphism between open sets in $\;^n$ and (4.6.5), without the subscript α, gives

$$k \circ \psi^{-1} = ((h \circ \varphi^{-1}) \circ g) \det(g').$$

Now, since both (U, φ) and (V, ψ) are consistent with the orientation of X, the Jacobian determinant $\det(g')$ is positive so we may, if we wish, write this as

$$k \circ \psi^{-1} = ((h \circ \varphi^{-1}) \circ g) \Big| \det(g') \Big|.$$

But now the Change of Variables Formula (4.6.3) with $f = h \circ \varphi^{-1}$ and $M = \psi(U \cap V)$ assures the integrability of $k \circ \psi^{-1}$ and gives

$$\int_{\varphi(U\cap V)} h \circ \varphi^{-1} dm = \int_{\psi(U\cap V)} ((h \circ \varphi^{-1}) \circ g) \Big| \det(g') \Big| dm$$

$$= \int_{\psi(U\cap V)} k \circ \psi^{-1} dm$$

as required.

Before proceeding with the general case we make a few observations about this one. First notice that if $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are two n-forms of the prescribed type with supports both contained in U, then $\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ is also of this type and

$$\int_X \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 = \int_X \boldsymbol{\omega}_1 + \int_X \boldsymbol{\omega}_2.$$

Furthermore, if $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ agree except perhaps at a set of points of measure zero in X, then

$$\int_X \boldsymbol{\omega}_1 = \int_X \boldsymbol{\omega}_2 \qquad (\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2 \text{ a.e.})$$

Next suppose that $\boldsymbol{\omega}$ is an n-form of the type we are considering and let $\{\phi_k\}_{k=1,2,\ldots}$ be a partition of unity subordinate to an oriented atlas for X. Since $\operatorname{supp}\boldsymbol{\omega}$ is compact and $\{\operatorname{supp}\phi_k\}_{k=1,2,\ldots}$ is locally finite there are only finitely many values of k for which $\phi_k\boldsymbol{\omega}$ is *not* identically zero. Thus, $\sum_{k=1}^{\infty} \phi_k\,\boldsymbol{\omega}$ is actually a finite sum and

$$\sum_{k=1}^{\infty} \phi_k\,\boldsymbol{\omega} \;=\left(\sum_{k=1}^{\infty} \phi_k\right)\boldsymbol{\omega} = 1\boldsymbol{\omega} \;= \boldsymbol{\omega}.$$

Each $\phi_k\,\boldsymbol{\omega}$ is an n-form with compact support contained in a coordinate neighborhood. If $\boldsymbol{\omega}$ is integrable, then each $\phi_k\,\boldsymbol{\omega}$ is integrable and linearity implies

$$\int_X \boldsymbol{\omega} = \sum_{k=1}^{\infty} \int_X \phi_k\,\boldsymbol{\omega}. \tag{4.6.7}$$

Now let us turn to the general case and consider a measurable n-form $\boldsymbol{\omega}$ on X($\operatorname{supp}\boldsymbol{\omega}$ need not be compact and need not be contained in a coordinate neighborhood). Choose an oriented atlas for X and a partition of unity $\{\phi_k\}_{k=1,2,\ldots}$ subordinate to it. Then each $\phi_k\,\boldsymbol{\omega}$ is an n-form with compact support contained in a coordinate neighborhood. Since $\{\operatorname{supp}\phi_k\}_{k=1,2,\ldots}$ is locally finite the sum $\sum_{k=1}^{\infty} \phi_k\boldsymbol{\omega}$ is finite on some neighborhood of each point so

$$\boldsymbol{\omega} \;= \sum_{k=1}^{\infty} \phi_k\,\boldsymbol{\omega}$$

on X. We will say that $\boldsymbol{\omega}$ is **integrable on** X if each $\phi_k\,\boldsymbol{\omega}$ is integrable on X and if the series

$$\sum_{k=1}^{\infty} \int_X \phi_k\,\boldsymbol{\omega}$$

converges absolutely. In this case we define

$$\int_X \boldsymbol{\omega} = \sum_{k=1}^{\infty} \int_X \phi_k\,\boldsymbol{\omega}.$$

Note that if $\boldsymbol{\omega}$ happens to have compact support contained in some coordinate neighborhood, then we proved in (4.6.7) that this new definition agrees with the old one. To show that the definition does not depend on the chosen atlas or partition of unity we suppose $\{\phi'_j\}_{j=1,2,\ldots}$ is a partition of unity subordinate to some other oriented atlas for X. Then $\boldsymbol{\omega} = \sum_{j=1}^{\infty} \phi'_j\boldsymbol{\omega}$ on X (the sum is finite on some neighborhood of each point). Each of the n-forms

$\phi'_j\boldsymbol{\omega}$ has compact support contained in a coordinate neighborhood so we may apply (4.6.7) to it

$$\int_X \phi'_j\,\boldsymbol{\omega} = \sum_{k=1}^{\infty}\int_X \phi_k\,\phi'_j\,\boldsymbol{\omega}.$$

Similarly,

$$\int_X \phi_k\,\boldsymbol{\omega} = \sum_{j=1}^{\infty}\int_X \phi'_j\,\phi_k\,\boldsymbol{\omega}.$$

Thus,

$$\sum_{j=1}^{\infty}\int_X \phi'_j\,\boldsymbol{\omega} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\int_X \phi_k\,\phi'_j\,\boldsymbol{\omega}$$

and

$$\sum_{k=1}^{\infty}\int_X \phi_k\,\boldsymbol{\omega} = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\int_X \phi'_j\,\phi_k\,\boldsymbol{\omega}. \tag{4.6.8}$$

Exercise 4.6.4 Show that absolute convergence of $\sum_{k=1}^{\infty}\int_X \phi_k\,\boldsymbol{\omega}$ implies the convergence of

$$\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|\int_X \phi'_j\,\phi_k\,\boldsymbol{\omega}\right|$$

and use this absolute convergence to interchange the order of summation in (4.6.8) and obtain

$$\sum_{k=1}^{\infty}\int_X \phi_k\,\boldsymbol{\omega} \; = \sum_{j=1}^{\infty}\int_X \phi'_j\boldsymbol{\omega}.$$

Finally, conclude that $\sum_{j=1}^{\infty}\int_X \phi'_j\boldsymbol{\omega}$ converges absolutely.

Remark: If $\boldsymbol{\omega}$ has compact support (not necessarily contained in a coordinate neighborhood), then all of the sums dealt with above are finite so absolute convergence is assured. In particular, this is always the case when X is compact.

To complete our sequence of definitions we let $\boldsymbol{\omega}$ be a measurable n-form on X and $M \subseteq X$ a measurable set. We say that $\boldsymbol{\omega}$ is **integrable on** M if $\chi_M\,\boldsymbol{\omega}$ is integrable on X and in this case, we define

$$\int_M \boldsymbol{\omega} = \int_X \chi_M\,\boldsymbol{\omega}.$$

Exercise 4.6.5 Establish each of the following properties.

(a) If $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are integrable on X, then $\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ is integrable on X and

$$\int_X \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 = \int_X \boldsymbol{\omega}_1 + \int_X \boldsymbol{\omega}_2.$$

(b) If $\boldsymbol{\omega}$ is integrable on X and $a \in$, then $a\boldsymbol{\omega}$ is integrable on X and

$$\int_X a\,\boldsymbol{\omega} = a\int_X \boldsymbol{\omega}.$$

(c) If $\boldsymbol{\omega}_1$ is integrable on X, and $\boldsymbol{\omega}_2 = \boldsymbol{\omega}_1$ a.e., then $\boldsymbol{\omega}_2$ is integrable on X and

$$\int_X \boldsymbol{\omega}_2 = \int_X \boldsymbol{\omega}_1.$$

(d) If $\boldsymbol{\omega}$ is integrable on M and M' differs from M by a set of measure zero, then $\boldsymbol{\omega}$ is integrable on M' and

$$\int_{M'} \boldsymbol{\omega} = \int_M \boldsymbol{\omega}.$$

(e) If $\boldsymbol{\omega}$ is integrable on M and $M = M_1 \cup M_2$, where M_1 and M_2 are measurable and $M_1 \cap M_2$ has measure zero, then $\boldsymbol{\omega}$ is integrable on M_1 and M_2 and

$$\int_M \boldsymbol{\omega} = \int_{M_1} \boldsymbol{\omega} + \int_{M_2} \boldsymbol{\omega}.$$

(f) If $\boldsymbol{\omega}$ is integrable on X and U is an open submanifold X with inclusion $\iota : U \hookrightarrow X$, then

$$\int_U \boldsymbol{\omega} = \int_U \iota^*\boldsymbol{\omega}.$$

Note: $\int_U \boldsymbol{\omega} = \int_X \chi_U\,\boldsymbol{\omega}$, but, for $\int_U \iota^*\boldsymbol{\omega}$, U is regarded as a manifold in its own right.

One last property of integrals that is particularly useful in calculations is contained in the next result.

Theorem 4.6.1 *Let X and Y be two smooth, oriented, n-dimensional manifolds and $f : X \longrightarrow Y$ an orientation preserving diffeomorphism of X onto Y. Let $\boldsymbol{\omega}$ be an integrable n-form on Y. Then $f^*\boldsymbol{\omega}$ is an integrable n-form on X and*

$$\int_X f^*\boldsymbol{\omega} = \int_Y \boldsymbol{\omega}.$$

Proof: Let $\{(U_\alpha, \varphi_\alpha)\}_{\alpha\in\mathcal{A}}$ be an oriented atlas for Y. Since f is an orientation preserving diffeomorphism $\{(f^{-1}(U_\alpha), \varphi_\alpha \circ f)\}_{\alpha\in\mathcal{A}}$ is an oriented atlas for X. Moreover, if $\{\phi_k\}_{k=1,2,\ldots}$ is a partition of unity subordinate to $\{U_\alpha\}_{\alpha\in\mathcal{A}}$, then $\{\phi_k \circ f\}_{k=1,2,\ldots}$ is a partition of unity subordinate to $\{f^{-1}(U_\alpha)\}_{\alpha\in\mathcal{A}}$. By assumption,

$$\int_Y \boldsymbol{\omega} = \sum_{k=1}^{\infty} \int_Y \phi_k\,\boldsymbol{\omega},$$

where the convergence on the right-hand side is absolute. We must show that

$$\sum_{k=1}^{\infty} \int_X (\phi_k \circ f) f^* \boldsymbol{\omega} = \sum_{k=1}^{\infty} \int_Y \phi_k \, \boldsymbol{\omega}$$

and that the convergence on the left-hand side is absolute. Both of these will be proved if we show that

$$\int_X (\phi_k \circ f) f^* \boldsymbol{\omega} \; = \int_Y \phi_k \, \boldsymbol{\omega}$$

for each k. Notice that $(\phi_k \circ f) f^* \boldsymbol{\omega} = f^*(\phi_k \, \boldsymbol{\omega})$ so we must prove that

$$\int_X f^* \; (\phi_k \, \boldsymbol{\omega}) = \int_Y \phi_k \, \boldsymbol{\omega}.$$

Now, $\phi_k \, \boldsymbol{\omega}$ has compact support contained in some chart $(U_\alpha, \varphi_\alpha)$ for Y so if we let $x^1, \ldots, x^n$ be the coordinate functions of φ_α and write $\phi_k \, \boldsymbol{\omega} = h \, dx^1 \wedge \cdots \wedge dx^n$ on U_α we have

$$\int_Y \phi_k \, \boldsymbol{\omega} = \int_{U_\alpha} \phi_k \, \boldsymbol{\omega} \; = \int_{\varphi_\alpha(U_\alpha)} h \circ \varphi_\alpha^{-1} \, dm.$$

Moreover, on $f^{-1}(U_\alpha)$,

$$f^*(\phi_k \, \boldsymbol{\omega}) = f^*(h dx^1 \wedge \cdots \wedge dx^n) = (h \circ f) dy^1 \wedge \cdots \wedge dy^n,$$

where $y^i = x^i \circ f$, $i = 1, \ldots, n$, are local coordinates on $f^{-1}(U_\alpha)$. Since $f^*(\phi_k \, \boldsymbol{\omega})$ has compact support contained in $f^{-1}(U_\alpha)$,

$$\begin{aligned}
\int_X f^*(\phi_k \, \boldsymbol{\omega}) &= \int_{f^{-1}(U_\alpha)} f^*(\phi_k \, \boldsymbol{\omega}) \\
&= \int_{(\varphi_\alpha \circ f)(f^{-1}(U_\alpha))} (h \circ f) \circ (\varphi_\alpha \circ f)^{-1} \, dm \\
&= \int_{\varphi_\alpha(U_\alpha)} h \circ f \circ f^{-1} \circ \varphi_\alpha^{-1} \, dm = \int_{\varphi_\alpha(U_\alpha)} h \circ \varphi_\alpha^{-1} \, dm \\
&= \int_Y \phi_k \, \boldsymbol{\omega}
\end{aligned}$$

as required. ∎

Exercise 4.6.6 Show that, if $f : X \longrightarrow Y$ is an orientation reversing diffeomorphism, then

$$\int_X f^* \boldsymbol{\omega} \; = - \int_Y \boldsymbol{\omega}.$$

It's time now to actually compute a few integrals. We recall (Theorem 4.3.2) that any oriented Riemannian n-manifold X has defined on it a unique metric volume form $\boldsymbol{\omega}$ with the property that $\boldsymbol{\omega}_x(e_1, \ldots, e_n) = 1$ whenever $\{e_1, \ldots, e_n\}$ is an oriented, orthonormal basis for $T_x(X)$. This n-form is smooth and, if X is compact, it has compact support and therefore is integrable over X.

Exercise 4.6.7 Show that, if $\boldsymbol{\omega}$ is the metric volume form on a compact, oriented, Riemannian manifold X, then $\int_X \boldsymbol{\omega} > 0$.

The value of the integral in Exercise 4.6.7 is called the **volume** of X. We will compute the volumes of a few spheres. The volume form on S^n is $\iota^* \tilde{\boldsymbol{\omega}}$, where $\tilde{\boldsymbol{\omega}}$ is given by (4.3.2) and $\iota : S^n \hookrightarrow {}^{n+1}$ is the inclusion. We begin with S^1.

Remark: Scrupulous honesty in calculations such as these can result in an avalanche of unnecessary notation. On the other hand, too cavalier an attitude can leave one wondering where all of the machinery went with which we built the integral. We will attempt to follow an intermediate path, perhaps a bit too pedantic early on, but gaining bravado as we proceed.

The volume (length) of S^1 is (by (4.3.3)) given by

$$\int_{S^1} \boldsymbol{\omega} = \int_{S^1} \iota^* \tilde{\boldsymbol{\omega}} = \int_{S^1} \iota^*(x^1 dx^2 - x^2 dx^1).$$

Since S^1 and $S^1 - \{(-1,0)\}$ differ by a set of measure zero,

$$\int_{S^1} \boldsymbol{\omega} = \int_{S^1 - \{(-1,0)\}} \iota^*(x^1 dx^2 - x^2 dx^1).$$

Remark: Technically, the integral on the right is

$$\int_{S^1} \chi_{S^1 - \{(-1,0)\}} \iota^*(x^1 dx^2 - x^2 dx^1),$$

but we intend to appeal immediately to Exercise 4.6.5 (f) and regard it as an integral over the open submanifold $S^1 - \{(-1,0)\}$ of S^1. We should then add another inclusion map $S^1 - \{(-1,0)\} \hookrightarrow S^1$, but we will permit ourselves sufficient bravado even now to ignore this one. If this bothers you, by all means put it in.

Now, define a map $f : (-\pi, \pi) \longrightarrow S^1 - \{(-1,0)\}$ by

$$(\iota \circ f)(\theta) = (\cos\theta, \sin\theta),$$

where θ is the standard coordinate on $(-\pi, \pi) \subseteq$.

Exercise 4.6.8 Show that f is an orientation preserving diffeomorphism. **Hint**: Show first that it is a diffeomorphism. Then "orientation preserving" need only be checked at one point. Now show that

$$(\iota \circ f)_{*0}\left(\left.\frac{d}{d\theta}\right|_0\right) = \left.\frac{\partial}{\partial x^2}\right|_{(1,0)}.$$

According to Theorem 4.6.1 we therefore have

$$\begin{aligned}\int_{S^1} \boldsymbol{\omega} &= \int_{(-\pi,\pi)} f^* \Big(\iota^* (x^1 dx^2 - x^2 dx^1) \Big) \\ &= \int_{(-\pi,\pi)} (\iota \circ f)^* (x^1 dx^2 - x^2 dx^1).\end{aligned}$$

Now,

$$\begin{aligned}(\iota \circ f)^* (x^1 dx^2 - x^2 dx^1) &= (x^1 \circ \iota \circ f)\; d(x^2 \circ \iota \circ f) \\ &\quad - \;(x^2 \circ \iota \circ f)\; d(x^1 \circ \iota \circ f) \\ &= \cos\theta \, d(\sin\theta) - \sin\theta \, d(\cos\theta) \\ &= \cos^2\theta \, d\theta + \sin^2\theta \, d\theta \\ &= d\theta\end{aligned}$$

so

$$\int_{S^1} \boldsymbol{\omega} \;= \int_{(-\pi,\pi)} d\theta.$$

Finally, $d\theta = 1\, d\theta$ is the standard coordinate expression for a continuous 1-form on and $(-\pi,\pi)$ differs from $[-\pi,\pi]$ by a set of measure zero so

$$\int_{S^1} \boldsymbol{\omega} = \int_{[-\pi,\pi]} 1\, d\theta = \int_{[-\pi,\pi]} 1\, dm = \theta|_{-\pi}^{\pi} = 2\pi.$$

Exercise 4.6.9 Define $h : \;\; \longrightarrow S^1$ by $(\iota \circ h)(\theta) = (\cos\theta, \sin\theta)$. Show that if $\boldsymbol{\omega}$ is any continuous 1-form on S^1, then

$$\int_{S^1} \boldsymbol{\omega} = \int_{[\alpha,\alpha+2\pi]} h^* \boldsymbol{\omega}$$

for any real number α.

The calculations for S^2 are analogous. By (4.3.4), the volume (area) of S^2 is given by

$$\int_{S^2} \boldsymbol{\omega} = \int_{S^2} \iota^* (x^1 dx^2 \wedge dx^3 - x^2 dx^1 \wedge dx^3 + x^3 dx^1 \wedge dx^2).$$

Define a map f from $(0,\ \pi) \times (-\pi,\pi)$ into S^2 by

$$(\iota \circ f)\ (\phi,\theta) = (\sin\phi\ \cos\theta,\ \sin\phi\ \sin\theta,\ \cos\phi).$$

The image of f is the open submanifold of S^2 consisting of the complement of the semicircle $\{(-\sin\phi,\ 0,\ \cos\phi) : 0 \leq \phi \leq \pi\}$ in S^2. This semicircle has measure zero so we may integrate $\boldsymbol{\omega}$ over its complement to get the volume.

Exercise 4.6.10 Show that f is an orientation preserving diffeomorphism of $(0,\pi) \times (-\pi,\pi)$ onto its image in S^2.

Thus,

$$\int_{S^2} \boldsymbol{\omega} \;=\; \int_{(0,\pi)\times(-\pi,\pi)} (\iota \circ f)^* \;\big(x^1 dx^2 \wedge dx^3 \\ - x^2 dx^1 \wedge dx^3 + x^3 dx^1 \wedge dx^2\big).$$

Now,

$$\begin{aligned}(\iota \circ f)^* \;(x^1 dx^2 \wedge dx^3) &= \sin\phi \;\cos\theta \;(d(\sin\phi\,\sin\theta) \wedge d(\cos\phi))\\ &= \sin\phi\;\cos\theta\;((\cos\phi\,\sin\theta\,d\phi + \sin\phi\,\cos\theta\,d\theta)\\ &\qquad \wedge(-\sin\phi\,d\theta))\\ &= \sin^3\phi\,\cos^2\theta\,d\phi\wedge d\theta.\end{aligned}$$

Exercise 4.6.11 Compute the remaining pullbacks and conclude that

$$\int_{S^2} \boldsymbol{\omega} = \int_{(0,\pi)\times(-\pi,\pi)} \sin\phi\; d\phi \wedge d\theta.$$

Now, $\sin\phi\; d\phi\wedge d\theta$ is the standard coordinate expression for a continuous 2-form on ^{2} and $(0,\pi)\times(-\pi,\pi)$ differs from $[0,\pi]\times[-\pi,\pi]$ by a set of measure zero so

$$\int_{S^2} \boldsymbol{\omega} = \int_{[0,\pi]\times[-\pi,\pi]} \sin\phi\,dm = \int_{-\pi}^{\pi}\int_0^{\pi} \sin\phi\,d\phi\,d\theta = 4\pi.$$

Exercise 4.6.12 Show that the volume of S^3 is $2\pi^2$. **Hint**: The appropriate orientation preserving diffeomorphism this time is given by

$$(\iota\circ f)^*(\xi,\phi,\theta) = (\sin\xi\sin\phi\,\cos\theta,\;\sin\xi\,\sin\phi\,\sin\theta,\;\sin\xi\,\cos\phi,\;\cos\xi),$$

where $(\xi,\phi,\theta)\in(0,\pi)\times(0,\pi)\times(-\pi,\pi)$.

Exercise 4.6.13 Let $\boldsymbol{\omega}$ be the closed, nonexact 1-form on ${}^2 - \{(0,0)\}$ constructed from two polar coordinate functions θ_1 and θ_2 in Section 4.4 (page 252). Show that

$$\int_{S^1} \boldsymbol{\omega} = 2\pi.$$

Note: It will follow from Stokes' Theorem (Section 4.7) that the integral of $\boldsymbol{\omega}$ over a circle in ${}^2 - \{(0,0)\}$ is 2π if the circle encloses the origin and 0 if it does not.

We conclude this section with a calculation that yields a special case of Stokes' Theorem.

Theorem 4.6.2 *Let X be a smooth, oriented, n-dimensional manifold and suppose $\boldsymbol{\omega}\in\Lambda^{n-1}(X)$ is a smooth $(n-1)$-form on X with compact support. Then*

$$\int_X d\boldsymbol{\omega} \;= 0.$$

Proof: Since $d\boldsymbol{\omega}$ is smooth and has compact support (because $\boldsymbol{\omega}$ does), it is integrable on X.

Exercise 4.6.14 Show that it is enough to prove the result for smooth $(n-1)$-forms with compact support contained in some coordinate neighborhood U of X.

Thus, suppose (U, φ) is a chart with $\operatorname{supp} \boldsymbol{\omega} \subseteq U$ and let $x^1, \ldots, x^n$ be its coordinate functions. By composing with a diffeomorphism of $\;^n$ onto the open unit ball in $\;^n$, if necessary, we may assume $\varphi(U)$ is a bounded set in $\;^n$. Write $\boldsymbol{\omega} = \sum_{i=1}^n \omega_i \, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n$ on U. Then

$$d\boldsymbol{\omega} = \left(\sum_{i=1}^n (-1)^{i-1} \frac{\partial}{\partial x^i} (\omega_i \circ \varphi^{-1}) \circ \varphi \right) dx^1 \wedge \cdots \wedge dx^n$$

on U so

$$\begin{aligned} \int_X d\boldsymbol{\omega} &= \int_{\varphi(U)} \left(\sum_{i=1}^n (-1)^{i-1} \frac{\partial(\omega_i \circ \varphi^{-1})}{\partial x^i} \right) dm \\ &= \sum_{i=1}^n (-1)^{i-1} \int_{\varphi(U)} \frac{\partial(\omega_i \circ \varphi^{-1})}{\partial x^i} dm. \end{aligned}$$

Take each $\omega_i \circ \varphi^{-1}$ to be zero outside $\varphi(U)$ and choose a cube $C = \{(x^1, \ldots, x^n) \in \;^n : |x^i| \leq a, \; i = 1, \ldots, n\}$ containing $\varphi(U)$. Then

$$\int_X d\boldsymbol{\omega} = \sum_{i=1}^n (-1)^{i-1} \int_C \frac{\partial(\omega_i \circ \varphi^{-1})}{\partial x^i} dm.$$

But each $\int_C \frac{\partial(\omega_i \circ \varphi^{-1})}{\partial x^i} \, dm$ is zero because, by Fubini's Theorem, it can be computed as an iterated integral, integrating first with respect to x^i and $\omega_i \circ \varphi^{-1}$ vanishes on the boundary of the cube. ∎

4.7 Stokes' Theorem

In this final section we prove a far-reaching and quite beautiful generalization of the three classical integral theorems of vector calculus (Green's Theorem, Stokes' Theorem and The Divergence Theorem). Each of these asserts the equality of two integrals, one of them over an n-dimensional region D ($n = 2, 3$) and the other over its $(n-1)$-dimensional boundary ∂D. The Divergence Theorem, for example, states that the outward flux of a smooth vector field $\vec{F}$ on $\;^3$ through a closed surface ∂D equals the integral of the vector field's divergence $\operatorname{div} \vec{F}$ over the region D that the surface bounds:

$$\iint\limits_{\partial D} \vec{F} \cdot \vec{N} \, dS = \iiint\limits_{D} \operatorname{div} \vec{F} \, dV.$$

With a glace back at Exercise 4.4.8 it may come as no surprise that these theorems are really statements about $(n-1)$-forms and their exterior derivatives. Remarkably, these are all, in a sense, the same statement and it is our objective here to derive the one simple, elegant equality from which they all follow:

$$\int_{\partial D} \iota^* \boldsymbol{\omega} = \int_D d\boldsymbol{\omega}. \tag{4.7.1}$$

The first order of business is to define precisely the appropriate regions of integration.

Throughout this section X will denote a smooth, n-dimensional, oriented manifold. A subset D of X will be called a **domain with smooth boundary** in X if, for each $p \in X$, one of the following is true:

(1) There is an open neighborhood of p in X which is contained entirely in $X - D$ (such points are said to be in the **exterior** of D).

(2) There is an open neighborhood of p in X which is contained entirely in D (these are called the **interior** points of D).

(3) There exists a chart (U, φ) for X at p with $\varphi(p) = 0 \in \mathbb{R}^n$ and $\varphi(U \cap D) = \varphi(U) \cap \mathbb{R}^n_+$, where $\mathbb{R}^n_+ = \{(x^1, \ldots, x^n) \in \mathbb{R}^n : x^n \geq 0\}$ is the closed upper half-space in $\mathbb{R}^n$ (these are called the **boundary points** of D and the set of all such is denoted ∂D and called the **boundary** of D).

Remark: These definitions make sense for any $n \geq 1$, but, when $n = 1$, ∂D is a set of isolated points. Orientations for 0-dimensional manifolds and integrals over such can all be defined in such a way that the results of this section remain valid in this case, but we will have no need of them and so will assume henceforth that $n \geq 2$. Notice also that the set Int D of all interior points of D is open in X and that there is nothing in the definition to preclude the possibility that $\partial D = \emptyset$ (so X itself qualifies as a domain with smooth boundary).

Exercise 4.7.1 Show that these three conditions are mutually exclusive and that, for any chart (U, φ) of the type described in (3), $\varphi(U \cap \partial D) = \{x \in \varphi\ (U) : x^n = 0\}$. Conclude (e.g., from Exercise 5.6.1, [**N4**]) that ∂D is an $(n-1)$-dimensional submanifold of X.

Exercise 4.7.2 Show that the n-dimensional disc $D^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is a domain with smooth boundary in $\mathbb{R}^n$ and $\partial D^n = S^{n-1}$.

Exercise 4.7.3 Show that if D is a domain with smooth boundary in X, then ∂D has measure zero in X and conclude that D is measurable in X.

Next we wish to show that, when D is a domain with smooth boundary in X, the submanifold ∂D inherits a natural orientation from X.

Remark: This induced orientation on ∂D is “natural” in the somewhat indirect sense that it supplies ∂D with the orientation it must have in order for Stokes' Theorem to be true. This sort of thing is to be expected, of course, from the classical version of Stokes' Theorem

$$\int_{\partial D} \vec{F} \cdot \vec{T}\, ds = \iint_{D} \vec{N} \cdot \operatorname{curl} \vec{F}\, dS$$

in which the orientation (tangent vector $\vec{T}$) to the closed curve ∂D must be chosen consistent with the orientation (normal vector $\vec{N}$) of the surface D. Thus, for a given orientation of S^2 (outward normal, say), two applications of Stokes' Theorem, one to the upper hemisphere S^2_+ and one to the lower hemisphere S^2_-, will necessitate opposite orientations on their common boundary S^1 (counterclockwise and clockwise, respectively). The "natural" orientation for ∂D depends on the rest of D.

The induced orientation on ∂D is defined at each $p \in \partial D$ in terms of an "outward" vector in $T_p(X)$. Since ∂D is a submanifold of X we can identify $T_p(\partial D)$ with a subspace of $T_p(X)$ in the usual way. Choose a chart (U, φ) at p in X, consistent with the orientation of X, and with $\varphi(p) = 0$ and $\varphi(U \cap D) = \varphi(U) \cap {}^n_+$. Let $x^1, \ldots, x^{n-1}$, x^n be the coordinate functions of φ. By Exercise 4.7.1, $\varphi(U \cap \partial D) = \{x \in \varphi(U) : x^n = 0\}$ is a coordinate neighborhood at p in ∂D so $T_p(\partial D)$ is spanned by

$$\left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^{n-1}}\right|_p.$$

A tangent vector $\boldsymbol{u} = a^i \frac{\partial}{\partial x^i}|_p$ in $T_p(X)$ with $a^n \neq 0$ is not in $T_p(\partial D)$ and we will say that $\boldsymbol{u}$ is **outward pointing** if $a^n < 0$ (think about $D = {}^n_+$ in $X = {}^n$).

Exercise 4.7.4 Show that this definition does not depend on the choice of (U, φ), i.e., that if (V, ψ) is another chart at p in X, consistent with the orientation of X, and with $\psi(p) = 0$ and $\psi(V \cap D) = \psi(V) \cap {}^n_+$ and $\boldsymbol{u} = b^i \frac{\partial}{\partial y^i}|_p$, where $y^1, \ldots, y^{n-1}$, y^n are the coordinate functions of ψ, then a^n and b^n have the same sign.

Now we define the **induced orientation for** $T_p(\partial D)$ by decreeing that a basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ for $T_p(\partial D)$ is in this orientation if and only if $\{\boldsymbol{v}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ is an oriented basis for $T_p(X)$ for any outward pointing tangent vector $\boldsymbol{u} \in T_p(X)$.

Exercise 4.7.5 Show that the induced orientation for $T_p(\partial D)$ is well-defined, i.e., does not depend on the choice of outward pointing vector $\boldsymbol{u} \in T_p(X)$.

Exercise 4.7.6 Let $X = {}^n$ with its standard orientation and $D = {}^n_+$ so that $\partial D \cong {}^{n-1}$. Show that, at each $p \in \partial D$, the induced orientation for $T_p(\partial D)$ is the standard orientation of ${}^{n-1}$ if and only if n is even.

Exercise 4.7.7 Let $X = S^n$ (with either orientation), $D_1 = S^n_+ = \{(x^1, \ldots, x^{n+1}) \in S^n : x^{n+1} \geq 0\}$ and $D_2 = S^n_- = \{(x^1, \ldots, x^{n+1}) \in S^n : x^{n+1} \leq 0\}$. Then $\partial D_1 \cong S^{n-1}$ and $\partial D_2 \cong S^{n-1}$. Show that, at each $p \in S^{n-1}$, the induced orientations on ∂D_1 and ∂D_2 are opposite.

Now we show that there is a global orientation (nonzero $(n-1)$-form) on ∂D that determines the induced orientation on each $T_p(\partial D)$. First we define a smooth vector field $\boldsymbol{V}$ on an open neighborhood of ∂D in X which, at each $p \in \partial D$, is outward pointing. This is easy to do locally: For each $p \in \partial D$ select an oriented chart (U_p, φ_p) for X at p with $\varphi_p(p) = 0$ and $\varphi_p(U_p \cap D) = \varphi_p(U_p) \cap {}^n_+$. If $x^1, \ldots, x^{n-1}, x^n$ are the corresponding coordinate functions, then

$$\boldsymbol{V}_p = -\frac{\partial}{\partial x^n}$$

is a smooth vector field on U_p that is outward pointing at each point of $U_p \cap \partial D$. Then $\{U_p : p \in \partial D\}$ covers ∂D so we can select a countable subcover $\{U_n\}_{n=1,2,\ldots}$ and a partition of unity $\{\phi_n\}_{n=1,2,\ldots}$ with $\operatorname{supp} \phi_n \subseteq U_n$ for each $n = 1, 2, \ldots$. If $\boldsymbol{V}_n$ is the outward pointing vector field on $U_n \cap \partial D$, then we define $\boldsymbol{V}$ on $U = \bigcup_{n=1}^{\infty} U_n$ by

$$\boldsymbol{V} = \sum_{n=1}^{\infty} \phi_n \boldsymbol{V}_n.$$

This sum is locally finite and therefore defines a smooth vector field on $U \supseteq \partial D$ which is outward pointing on ∂D since $0 \leq \phi_n \leq 1$.

Now, let $\boldsymbol{\omega}$ be a nonzero n-form on X that determines the orientation of X (i.e., $\boldsymbol{\omega}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}, \boldsymbol{v}_n) > 0$ whenever $p \in X$ and $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}, \boldsymbol{v}_n\}$ is an oriented basis for $T_p(X)$). Regard $\boldsymbol{\omega}$ as a $C^\infty(X)$-multilinear map on $\mathcal{X}(X) \times \overset{n}{\cdots} \times \mathcal{X}(X)$. The restriction (pullback) of $\boldsymbol{\omega}$ to U, which we will also denote $\boldsymbol{\omega}$, operates on $\mathcal{X}(U) \times \overset{n}{\cdots} \times \mathcal{X}(U)$. Thus, we may define an $(n-1)$-form $\tilde{\boldsymbol{\omega}}$ on U by

$$\tilde{\boldsymbol{\omega}}(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{n-1}) = \boldsymbol{\omega}(\boldsymbol{V}, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{n-1}),$$

where $\boldsymbol{V}_i \in \mathcal{X}(U)$ for $i = 1, \ldots, n-1$ and $\boldsymbol{V}$ is the vector field on U defined above that is outward pointing on ∂D. For $p \in \partial D$ and $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ a basis for $T_p(\partial D)$, $\tilde{\boldsymbol{\omega}}_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}) > 0$ if and only if $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}\}$ is in the induced orientation for $T_p(\partial D)$. The $(n-1)$-form $\tilde{\boldsymbol{\omega}}$ determines the **induced orientation on** ∂D.

Theorem 4.7.1 (Stokes' Theorem) *Let X be a smooth, oriented, n-dimensional manifold, D a domain with smooth boundary in X and $\iota : \partial D \hookrightarrow X$ the inclusion map. Provide ∂D with the induced orientation. If $\boldsymbol{\omega}$ is any smooth $(n-1)$-form on X with compact support on X, then*

$$\int_{\partial D} \iota^* \boldsymbol{\omega} = \int_D d\boldsymbol{\omega}. \tag{4.7.2}$$

Proof: Observe first that, since $\boldsymbol{\omega}$ is smooth with compact support, the same is true of $\iota^* \boldsymbol{\omega}$ and $d\boldsymbol{\omega}$ so both of these are integrable. Now, choose a countable, oriented atlas $\{ (U_k, \varphi_k) \}_{k=1,2,\ldots}$ for X such that each $\varphi_k(U_k)$ is a bounded subset of $\mathbb{R}^n$ and either $U_k \cap \partial D = \emptyset$ (if $U_k \cap D \neq \emptyset$ we assume in this case that $U_k \subseteq \operatorname{Int} D$), or, if $U_k \cap \partial D \neq \emptyset$, then $\varphi_k(U_k \cap D) = \varphi_k(U_k) \cap \mathbb{R}^n_+$. Let $\{ \phi_k \}_{k=1,2,\ldots}$ be a family of real-valued functions on X of the sort guaranteed by Corollary 3.1.5. Then $\boldsymbol{\omega} = \sum_{k=1}^{\infty} \phi_k \boldsymbol{\omega}$, where the sum has only finitely many nonzero terms. Thus $d\boldsymbol{\omega} = \sum_{k=1}^{\infty} d(\phi_k \boldsymbol{\omega})$ and $\iota^* \boldsymbol{\omega} = \sum_{k=1}^{\infty} \iota^*(\phi_k \boldsymbol{\omega})$ are finite sums, as are each of the following:

$$\int_{\partial D} \iota^* \boldsymbol{\omega} = \sum_{k=1}^{\infty} \int_{\partial D} \iota^*(\phi_k \boldsymbol{\omega})$$

$$\int_D d\boldsymbol{\omega} = \sum_{k=1}^{\infty} \int_D d(\phi_k \boldsymbol{\omega}).$$

Thus, it will suffice to prove that

$$\int_{\partial D} \iota^* (\phi_k \boldsymbol{\omega}) = \int_D d (\phi_k \boldsymbol{\omega})$$

for each k and this simply amounts to proving (4.7.2) in the special case in which $\operatorname{supp} \boldsymbol{\omega}$ is contained in some U_k.

Assume then that $\operatorname{supp} \boldsymbol{\omega} \subseteq U_k$ and consider first the case in which $U_k \cap \partial D = \emptyset$. Then $\int_{\partial D} \iota^* \boldsymbol{\omega} = 0$ and we must show that $\int_D d\boldsymbol{\omega}$ is zero as well. This is obvious if $U_k \cap D = \emptyset$ since $\operatorname{supp} d\boldsymbol{\omega} \subseteq \operatorname{supp} \boldsymbol{\omega} \subseteq U_k$. If $U_k \cap D \neq \emptyset$, then $U_k \subseteq \operatorname{Int} D$ so

$$\int_D d\boldsymbol{\omega} = \int_{U_k} d\boldsymbol{\omega} = \int_X d\boldsymbol{\omega}$$

and this is zero by Theorem 4.6.2.

Finally, we consider the case in which $\operatorname{supp} \boldsymbol{\omega} \subseteq U_k$ and $U_k \cap \partial D \neq \emptyset$. Thus, $\varphi_k(U_k \cap D) = \varphi_k(U_k) \cap \mathbb{R}^n_+$. Let $x^1, \ldots, x^n$ be the coordinate functions of φ_k and write

$$\boldsymbol{\omega} = \sum_{i=1}^{n} (-1)^{i-1} \omega_i \, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n$$

on U_k (including the $(-1)^{i-1}$ here will spare us a few of these down the road). Thus,

$$d\boldsymbol{\omega} = \left(\sum_{i=1}^{n} \frac{\partial}{\partial x^i} (\omega_i \circ \varphi_k^{-1}) \circ \varphi_k \right) dx^1 \wedge \cdots \wedge dx^n \tag{4.7.3}$$

on U_k, and, since $\iota^*(dx^n) = d(x^n \circ \iota) = 0$ by Exercise 4.7.1,

$$\begin{aligned}\iota^*\boldsymbol{\omega} &= (-1)^{n-1}(\omega_n \circ \iota)\, \iota^*\,(dx^1 \wedge \cdots \wedge dx^{n-1})\\ \iota^*\boldsymbol{\omega} &= (-1)^{n-1}(\omega_n \circ \iota)\, d\,(x^1 \circ \iota) \wedge \cdots \wedge d(x^{n-1} \circ \iota)\end{aligned} \tag{4.7.4}$$

on $U_k \cap \partial D$. Now, the map $\varphi\colon U_k \cap \partial D \longrightarrow \ {}^{n-1}$ with coordinate functions $(x^1 \circ \iota, \ldots, x^{n-1} \circ \iota)$ is a chart on ∂D.

Exercise 4.7.8 Show that the orientation $d(x^1 \circ \iota) \wedge \cdots \wedge d(x^{n-1} \circ \iota)$ on $U_k \cap \partial D$ is $(-1)^n$ times the orientation induced by $dx^1 \wedge \cdots \wedge dx^{n-1} \wedge dx^n$ on U_k. **Hint**: $dx^1 \wedge \cdots \wedge dx^{n-1} \wedge dx^n = (-1)^n(-dx^n) \wedge dx^1 \wedge \cdots \wedge dx^{n-1}$.

Since $\iota^*\boldsymbol{\omega}$ has compact support contained in $U_k \cap \partial D$, Exercise 4.7.8 and (4.7.4) give

$$\begin{aligned}\int_{\partial D} \iota^*\boldsymbol{\omega} &= (-1)^n \int_{\varphi(U_k \cap \partial D)} (-1)^{n-1}(\omega_n \circ \iota) \circ \varphi^{-1}\, dm'\\ &= -\int_{\varphi(U_k \cap \partial D)} \omega_n \circ (\iota \circ \varphi^{-1}) dm'\end{aligned} \tag{4.7.5}$$

where m' denotes Lebesgue measure on ${}^{n-1}$. Notice that $\iota \circ \varphi^{-1}$ is the φ-coordinate expression for the inclusion map which, in φ_k-coordinates, is just $(x^1, \ldots, x^{n-1}) \longrightarrow (x^1, \ldots, x^{n-1}, 0)$. Thus, $(\omega_n \circ (\iota \circ \varphi^{-1}))(x^1, \ldots, x^{n-1}) = (\omega_n \circ \varphi^{-1})(x^1, \ldots, x^{n-1}, 0)$ and, since $\iota^*\,\boldsymbol{\omega}$ has compact support contained in $U_k \cap \partial D$, (4.7.5) can be written

$$\int_{\partial D} \iota^*\,\boldsymbol{\omega} = \int_{\ {}^{n-1}} (\omega_n \circ \varphi_k^{-1})(x^1, \ldots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1}. \tag{4.7.6}$$

Now we turn to the integral over D of $d\,\boldsymbol{\omega}$. By (4.7.3), the fact that $\varphi_k(U_k \cap D) = \varphi_k(U_k) \cap \ {}^n_+$. and our assumption that $\boldsymbol{\omega}$ (and therefore $d\,\boldsymbol{\omega}$) has compact support contained in U_k, this is given by

$$\int_D d\,\boldsymbol{\omega} = \sum_{i=1}^{n} \int_{\ {}^n} \chi_{\ {}^n_+} \frac{\partial(\omega_i \circ \varphi_k^{-1})}{\partial x^i}\, dm\,,$$

where $\chi_{\ {}^n_+}$ is the characteristic function of ${}^n_+$, i.e., 1 if $x^n \geq 0$ and 0 if $x_n < 0$, and, as in the proof of Theorem 4.6.2, we have taken $\omega_i \circ \varphi_k^{-1}$ to be zero outside of $\varphi_k^{-1}(U_k)$. Also as in the proof of Theorem 4.6.2, we now choose a cube $C = \{\,(x^1, \ldots, x^n) \in \ {}^n : |x^i| \leq a,\ i = 1, \ldots, n\}$ sufficiently large to contain $\varphi_k(U_k)$. Thus,

$$\int_D d\,\boldsymbol{\omega} = \sum_{i=1}^{n} \int_C \chi_{\ {}^n_+} \frac{\partial(\omega_i \circ \varphi_k^{-1})}{\partial x^i}\, dm. \tag{4.7.7}$$

Exercise 4.7.9 Argue as in the proof of Theorem 4.6.2 to show that, for $i = 1, \ldots, n-1$ (but *not* n),

$$\int_C \chi_{\ {}^n_+} \frac{\partial(\omega_i \circ \varphi_k^{-1})}{\partial x^i}\, dm = 0 \tag{4.7.8}$$

and conclude that

$$\int_D d\boldsymbol{\omega} = \int_C \chi_{\;{}^n_+} \frac{\partial(\omega_n \circ \varphi_k^{-1})}{\partial x^n}\, dm\,. \tag{4.7.9}$$

The argument used to establish (4.7.8) works because, if only $x^i (i = 1, \ldots, n-1)$ is varying in $\varphi_k(U_k \cap D)$, then $\chi_{\;{}^n_+}$ is equal to 1 everywhere. In the x^n-direction, of course, multiplying by $\chi_{\;{}^n_+}$ will yield a function that is zero on $(-\infty, 0)$ and unchanged on $[0, \infty)$. Thus, when we express the right-hand side of (4.7.9) as an iterated integral, integrating first with respect to x^n, the integrand becomes

$$\begin{aligned}\int_0^a \frac{\partial(\omega_n \circ \varphi_k^{-1})}{\partial x^n}\, dx^n &= (\omega_n \circ \varphi_k^{-1})\,(x^1, \ldots, x^{n-1}, a) \\ &\qquad - (\omega_n \circ \varphi_k^{-1})\,(x^1, \ldots, x^{n-1}, 0) \\ &= -(\omega_n \circ \varphi_k^{-1})\,(x^1, \ldots, x^{n-1}, 0)\end{aligned}$$

so

$$\int_D d\boldsymbol{\omega} = -\int_{\;{}^{n-1}} (\omega_n \circ \varphi_k^{-1})\,(x^1, \ldots, x^{n-1}, 0)\, dx^1 \cdots dx^{n-1} = \int_{\partial D} \iota^* \boldsymbol{\omega}$$

as required. ■

We conclude with an analogue of the familar principle of deformation of paths for line integrals in vector calculus. Let X be a smooth, oriented n-dimensional manifold and let N_1 and N_2 be two oriented k-dimensional submanifolds, where $1 \leq k \leq n$. We say that N_1 can be smoothly **deformed** into N_2 if there exists a smooth map $H : N \times (0,3) \longrightarrow X$ such that $H|N \times \{1\} \longrightarrow N_1$ and $H|N \times \{2\} \longrightarrow N_2$ are orientation preserving diffeomorphisms. We claim that, if $\boldsymbol{\omega}$ is any closed k-form on X and $\iota_1 : N_1 \hookrightarrow X$ and $\iota_2 : N_2 \hookrightarrow X$ are the inclusion maps, then

$$\int_{N_1} \iota_1^* \boldsymbol{\omega} = \int_{N_2} \iota_2^* \boldsymbol{\omega}\,. \tag{4.7.10}$$

Exercise 4.7.10 Prove (4.7.10). **Hint**: $H^* \boldsymbol{\omega}$ is a closed k-form on the $(k+1)$-manifold $N \times (0,3)$ and $N \times [0,2]$ is a domain with smooth boundary in $N \times (0,3)$.

5
de Rham Cohomology

Introduction

The plane ^{2} and the punctured plane $^2 - \{(0,0)\}$ are not diffeomorphic, nor even homeomorphic. There are various means by which one can prove this, but the most instructive among these detect the "hole" in the punctured plane (and none in ^{2}) by distinguishing topologically certain "types" of circles that can live in the two spaces. One way of formalizing this idea is to show that ^{2} and $^2 - \{(0,0)\}$ have different fundamental groups (see pages 107 and 133 of [**N4**]). Fundamental groups (and their higher dimensional generalizations) are notoriously difficult to calculate, however. In this chapter we will investigate another method of attaching to any smooth manifold certain algebraic objects which likewise "detect holes". The idea is quite simple and, for ^{2} and $^2 - \{(0,0)\}$, has already been hinted at in Section 4.4. There we found that any closed 1-form on ^{2} is necessarily exact (Poincaré Lemma), but constructed a closed 1-form $\boldsymbol{\omega}$ on $^2 - \{(0,0)\}$ that is not exact. Stokes' Theorem implies that the integral of any 1-form on ^{2} over any circle (compact, connected, 1-dimensional submanifold) is zero and the same is true of $\boldsymbol{\omega}$ provided the circle does not contain $(0,0)$ in its interior. For a circle S^1 that does contain $(0,0)$ in its interior (and has the orientation induced from $^2 - \{(0,0)\}$) we computed $\int_{S^1} \boldsymbol{\omega}$ and obtained 2π. The idea then is that the closed, nonexact 1-form $\boldsymbol{\omega}$ detects the hole in $^2 - \{(0,0)\}$ via its integrals over circles. The prospect then arises that one might distill topological information from the existence of closed, nonexact forms on a manifold. Unfortunately, the collection of all such forms (even of some fixed degree) can be a huge, unmanageable object. Observe, however, that another closed 1-form $\boldsymbol{\omega}'$ on $^2 - \{(0,0)\}$ that differs from $\boldsymbol{\omega}$ by an exact form $(\boldsymbol{\omega}' = \boldsymbol{\omega} + d\boldsymbol{\eta})$ has, by Stokes' Theorem, the same integral as $\boldsymbol{\omega}$ over any circle and so for our purposes, there is no reason to distinguish them. The natural thing to do then is define an equivalence relation on the set of closed 1-forms whereby two such forms are equivalent (we shall say "cohomologous") if they differ by an exact form. The resulting set of equivalence classes inherits a natural vector space structure from the space of 1-forms and, with this structure, is called the 1^{st} de Rham cohomology group of the manifold. For ^{2} it has dimension 0 (no holes), while for $^2 - \{(0,0)\}$ it has dimension 1 (one hole). The precise construction of this vector space and its higher dimensional analogues is the task we set for ourselves in the next section. There are many formal similarities with homology theory (Chapter 3 of [**N4**]), but the arguments tend to be substantially easier.

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0_5,

5.1 The de Rham Cohomology Groups

Throughout this section X will denote an n-dimensional smooth manifold and, for each integer $k \geq 1$, $\Lambda^k(X)$ is its vector space of smooth, real-valued k-forms. Then $\Lambda^k(X) = 0$ for $k \geq n+1$. For convenience, we will take $\Lambda^0(X) = C^\infty(X)$ and $\Lambda^k(X) = 0$ for $k < 0$. The exterior differentiation maps

$$d^k : \Lambda^k(X) \longrightarrow \Lambda^{k+1}(X), \quad k = 0, 1, \ldots, n$$

will, unless particular emphasis is required, all be written d and we will take d^k to be the trivial homomorphism when $k < 0$ and $k \geq n$. Thus, we have a sequence

$$\cdots \xrightarrow{d} \Lambda^{-1}(X) \xrightarrow{d} \Lambda^0(X) \xrightarrow{d} \Lambda^1(X) \xrightarrow{d} \cdots$$
$$\xrightarrow{d} \Lambda^{n-1}(X) \xrightarrow{d} \Lambda^n(X) \xrightarrow{d} \cdots$$

of vector spaces and linear maps. Since $d^k \circ d^{k-1} = 0$ for each k (Theorem 4.1.1 (3)) we have

$$\text{Image}\,(d^{k-1}) \subseteq \ker\,(d^k).$$

The elements of $\ker\,(d^k)$ are the closed k-forms, or **de Rham k-cocycles**, on X, while the elements of Image (d^{k-1}) are the exact k-forms, or **de Rham k-coboundaries**, on X. Both Image (d^{k-1}) and $\ker\,(d^k)$ are linear subspaces of $\Lambda^k(X)$ and so Image (d^{k-1}) is a subspace of $\ker\,(d^k)$. The quotient vector space

$$H^k_{\text{de R}}(X) = \ker\,(d^k)\,/\,\text{Image}\,(d^{k-1})$$

is called the **k^{th} de Rham cohomology group** of X.

Remark: $H^k_{\text{de R}}(X)$ is a *vector space*, but the tradition of referring to it as a cohomology *group* is so firmly embedded in the literature that any attempt to correct the terminology here would be pointless.

The elements of $H^k_{\text{de R}}(X)$ are equivalence classes of closed k-forms on X where the equivalence relation is defined as follows: If $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ are closed k-forms on X, then $\boldsymbol{\omega}' \sim \boldsymbol{\omega}$ if and only if $\boldsymbol{\omega}' - \boldsymbol{\omega} = d\boldsymbol{\eta}$ for some $\boldsymbol{\eta} \in \Lambda^{k-1}(X)$ (in this case we say that $\boldsymbol{\omega}'$ and $\boldsymbol{\omega}$ are **cohomologous**). The equivalence class containing $\boldsymbol{\omega}$ will be denoted $[\boldsymbol{\omega}]$ and called the **cohomology class** of $\boldsymbol{\omega}$.

Since $H^k_{\text{de R}}(X) = 0$ for $k \geq n+1$ and for $k \leq -1$, the only (possibly) nontrivial cohomology groups are

$$H^0_{\text{de R}}(X),\ H^1_{\text{de R}}(X), \ldots,\ H^n_{\text{de R}}(X) \qquad (n = \dim X).$$

$H^0_{\text{de R}}(X)$ is easily calculated. Since Image $(d^{-1}) = 0$, $H^0_{\text{de R}}(X) = \ker\,(d^0)$ and this is the set of all $f \in \Lambda^0(X) = C^\infty(X)$ for which $df = 0$.

Exercise 5.1.1 Show that $df = 0$ if and only if f is constant on each connected component of X.

In particular, if X is connected, then $H^0_{\text{de R}}(X)$ is just the subspace of $C^\infty(X)$ consisting of constant functions and this is clearly isomorphic to . If X is not connected, then (since we assume all manifolds are second countable) it has at most countably many connected components $C_1, C_2, \ldots$. Each characteristic function χ_{C_i} is an element of $H^0_{\text{de R}}(X)$ as is each (locally finite) sum $\sum_{i=1}^\infty r_i \chi_{C_i}$, where $r_i \in$. Indeed, every element of $H^0_{\text{de R}}(X)$ is such a sum. Thus, if X has only finitely many components (say, l), then $H^0_{\text{de R}}(X) \cong {}^l$. Otherwise, $H^0_{\text{de R}}(X)$ is the infinite dimensional vector space $\times \times \cdots = {}^{\aleph_0}$. Combining these observations with the Poincaré Lemma yields our first nontrivial result on cohomology groups.

Theorem 5.1.1 *Let X be an open, star-shaped subset of ^{n} (e.g., ^{n} itself). Then $H^0_{\text{de R}}(X) \cong$ and $H^k_{\text{de R}}(X) = 0$ for all $k \neq 0$.*

Remark: A star-shaped subset of ^{n} is contractible, but the converse is not true. We will, however, eventually show that any contractible manifold X has $H^k_{\text{de R}}(X) = 0$ for all $k \neq 0$.

Consider a compact, connected, oriented n-manifold X. One of our major results in this chapter is that $H^n_{\text{de R}}(X)$ is 1-dimensional, i.e., isomorphic to . This will allow us, in particular, to define the notion of the *degree* of a map between such manifolds. For the time being we will content ourselves with showing that $H^n_{\text{de R}}(X)$ cannot be trivial (keeping in mind the previous Remark, this alone will prove that no compact, connected, orientable manifold of dimension $n > 0$ is contractible). To find a nontrivial cohomology class in $H^n_{\text{de R}}(X)$ it will suffice to find a closed n-form on X that is not exact. According to Stokes' Theorem any exact n-form $\boldsymbol{\omega}$ on X satisfies $\int_X \boldsymbol{\omega} = 0$ and so we need only find a closed n-form on X whose integral over X is not zero. But *any* n-form on an n-manifold is closed because $\Lambda^{n+1}(X) = 0$ so we need only find *some* n-form on X which does not integrate to zero over X.

Exercise 5.1.2 Let $\boldsymbol{\mu}$ be an n-form determining the orientation of X. Show that $\int_X \boldsymbol{\mu} \neq 0$ and conclude that the cohomology class $[\boldsymbol{\mu}] \in H^n_{\text{de R}}(X)$ is nontrivial.

Noting that the argument we have just given does not require connectivity, we record the following lemma.

Lemma 5.1.2 *Let X be a compact, orientable, n-manifold. Then* $\dim H^n_{\text{de R}}(X) \geq 1$.

The efficient calculation of cohomology groups will require that we first deal with some of the more formal aspects of the subject. There is, however, at least one small example accessible to us at this stage. We know that $H^0_{\text{de R}}(S^1) =$ and $H^k_{\text{de R}}(S^1) = 0$ for $k \geq 2$ and $k \leq -1$. Furthermore, $H^1_{\text{de R}}(S^1)$ has dimension at least one by Lemma 5.1.2. We show that, in fact,

$$H^1_{\text{de R}}(S^1) \cong \ . \qquad (5.1.1)$$

Remark: We will show later that S^1 is a (smooth) deformation retract of $\ ^2 - \{(0,0)\}$ and that this implies that they have the same de Rham cohomology groups. Consequently, it will follow from (5.1.1) that $H^k_{\text{de R}}(\ ^2 - \{(0,0)\}) \cong$ as we claimed in the Introduction.

Since we know from Lemma 5.1.2 that $\dim H^1_{\text{de R}}(S^1) \geq 1$ we need only prove that $\dim H^1_{\text{de R}}(S^1) \leq 1$ in order to establish (5.1.1). For this it will suffice to find a nontrivial cohomology class $[\boldsymbol{\eta}] \in H^1_{\text{de R}}(S^1)$ such that, for any 1-form $\boldsymbol{\alpha}$ on S^1, there exists a constant α for which $\boldsymbol{\alpha} - \alpha\boldsymbol{\eta}$ is exact (since then $[\boldsymbol{\alpha}] = [\alpha\boldsymbol{\eta}] = \alpha[\boldsymbol{\eta}]$). For this we let $\iota : S^1 \hookrightarrow \ ^2 - \{(0,0)\}$ be the inclusion of S^1 into the punctured plane and define $\boldsymbol{\eta} = \iota^* \boldsymbol{\omega}$, where $\boldsymbol{\omega}$ is the 1-form on $\ ^2 - \{(0,0)\}$ defined (in Section 4.4) from two angular coordinate functions on $\ ^2 - \{(0,0)\}$. Specifically,

$$\boldsymbol{\omega} = -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy$$

so $\boldsymbol{\eta}$ is just the standard volume form on S^1. In Section 4.6 we found that $\int_{S^1} \boldsymbol{\eta} = 2\pi$ so, in particular, $[\boldsymbol{\eta}] \in H^1_{\text{de R}}(S^1)$ is nontrivial. Now let $\boldsymbol{\alpha}$ be an arbitrary 1-form on S^1 and set

$$\alpha = \frac{1}{2\pi}\int_{S^1} \boldsymbol{\alpha}.$$

We claim that $\boldsymbol{\alpha} - \alpha\boldsymbol{\eta}$ is exact. Of course,

$$\int_{S^1} (\boldsymbol{\alpha} - \alpha\boldsymbol{\eta}) = 0$$

so our result will follow from the next lemma.

Lemma 5.1.3 *A 1-form $\boldsymbol{\nu}$ on S^1 is exact if and only if $\int_{S^1} \boldsymbol{\nu} = 0$.*

Proof: We have already seen that any exact form integrates to zero so suppose conversely that $\int_{S^1} \boldsymbol{\nu} = 0$. Define $h : \ \to S^1$ by $h(t) = (\cos t,\ \sin t)$ and $g : \ \to \ $ by

$$g(t) = \int_0^t h^* \boldsymbol{\nu}.$$

We have shown in Section 4.6 that $\int_{S^1} \boldsymbol{\nu}$ is equal to the integral of $h^*\boldsymbol{\nu}$ over any interval of length 2π so our assumption implies that

$$g(t + 2\pi) = g(t)$$

for any $t \in \ $. Consequently, there is a well-defined smooth function $f : S^1 \to \ $ satisfying

$$g = f \circ h$$

(for each $p \in S^1$, $f(p) = g(t)$, where t is any real number with $h(t) = p$).

Exercise 5.1.3 Show that $df = \boldsymbol{\nu}$. ∎

Remark: We will show later (Theorem 5.4.3) that an n-form $\boldsymbol{\nu}$ on S^n is exact if and only if $\int_{S^n} \boldsymbol{\nu} = 0$ and, later still (Corollary 5.5.5), that an n-form $\boldsymbol{\nu}$ on any compact, orientable n-manifold X is exact if and only if $\int_X \boldsymbol{\nu} = 0$.

To develop efficient techniques for calculating the cohomology groups of other manifolds we must delve more deeply into some of the more formal, algebraic aspects of the subject. We will take this up in the next few sections. Before doing so, however, we must point out that in the next chapter we will find ourselves briefly in need of a rather obvious generalization of de Rham groups as constructed here. Our objective in Chapter 6 is to associate with each principal bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ certain cohomology classes of the base manifold X called the Chern classes of the bundle. These are the cohomology classes of certain closed forms constructed from a connection on the bundle. The construction is simple enough, but *a priori* it would seem to yield complex, rather than real forms. One can show that they are, in fact, real, but until this is done one must carry out the development within the context of the "complex de Rham cohomology groups $H^k_{\text{de R}}(X;\quad)$." These are constructed in precisely the same way as the groups $H^k_{\text{de R}}(X)$ except that one begins with complex-valued forms $\Lambda^k(X;\quad)$ and ends up with complex vector spaces. We leave it to the reader to write out the details of the construction and to generalize all of the algebraic machinery we will assemble. Here's your first opportunity. We shall denote by $H^*_{\text{de R}}(X)$ the direct sum $\bigoplus_{k=0}^{\infty} H^k_{\text{de R}}(X)$ of the cohomology groups of X (of course, $H^k_{\text{de R}}(X) \cong 0$ for $k > \dim X$).

Remark: Recall that if $\mathcal{V}_0, \mathcal{V}_1, \ldots$ are vector spaces, then their direct product $\prod_{k=0}^{\infty} \mathcal{V}_k$ is the Cartesian product $\mathcal{V}_0 \times \mathcal{V}_1 \times \cdots$ with its coordinatewise linear structure, while their direct sum $\bigoplus_{k=0}^{\infty} \mathcal{V}_k$ is the subspace of $\prod_{k=0}^{\infty} \mathcal{V}_k$ consisting of those elements with at most finitely many coordinates nonzero. Each $\mathcal{V}_k$ can then be canonically identified with a subspace of $\bigoplus_{k=0}^{\infty} \mathcal{V}_k$ and this direct sum can then be viewed as the *internal* direct sum of these subspaces. In this way we view the elements of $\bigoplus_{k=0}^{\infty} \mathcal{V}_k$ as finite "formal sums" of elements from the various $\mathcal{V}_k,\ k = 0, 1, \ldots$. The elements of $H^*_{\text{de R}}(X)$ are therefore to be thought of as finite formal sums of cohomology classes of various degrees.

Exercise 5.1.4 Show that if $[\boldsymbol{\omega}_1] \in H^k_{\text{de R}}(X)$ and $[\boldsymbol{\omega}_2] \in H^l_{\text{de R}}(X)$, then

$$[\boldsymbol{\omega}_1] \wedge [\boldsymbol{\omega}_2] = [\boldsymbol{\omega}_1 \wedge \boldsymbol{\omega}_2]$$

is a well-defined element of $H^{k+l}_{\text{de R}}(X)$. Now extend the product $\wedge$ to all of $H^*_{\text{de R}}(X)$ (by the "distributive law") and show that $H^*_{\text{de R}}(X)$ thereby acquires the structure of an algebra with identity.

In the same way one constructs the complex de Rham cohomology algebra $H^*(X; \quad)$.

5.2 Induced Homomorphisms

Let X and Y be smooth manifolds and $f : X \to Y$ a smooth map. For each k, the pullback $f^* : \Lambda^k(Y) \to \Lambda^k(X)$ carries closed forms to closed forms ($d\boldsymbol{\omega} = 0$ implies $d(f^*\boldsymbol{\omega}) = f^*(d\boldsymbol{\omega}) = f^*(0) = 0$) and exact forms to exact forms ($\boldsymbol{\omega} = d\boldsymbol{\eta}$ implies $f^*\boldsymbol{\omega} = f^*(d\boldsymbol{\eta}) = d(f^*\boldsymbol{\eta})$). Thus, we may define a linear map

$$f^\# : H^k_{\text{de R}}(Y) \longrightarrow H^k_{\text{de R}}(X)$$

by

$$f^\#([\boldsymbol{\omega}]) = [f^*\boldsymbol{\omega}] \tag{5.2.1}$$

for each $[\boldsymbol{\omega}] \in H^k_{\text{de R}}(Y)$. Notice that this definition makes sense because $f^*\boldsymbol{\omega}$ is closed and is independent of the choice of representative $\boldsymbol{\omega}$ of the cohomology class $[\boldsymbol{\omega}]$ because $\boldsymbol{\omega'} - \boldsymbol{\omega} = d\boldsymbol{\eta}$ implies $f^*\boldsymbol{\omega'} - f^*\boldsymbol{\omega} = d(f^*\boldsymbol{\eta})$. Observe also that if $Y = X$ and $f = id_X$, then $f^\#$ is the identity on each $H^k_{\text{de R}}(X)$, i.e.,

$$(id_X)^\# = id_{H^k_{\text{de R}}(X)} \tag{5.2.2}$$

for each k. Moreover, from the corresponding property of pullbacks we find that, if $f : X \to Y$ and $g : Y \to Z$ are smooth, then

$$(g \circ f)^\# = f^\# \circ g^\#. \tag{5.2.3}$$

In particular, if $f : X \to Y$ and $g : Y \to X$ are inverse diffeomorphisms, then

$$f^\# : H^k_{\text{de R}}(Y) \longrightarrow H^k_{\text{de R}}(X) \quad \text{and} \quad g^\# : H^k_{\text{de R}}(X) \longrightarrow H^k_{\text{de R}}(Y)$$

are inverse isomorphisms for each k. Consequently, the de Rham cohomology groups are diffeomorphism invariants.

Remark: The de Rham cohomology groups are actually invariants of homotopy type, but this is by no means clear (see the Remark following Corollary 5.2.4).

The maps $f^\#$ are said to be those **induced in cohomology** by f. Properties (5.2.2) and (5.2.3) of these induced maps should be compared with the analogous properties of induced maps in homotopy groups (Theorems 2.2.5 and 2.5.4 of [**N4**]) and singular homology groups (Section 3.2 of [**N4**]). They differ only in the direction of the arrows and express the "functorial" nature of the construction of the de Rham cohomology groups (in the jargon of category theory the direction of the arrows is accounted for by referring to the homotopy and homology functors as *covariant* and the cohomology functors as *contravariant*).

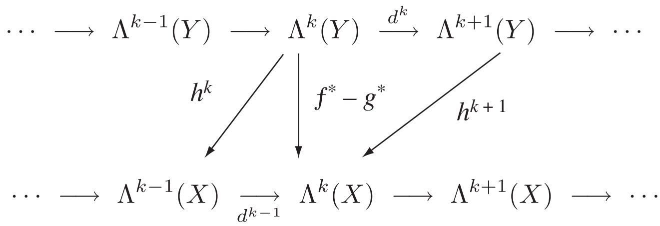
Given two smooth maps $f, g : X \to Y$ we will be interested in determining whether or not they induce the same linear maps in cohomology (i.e., $f^{\#} = g^{\#}$ for each k). A common method of proving that these induced maps are the same is based on an idea that we encountered in the proof of the Poincaré Lemma (Theorem 4.4.2) and is entirely analogous to the chain homotopies of homology theory (Section 3.3 of [**N4**]). Suppose there exists a family of linear maps

$$h^k : \Lambda^k(Y) \longrightarrow \Lambda^{k-1}(X)$$

such that

$$h^{k+1} \circ d^k + d^{k-1} \circ h^k = f^* - g^* \tag{5.2.4}$$

for each k.



Remark: This is *not* intended to be a commutative diagram. Rather, $f^* - g^*$ is the sum of the other edges in the adjacent triangles.

Then, evaluating both sides of (5.2.4) at some closed k-form $\boldsymbol{\omega}$ on Y gives

$$h^{k+1}\,(d^k\boldsymbol{\omega}) + d^{k-1}(h^k\boldsymbol{\omega}) = f^*\boldsymbol{\omega} - g^*\boldsymbol{\omega},$$

i.e.,

$$f^*\boldsymbol{\omega} - g^*\boldsymbol{\omega} = d(h^k\boldsymbol{\omega})$$

so $f^*\boldsymbol{\omega}$ and $g^*\boldsymbol{\omega}$ are cohomologous. Consequently, $f^{\#}([\boldsymbol{\omega}]) = g^{\#}([\boldsymbol{\omega}])$. Since $\boldsymbol{\omega}$ was arbitrary we conclude that $f^{\#} = g^{\#}$.

A family of maps $h^k : \Lambda^k(Y) \to \Lambda^{k-1}(X)$ satisfying (5.2.4) is called an **algebraic homotopy** (or **cochain homotopy**) between f and g and we have just shown that the existence of such a thing implies that f and g induce the same maps in cohomology. Our major result in this section will justify the rather peculiar terminology by showing that if f and g are (smoothly) homotopic maps of X into Y, then an algebraic homotopy between f and g always exists.

Two smooth maps $f, g : X \to Y$ are said to be **smoothly homotopic** if there exists a smooth map $F : X \times \quad \to Y$ such that $F(x, 0) = f(x)$

and $F(x,1) = g(x)$ for all $x \in X$. F is then a **smooth homotopy** from f to g. A map that is smoothly homotopic to a constant map is said to be **smoothly nullhomotopic**.

Remark: In Section 3.2 we proved that, when X and Y are open submanifolds of Euclidean spaces, two smooth maps that are homotopic in the usual topological sense (Section 2.3 of [**N4**]) are also smoothly homotopic. The same is true for any two manifolds X and Y, but the proof requires some rather substantial machinery (see [**Hir**]).

The proof that f and g induce the same map in cohomology is somewhat indirect and we begin by considering the following situation: Let U be an open subset of $\quad^n$ (with standard coordinate functions $x^1, \ldots, x^n$) and let t denote the standard coordinate function on $\quad$. Then $x^1, \ldots, x^n, t$ are standard coordinates on $U \times \quad \subseteq \quad^n \times \quad = \quad^{n+1}$. For each $k = 0, 1, \ldots$ we define

$$P : \Lambda^k(U \times \quad) \longrightarrow \Lambda^{k-1}(U \times$$

as follows: If $k = 0$ we take P to be identically zero. Each $\boldsymbol{\omega} \in \Lambda^k(U \times \quad)$ with $k \geq 1$ is uniquely expressible as

$$\boldsymbol{\omega}(x,t) = \sum_I \omega_I\,(x,t)\, dt \wedge dx^I + \sum_J \omega_J\,(x,t)\, dx^J$$

where $x = (x^1, \ldots, x^n)$ and I and J are increasing index sets of length $k-1$ and k, respectively (if $I = (i_1, \ldots, i_{k-1})$ with $1 \leq i_1 < \cdots < i_{k-1} \leq n$, then $dx^I = dx^{i_1} \wedge \cdots \wedge dx^{i_{k-1}}$, etc.). Now, if a is a fixed, but arbitrary real number we define $P\boldsymbol{\omega}$ by

$$P\boldsymbol{\omega}(x,t) = \sum_I \left[\int_a^t \omega_I\,(x,s)\, ds \right] dx^I. \tag{5.2.5}$$

Note that P sends the term $\sum_J \omega_J(x,t) dx^J$ to zero and that $P\boldsymbol{\omega}$ has no dt (although its components depend on t). Since these components are clearly C^∞, $P\boldsymbol{\omega}$ is indeed a $(k-1)$-form.

Now, suppose V is another open set in $\quad^n$ and $\phi : V \to U$ is a diffeomorphism. Then

$$\boldsymbol{\Phi} = \phi \times id \quad : V \times \quad \longrightarrow U \times$$

is a diffeomorphism and, for each $k = 0, 1, \ldots$

$$\boldsymbol{\Phi}^* : \Lambda^k(U \times \quad) \longrightarrow \Lambda^k(V \times \quad).$$

We claim that

$$\boldsymbol{\Phi}^*(P\boldsymbol{\omega}) = P(\boldsymbol{\Phi}^*\boldsymbol{\omega}) \tag{5.2.6}$$

for each $\boldsymbol{\omega} \in \Lambda^k(U \times \quad)$.

$$\begin{array}{ccc} \Lambda^k(U \times \mathbb{R}) & \xrightarrow{\;P\;} & \Lambda^{k-1}(U \times \mathbb{R}) \\ \Big\downarrow \Phi^* & & \Big\downarrow \Phi^* \\ \Lambda^k(V \times \mathbb{R}) & \xrightarrow[\;P\;]{} & \Lambda^{k-1}(V \times \mathbb{R}) \end{array}$$

To prove (5.2.6) we observe first that it is trivial if $k = 0$ so we assume $k \geq 1$. Next observe that dt is a 1-form on both $U \times$ and $V \times$ and

$$\mathbf{\Phi}^*(dt) = d(\mathbf{\Phi}^* t) = d(t \circ \mathbf{\Phi}) = dt$$

because $\mathbf{\Phi} = \phi \times id$. On the other hand,

$$\begin{aligned} \mathbf{\Phi}^*(dx^i) &= d(\mathbf{\Phi}^* x^i) = d(x^i \circ \mathbf{\Phi}) = d(x^i \circ \phi) \\ &= d\phi^i = \frac{\partial \phi^i}{\partial x^j}\, dx^j. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbf{\Phi}^*(dx^{i_1} \wedge \cdots \wedge dx^{i_k}) &= (\mathbf{\Phi}^*(dx^{i_1})) \wedge \cdots \wedge (\mathbf{\Phi}^*(dx^{i_k})) \\ &= \left(\frac{\partial \phi^{i_1}}{\partial x^{j_1}}\, dx^{j_1}\right) \wedge \cdots \wedge \left(\frac{\partial \phi^{i_k}}{\partial x^{j_k}}\, dx^{j_k}\right) \end{aligned}$$

so each $\mathbf{\Phi}^*(dx^I)$ can be uniquely expressed as

$$\mathbf{\Phi}^*(dx^I) = \sum_K \gamma_K(x) dx^K,$$

where K varies over increasing index sets of length k and the functions γ_K do not depend on t. An analogous statement is true of each $\mathbf{\Phi}^*(dx^J)$. Now, for $\boldsymbol{\omega} \in \Lambda^k(U \times \quad)$,

$$\begin{aligned} \mathbf{\Phi}^*\boldsymbol{\omega}(x,t) &= \mathbf{\Phi}^*\left(\sum_I \omega_I\,(x,t)\,dt \wedge dx^I + \sum_J \omega_J\,(x,t)\,dx^J\right) \\ &= \sum_I \omega_I\,(x,t)\mathbf{\Phi}^*\Big(dt \wedge dx^I\Big) + \sum_J \omega_J\,(x,t)\,\mathbf{\Phi}^*\Big(dx^J\Big) \\ &= \sum_I \omega_I\,(x,t)\,dt \wedge \mathbf{\Phi}^*\Big(dx^I\Big) + \sum_J \omega_J\,(x,t)\,\mathbf{\Phi}^*\Big(dx^J\Big). \end{aligned}$$

Exercise 5.2.1 Show that

$$P(\mathbf{\Phi}^* \boldsymbol{\omega}) = \sum_I \left[\int_a^t \omega_I\,(x,s) ds \right] \mathbf{\Phi}^*\,(dx^I) = \mathbf{\Phi}^*\,(P\boldsymbol{\omega}).$$

The fact that P, defined for open subsets of $\;^n$, commutes with pullbacks by diffeomorphisms of the type $\mathbf{\Phi} = \phi \times id\;$ permits us to define an analogous operation on manifolds. More precisely, we will prove the following: Let X be a smooth n-manifold. Then, for each $k = 0, 1, \ldots$ there exists a linear map

$$P : \Lambda^k\,(X \times \quad) \longrightarrow \Lambda^{k-1}(X \times \quad)$$

such that

1. If $X = U$, an open submanifold of $\;^n$, then P is given by (5.2.5).
2. If Y is another smooth n-manifold, $\phi : Y \to X$ is a diffeomorphism and $\mathbf{\Phi} = \phi \times id\;\; : Y \times \quad \to X \times \quad$, then
$$\mathbf{\Phi}^* \circ P = P \circ \mathbf{\Phi}^*. \tag{5.2.7}$$

To prove this we first take P to be identically zero when $k = 0$. Now let $\boldsymbol{\omega} \in \Lambda^k(X \times \quad)$, where $k \geq 1$. We will define $P\boldsymbol{\omega}$ locally in coordinates and then show that the definition is independent of the choice of coordinates. Let (U, φ) be a chart on X. Then $\varphi^{-1} : \varphi(U) \to U$ is a diffeomorphism and therefore so is $\varphi^{-1} \times id\;\; : \varphi(U) \times \quad \to U \times \quad$. Thus, $(\varphi^{-1} \times id\;\;)^* \boldsymbol{\omega}$ is a k-form on $\varphi(U) \times \quad \subseteq \quad^n \times \quad$ (technically we should write $\iota^* \boldsymbol{\omega}$ rather than $\boldsymbol{\omega}$, where $\iota : U \times \quad \hookrightarrow X \times \quad$, but we will suppress this inclusion). Thus, $P((\varphi^{-1} \times id\;\;)^* \boldsymbol{\omega})$ as defined above in Euclidean spaces, is a $(k-1)$-form on $\varphi(U) \times \quad$. We may therefore define $P\boldsymbol{\omega}$ on U by

$$P\boldsymbol{\omega} = (\varphi \times id\;\;)^* \Big(P\Big(\left(\varphi^{-1} \times id\;\;\right)^* \boldsymbol{\omega} \Big) \Big).$$

Now, let (V, ψ) be another chart on X with $U \cap V \neq \emptyset$. To prove that $P\boldsymbol{\omega}$ is a well-defined $(k-1)$-form on $X \times \quad$ we must show that, on $U \cap V$,

$$(\psi \times id\;\;)^* \Big(P\Big(\Big(\psi^{-1} \times id\;\;\Big)^* \boldsymbol{\omega} \Big) \Big) = (\varphi \times id\;\;)^* \Big(P\Big(\Big(\varphi^{-1} \times id\;\;\Big)^* \boldsymbol{\omega} \Big) \Big).$$

Define $\phi : \psi(U \cap V) \to \varphi(U \cap V)$ by $\phi = \varphi \circ \psi^{-1}$ and let

$$\mathbf{\Phi} = \phi \times id\;\; : \psi(U \cap V) \times \quad \longrightarrow \varphi(U \cap V) \times \quad .$$

Exercise 5.2.2 Show that

$$\mathbf{\Phi}^* \Big(P\Big(\left(\varphi^{-1} \times id\;\;\right)^* \boldsymbol{\omega} \Big) \Big) = P\Big(\left(\psi \times id\;\;\right)^* \boldsymbol{\omega} \Big).$$

Thus,

$$(\psi \times id\ )^* \Big(P\Big(\big(\psi^{-1} \times id\ \big)^* \boldsymbol{\omega} \Big) \Big)$$

$$= (\psi \times id\ )^* \Big(\boldsymbol{\Phi}^* \Big(P\Big((\varphi^{-1} \times id\ )^* \boldsymbol{\omega} \Big) \Big) \Big)$$

$$= \Big(\boldsymbol{\Phi} \circ (\psi \times id\ ) \Big)^* \Big(P\Big(\big(\varphi^{-1} \times id\ \big)^* \boldsymbol{\omega} \Big) \Big)$$

$$= (\varphi \times\ id\ )^* \Big(P\Big(\big(\varphi^{-1} \times id\ \big)^* \boldsymbol{\omega} \Big) \Big)$$

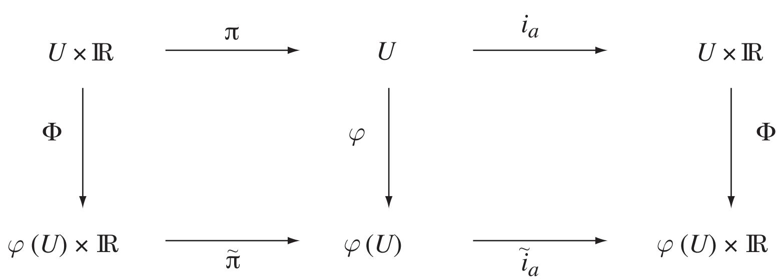
because $\boldsymbol{\Phi} \circ (\psi \times id\ ) = (\phi \times id\ ) \circ (\psi \times id\ ) = (\phi \circ \psi) \times id\ = \varphi \times id\ $. We conclude then that P is well-defined on X. The fact that P agrees with (5.2.5) when X is an open submanifold of ^{n} is obvious from its definition. Property #2 on page 308 is also clear since it is an equality between two $(k-1)$-forms on $Y \times$ which can be verified locally, in coordinates, using (5.2.6). We put the operator P to use in our next result.

Theorem 5.2.1 *Let $\pi : X \times \ \to X$ be the projection and, for any $a \in$, let $i_a : X \to X \times$ be the embedding $i_a(x) = (x, a)$. Then, for any form $\boldsymbol{\omega}$ on $X \times$,*

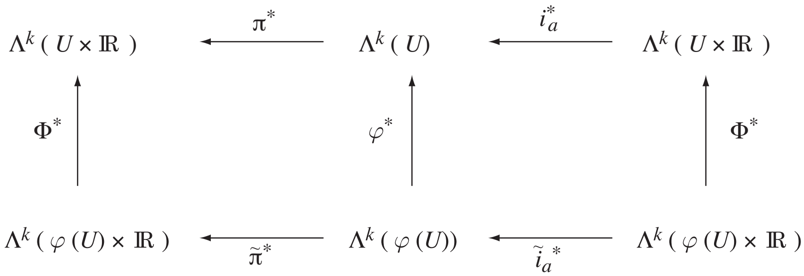
$$d \circ P(\boldsymbol{\omega}) - P \circ d(\boldsymbol{\omega}) = \boldsymbol{\omega} - (i_a \circ \pi)^* \boldsymbol{\omega}, \tag{5.2.8}$$

where $P : \Lambda^k(X \times \ ) \to \Lambda^{k-1}(X \times \ )$ is the operator defined as above using the local coordinate operator (5.2.5) (note the presence of a in (5.2.5)).

Proof: First we observe that it is enough to prove the result when X is an open subset of a Euclidean space. Indeed, suppose this has been proved. Let $\varphi : U \to \varphi(U) \subseteq \ ^n$ be a chart on X and set $\boldsymbol{\Phi} = \varphi \times id\ $. Then $\boldsymbol{\Phi}^* : \Lambda^k(\varphi(U) \times \ ) \to \Lambda^k(U \times \ )$ for each k. With $\pi : U \times \ \to U$ the projection and $i_a : U \to U \times$ the embedding we define $\tilde{\pi}$ and $\tilde{i}_a$ by the following commutative diagram:



For each k this induces a commutative diagram



For any $\tilde{\boldsymbol{\omega}} \in \Lambda^k(\varphi(U) \times \quad)$,

$$\pi^*(i_a^*(\boldsymbol{\Phi}^*\tilde{\boldsymbol{\omega}})) = \boldsymbol{\Phi}^*\Big(\tilde{\pi}^*(\tilde{i}_a{}^*\tilde{\boldsymbol{\omega}})\Big),$$

i.e.,

$$(i_a \circ \pi)^*(\boldsymbol{\Phi}^*\tilde{\boldsymbol{\omega}}) = \boldsymbol{\Phi}^*\Big((\tilde{i}_a \circ \tilde{\pi})^*\tilde{\boldsymbol{\omega}}\Big).$$

Now let $\boldsymbol{\omega} \in \Lambda^k(X \times \quad)$. On U, $\boldsymbol{\omega} = \boldsymbol{\Phi}^*\tilde{\boldsymbol{\omega}}$ for some $\tilde{\boldsymbol{\omega}} \in \Lambda^k(\varphi(U) \times \quad)$. Since d commutes with any pullback and P commutes with $\boldsymbol{\Phi}^*$,

$$\begin{aligned} d \circ P(\boldsymbol{\omega}) + P \circ d(\boldsymbol{\omega}) &= d \circ P(\boldsymbol{\Phi}^*\tilde{\boldsymbol{\omega}}) + P \circ d(\boldsymbol{\Phi}^*\tilde{\boldsymbol{\omega}}) \\ &= \boldsymbol{\Phi}^*(d \circ P(\boldsymbol{\omega}) + P \circ d(\boldsymbol{\omega})). \end{aligned}$$

Moreover,

$$\begin{aligned} \boldsymbol{\omega} - (i_a \circ \pi)^*\boldsymbol{\omega} &= \boldsymbol{\Phi}^*(\tilde{\boldsymbol{\omega}}) - (i_a \circ \pi)^*(\boldsymbol{\Phi}^*\tilde{\boldsymbol{\omega}}) \\ &= \boldsymbol{\Phi}^*(\tilde{\boldsymbol{\omega}} - (\tilde{i}_a \circ \tilde{\pi})^*\tilde{\boldsymbol{\omega}}). \end{aligned}$$

Thus, assuming we have proved that $d \circ P(\tilde{\boldsymbol{\omega}}) + P \circ d(\tilde{\boldsymbol{\omega}}) = \tilde{\boldsymbol{\omega}} - (\tilde{i}_a \circ \tilde{\pi})^*\tilde{\boldsymbol{\omega}}$, the result for X will follow on U and therefore on all of X.

Thus, we need only prove the result when $X = U$ is an open submanifold of $\ ^n$ (we will drop all of the tildas in the notation used above). Any $\boldsymbol{\omega} \in \Lambda^k(U \times \quad)$ has a unique representation in standard coordinates $(x,t) = (x^1, \ldots, x^n, t)$ of the form

$$\boldsymbol{\omega}(x,t) = \sum_I \omega_I(x,t)\, dt \wedge dx^I + \sum_J \omega_J(x,t)\, dx^J.$$

Then $P\boldsymbol{\omega}$ is given by (5.2.5). Since both sides of (5.2.8) are linear in $\boldsymbol{\omega}$ we need only consider the following two cases separately:

1. $\boldsymbol{\omega} = \omega_J(x,t)\, dx^J$.
2. $\boldsymbol{\omega} = \omega_I(x,t)\, dt \wedge dx^I$.

In the first case, $\boldsymbol{\omega} = \omega_J\,(x,t)\,dx^J$ implies $P\boldsymbol{\omega} = 0$ so $d \circ P(\boldsymbol{\omega}) = 0$. Moreover, $d\boldsymbol{\omega}$ is the sum of two terms, one of which contains no dt and therefore does not contribute to $P \circ d(\boldsymbol{\omega})$. The other term is $\frac{\partial \omega_J}{\partial t}\,dt \wedge dx^J$ so

$$\begin{aligned} P \circ d\,(\boldsymbol{\omega}) &= \left[\int_a^t \frac{\partial \omega_J}{\partial s}\,(x,s)\,ds\right] dx^J = [\omega_J\,(x,t) - \omega_J(x,a)]\,dx^J \\ &= \boldsymbol{\omega} - \omega_J\,(x,a)\,dx^J. \end{aligned}$$

Next observe that

$$\begin{aligned} (i_a \circ \pi)^*\boldsymbol{\omega} &= (i_a \circ \pi)^*(\omega_J\,(x,t)\,dx^J) \\ &= (i_a \circ \pi)^*(\omega_J\,(x,t)\,dx^{j_1} \wedge \cdots \wedge dx^{j_k}) \\ &= (i_a \circ \pi)^*(\omega_J\,(x,t))(i_a \circ \pi)^*(dx^{j_1}) \wedge \cdots \wedge (i_a \circ \pi)^*(dx^{j_k}). \end{aligned}$$

But

$$(i_a \circ \pi)^*(\omega_J\,(x,t)) = \omega_J \circ (i_a \circ \pi)(x,t) = \omega_J\,(x,a)$$

and

$$(i_a \circ \pi)^*(dx^j) = d\,(x^j \circ (i_a \circ \pi)) = dx^j$$

so

$$(i_a \circ \pi)^*\boldsymbol{\omega} = \omega_J\,(x,a)\,dx^J.$$

Thus,

$$d \circ P\,(\boldsymbol{\omega}) + P \circ d\,(\boldsymbol{\omega}) = 0 + \boldsymbol{\omega} - (i_a \circ \pi)^*\boldsymbol{\omega}$$

as required.

For the second case one has, as above, $(i_a \circ \pi)^*(\omega_I\,(x,t)) = \omega_I\,(x,a)$ and $(i_a \circ \pi)^*(dx^i) = dx^i$, but now

$$(i_a \circ \pi)^*(dt) = d(t \circ (i_a \circ \pi)) = d(t(x,a)) = d(a) = 0$$

so

$$(i_a \circ \pi)^*\boldsymbol{\omega} = 0.$$

Next,

$$P\boldsymbol{\omega} = \left[\int_a^t \omega_I\,(x,s)\,ds\right] dx^I$$

so

$$\begin{aligned} d \circ P(\boldsymbol{\omega}) &= \left[\left(\frac{\partial}{\partial x^\alpha}\int_a^t \omega_I\,(x,s)\,ds\right) dx^\alpha + \left(\frac{\partial}{\partial t}\int_a^t \omega_I\,(x,s)\,ds\right) dt\right] \wedge dx^I \\ &= \left[\int_a^t \frac{\partial \omega_I}{\partial x^\alpha}\,(x,s)\,ds\right] dx^\alpha \wedge dx^I + \omega_I\,(x,t)\,dt \wedge dx^I \\ &= \left[\int_a^t \frac{\partial \omega_I}{\partial x^\alpha}\,(x,s)\,ds\right] dx^\alpha \wedge dx^I + \boldsymbol{\omega}. \end{aligned}$$

Exercise 5.2.3 Show that $d\boldsymbol{\omega} = -\frac{\partial \omega_I}{\partial x^\alpha}(x,t)dt \wedge dx^\alpha \wedge dx^I$ and conclude that

$$d \circ P(\boldsymbol{\omega}) + P \circ d(\boldsymbol{\omega}) = \boldsymbol{\omega} - (i_a \circ \pi)^* \boldsymbol{\omega}. \qquad \blacksquare$$

Corollary 5.2.2 *Let $\pi : X \times \quad \to X$ be the projection and, for any $a \in \quad$, let $i_a : X \to X \times \quad$ be the embedding $i_a(x) = (x,a)$. Then the induced maps*

$$i_a^{\#} : H^k_{\text{de R}}(X \times \quad) \longrightarrow H^k_{\text{de R}}(X)$$

and

$$\pi^{\#} : H^k_{\text{de R}}(X) \longrightarrow H^k_{\text{de R}}(X \times \quad)$$

are inverses of each other for every $k = 0, 1, 2, \ldots$. In particular,

$$H^k_{\text{de R}}(X \times \quad) \cong H^k_{\text{de R}}(X).$$

Proof: Since $\pi \circ i_a = id_X$, $i_a{}^{\#} \circ \pi^{\#} = id_{H^k_{\text{de R}}(X)}$ for any k. Now, $i_a \circ \pi$ is not the identity, but we show that it nevertheless induces the identity in cohomology. Letting P denote the operator of Theorem 5.2.1 we have

$$d(P\boldsymbol{\omega}) + P(d\,\boldsymbol{\omega}) = \boldsymbol{\omega} - (i_a \circ \pi)^* \boldsymbol{\omega}.$$

If $\boldsymbol{\omega}$ is closed, this gives

$$\boldsymbol{\omega} - (i_a \circ \pi)^* \boldsymbol{\omega} = d(P\boldsymbol{\omega})$$

so $\boldsymbol{\omega} - (i_a \circ \pi)^* \boldsymbol{\omega}$ is exact. Thus, $[\boldsymbol{\omega}] = [(i_a \circ \pi)^* \boldsymbol{\omega}] = (i_a \circ \pi)^{\#}([\boldsymbol{\omega}])$ so $(i_a \circ \pi)^{\#} = id_{H^k_{\text{de R}}(X \times \quad)}$, i.e., $\pi^{\#} \circ i_a{}^{\#} = id_{H^k_{\text{de R}}(X \times \quad)}$ and the result follows. $\blacksquare$

Remark: Note that it follows from Corollary 5.2.2 that, if a and b are any two real numbers, then

$$i_a{}^{\#} = i_b{}^{\#}.$$

Corollary 5.2.3 *Smoothly homotopic maps induce the same maps in cohomology.*

Proof: Let X and Y be manifolds, $f, g : X \to Y$ smooth maps and $F : X \times \quad \to Y$ a smooth map with $F(x,0) = f(x)$ and $F(x,1) = g(x)$ for all $x \in X$. If $i_0, i_1 : X \to X \times \quad$ are the embeddings $i_0(x) = (x,0)$ and $i_1(x) = (x,1)$, then $f = F \circ i_0$ and $g = F \circ i_1$. Thus, $f^{\#} = i_0{}^{\#} \circ F^{\#}$ and $g^{\#} = i_1{}^{\#} \circ F^{\#}$. But $i_0{}^{\#} = i_1{}^{\#}$ so the result follows. $\blacksquare$

We will say that two manifolds X and Y are of the same **smooth homotopy type** if there exist smooth maps $h : X \to Y$ and $h' : Y \to X$ such that $h \circ h'$ and $h' \circ h$ are smoothly homotopic to id_Y and id_X, respectively. In this case we conclude from Corollary 5.2.3 that $(h')^{\#} \circ h^{\#} = id_{H^k_{\text{de R}}(Y)}$ and

$h^{\#} \circ (h')^{\#} = id_{H^k_{\mathrm{de\,R}}(X)}$ for any k so $h^{\#}$ and $(h')^{\#}$ are inverse isomorphisms. In particular, we have the following very useful computational device.

Corollary 5.2.4 *Two manifolds of the same smooth homotopy type have the same de Rham cohomology groups.*

Remark: From the Remarks following Corollary 3.2.3 it follows that the de Rham cohomology groups are homotopy invariants (and, in particular, homeomorphism invariants).

A manifold X is **smoothly contractible** if the identity map id_X is smoothly homotopic to a constant map from X to X.

Exercise 5.2.4 Show that X is smoothly contractible if and only if it has the same smooth homotopy type as a point (connected, 0-dimensional manifold). **Hint**: The corresponding topological result is Theorem 2.3.7 of [**N4**].

Corollary 5.2.5 *A smoothly contractible manifold X has trivial de Rham cohomology groups $H^k_{\mathrm{de\,R}}(X)$ for all $k \geq 1$.*

Let X be a manifold and A a smooth submanifold of X. Then the inclusion map $\iota : A \hookrightarrow X$ is smooth (in fact, an embedding, by Lemma 5.6.1 of [**N4**]). A smooth map $r : X \to A$ is called a **smooth retraction** of X onto A if $r \circ \iota = id_A$. In this case, $\iota^{\#} \circ r^{\#} = id_{H^k_{\mathrm{de\,R}}(A)}$ for each k. A smooth retraction r is called a **smooth deformation retraction** if $\iota \circ r : X \to A$ is smoothly homotopic to id_X. In this case, A and X obviously have the same smooth homotopy type and therefore isomorphic cohomology groups, but more is true.

Exercise 5.2.5 Show that, if $r : X \to A$ is a smooth deformation retraction of X onto A, then $\iota^{\#}$ and $r^{\#}$ are inverse isomorphisms for each k.

Exercise 5.2.6 Show that there is a smooth deformation retraction of $^2 - \{ (0,0) \}$ onto S^1 and then describe all of the de Rham cohomology groups of $^2 - \{ (0,0) \}$.

Corollary 5.2.2 generalizes easily to the case in which is replaced by any smoothly contractible manifold. Specifically, we suppose X and Y are manifolds with Y smoothly contractible. Let $F : Y \times \;\; \to Y$ be a smooth map with $F(y,0) = y$ and $F(y,1) = y_0 \in Y$ for every $y \in Y$. We also let $\pi : X \times Y \to X$ be the projection and $i_{y_0}(x) = (x, y_0)$. Then $\pi \circ i_{y_0} = id_X$.

Exercise 5.2.7 Show that $i_{y_0} \circ \pi$ is smoothly homotopic to $id_{X \times Y}$. **Hint**: Define $H : (X \times Y) \times \;\; \to X \times Y$ by $H((x,y),t) = (x, F(y,t))$.

It follows that $\pi^{\#} : H^k_{\mathrm{de\,R}}(X) \to H^k_{\mathrm{de\,R}}(X \times Y)$ and ${i_{y_0}}^{\#} : H^k_{\mathrm{de\,R}}(X \times Y) \to H^k_{\mathrm{de\,R}}(X)$ are inverse isomorphisms and, in particular, we have the following result.

Theorem 5.2.6 *Let X and Y be smooth manifolds with Y smoothly contractible. Then $H^k_{\mathrm{de\,R}}(X \times Y) \cong H^k_{\mathrm{de\,R}}(X)$ for every k.*

One more result along these lines will be of use. We wish to consider a smooth vector bundle over X. More precisely, we begin with a smooth principal G-bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ over X and a representation $\rho : G \to GL(\mathcal{V})$ of G on some finite dimensional vector space $\mathcal{V}$. Now consider the associated vector bundle

$$\mathcal{P}_\rho : P \times_\rho \mathcal{V} \longrightarrow X.$$

We will show that $P \times_\rho \mathcal{V}$ and X have the same de Rham cohomology groups. For this recall that each fiber $\mathcal{P}_\rho^{-1}(x) = \{\, [p, v] : v \in \mathcal{V} \,\}$, where p is any point in $\mathcal{P}^{-1}(x)$, has a natural vector space structure under which it is isomorphic to $\mathcal{V}$. In particular, each ${\mathcal{P}_\rho}^{-1}(x)$, has a copy $[p, 0]$ of the zero element in $\mathcal{V}$ and we may define a map $\sigma_0 : X \to P \times_\rho \mathcal{V}$ by $\sigma_0(x) = [p, 0]$.

Exercise 5.2.8 Show that σ_0 is a smooth global cross-section of $P \times_\rho \mathcal{V}$.

The cross-section σ_0 is called the **zero section** of $P \times_\rho \mathcal{V}$ and, because it is a cross-section, it satisfies $\mathcal{P}_\rho \circ \sigma_0 = id_X$. Defining $F : (P \times_\rho \mathcal{V}) \times \quad \to P \times_\rho \mathcal{V}$ by $F([p, v], t) = [p, tv]$ we obtain a smooth map with $F([p, v], 0) = [p, 0] = \sigma_0 \circ \mathcal{P}_\rho([p, v])$ and $F([p, v], 1) = [p, v] = id_{P \times_\rho \mathcal{V}}([p, v])$. Thus, $\sigma_0 \circ \mathcal{P}_\rho$ is homotopic to $id_{P \times_\rho \mathcal{V}}$. We conclude that ${\sigma_0}^\# : H^k_{\mathrm{de\,R}}(P \times_\rho \mathcal{V}) \to H^k_{\mathrm{de\,R}}(X)$ and ${P_\rho}^\# : H^k_{\mathrm{de\,R}}(X) \to H^k_{\mathrm{de\,R}}(P \times_\rho \mathcal{V})$ are inverse isomorphisms.

Theorem 5.2.7 *A smooth vector bundle and its base manifold have the same de Rham cohomology groups.*

The situation is very different for principal bundles, however, as we shall see in Section 5.4.

Corollary 5.2.4 is an important tool in the calculation of cohomology groups, but its effective use requires a means of putting together the cohomology of a manifold X from that of two open submanifolds of which it is the union. For example, $S^2 = U_N \cup U_S$, where U_N and U_S are S^2 minus the south and north poles, respectively. Both U_N and U_S are diffeomorphic to $\ ^2$ and so have trivial cohomology. Moreover, $U_N \cap U_S$ is diffeomorphic to the punctured plane $\ ^2 - \{\, (0,0) \,\}$ and so we know its cohomology as well (because it has the same homotopy type as S^1). What's needed then is some relationship between $H^k_{\mathrm{de\,R}}(U_N \cup U_S)$, $H^k_{\mathrm{de\,R}}(U_N)$, $H^k_{\mathrm{de\,R}}(U_S)$ and $H^k_{\mathrm{de\,R}}(U_N \cap U_S)$. This is provided by the "Mayer-Vietoris sequence" which is our next major objective. We take one preliminary step to conclude this section, an important algebraic detour in the next and finally prove the theorem in Section 5.4. This will all seem quite familiar to those who have been through the proof of Mayer-Vietoris for singular homology in Section 3.5 of [**N4**], but the argument here is much simpler and one should keep an eye out to spot the precise point at which one can account for this relative simplicity.

We consider then a smooth manifold $X = U \cup V$, where U and V are open submanifolds of X. Introduce the inclusions

$$\begin{aligned} i_U &: U \hookrightarrow X \\ i_V &: V \hookrightarrow X \\ j_U &: U \cap V \hookrightarrow U \\ j_V &: U \cap V \hookrightarrow V \end{aligned}$$

and, for each k, their pullback (restriction) maps

$$\begin{aligned} {i_U}^* &: \Lambda^k(X) \longrightarrow \Lambda^k(U) \\ {i_V}^* &: \Lambda^k(X) \longrightarrow \Lambda^k(V) \\ {j_U}^* &: \Lambda^k(U) \longrightarrow \Lambda^k(U \cap V) \\ {j_V}^* &: \Lambda^k(V) \longrightarrow \Lambda^k(U \cap V). \end{aligned}$$

We also define

$$\alpha^k = {i_U}^* \oplus {i_V}^* : \Lambda^k(X) \longrightarrow \Lambda^k(U) \oplus \Lambda^k(V)$$

by

$$\alpha^k(\boldsymbol{\omega}) = ({i_U}^*\boldsymbol{\omega},\ {i_V}^*\boldsymbol{\omega})$$

and

$$\beta^k = {j_U}^* - {j_V}^* : \Lambda^k(U) \oplus \Lambda^k(V) \longrightarrow \Lambda^k(U \cap V)$$

by

$$\beta^k(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) = {j_U}^*\boldsymbol{\lambda}_1 - {j_V}^*\boldsymbol{\lambda}_2.$$

Thus, for each k, we have a sequence

$$0 \longrightarrow \Lambda^k(X) \xrightarrow{\alpha^k} \Lambda^k(U) \oplus \Lambda^k(V) \xrightarrow{\beta^k} \Lambda^k(U \cap V) \longrightarrow 0 \tag{5.2.9}$$

which we claim is exact (i.e., the image of each map is the kernel of the next). Exactness at $\Lambda^k(X)$ is the statement that α^k is one-to-one. To see this suppose $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are distinct elements of $\Lambda^k(X)$. Since $X = U \cup V$, there is a p in U or in V (or in both) at which $\boldsymbol{\omega}_1(p) \neq \boldsymbol{\omega}_2(p)$. Since U and V are open in X, the tangent space at p to either U or V coincides with $T_p(X)$ so either ${i_U}^*\boldsymbol{\omega}_1(p) \neq {i_U}^*\boldsymbol{\omega}_2(p)$ if $p \in U$ or ${i_V}^*\boldsymbol{\omega}_1(p) \neq {i_V}^*\boldsymbol{\omega}_2(p)$ if $p \in V$. In either case, $\alpha^k(\boldsymbol{\omega}_1) \neq \alpha^k(\boldsymbol{\omega}_2)$.

To prove exactness at $\Lambda^k(U) \oplus \Lambda^k(V)$ we must show that Image $(\alpha^k) = \ker(\beta^k)$ and for this we prove containment in each direction. Since $\alpha^k(\boldsymbol{\omega}) = ({i_U}^*\boldsymbol{\omega}, {i_V}^*\boldsymbol{\omega})$, $\beta^k \circ \alpha^k(\boldsymbol{\omega}) = {j_U}^*({i_U}^*\boldsymbol{\omega}) - {j_V}^*({i_V}^*\boldsymbol{\omega}) = 0$ because both terms are the restrictions of $\boldsymbol{\omega}$ to $U \cap V$. Thus, Image $(\alpha^k) \subseteq \ker(\beta^k)$. Next we show that $\ker(\beta^k) \subseteq$ Image (α^k). Suppose $\beta^k(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = 0$. Then ${j_U}^*(\boldsymbol{\lambda}_1) = {j_V}^*(\boldsymbol{\lambda}_2)$ so $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ agree on $U \cap V$. Thus, we can define $\boldsymbol{\omega}$ on $X = U \cup V$ by taking ${i_U}^*\boldsymbol{\omega} = \boldsymbol{\lambda}_1$ and ${i_V}^*\boldsymbol{\omega} = \boldsymbol{\lambda}_2$. Then $\alpha^k(\boldsymbol{\omega}) = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ so $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \in$ Image (α^k).

Exactness at $\Lambda^k(U \cap V)$ is the statement that β^k is onto. To see this we let $\{\,\phi_U, \phi_V\,\}$ be a family of functions of the sort guaranteed by Corollary 3.1.5 for the open cover $\{\,U, V\,\}$ of X. Next let $\boldsymbol{\omega} \in \Lambda^k(U \cap V)$ be arbitrary. Then $\phi_V \boldsymbol{\omega} \in \Lambda^k(U \cap V)$ and, since $\operatorname{supp} \phi_V \subseteq V$ we can define $\boldsymbol{\lambda}_1 \in \Lambda^k(U)$ by

$$\boldsymbol{\lambda}_1 = \begin{cases} \phi_V\, \boldsymbol{\omega} & \text{on } U \cap V \\ 0 & \text{on } U - (U \cap V) \end{cases}.$$

Similarly, define $\boldsymbol{\lambda}_2 \in \Lambda^k(V)$ to be $-\phi_U\, \boldsymbol{\omega}$ on $U \cap V$ and 0 on $V - (U \cap V)$. Then

$$\begin{aligned} \beta^k(\boldsymbol{\lambda}_1,\, \boldsymbol{\lambda}_2) &= {j_U}^* \boldsymbol{\lambda}_1 - {j_V}^* \boldsymbol{\lambda}_2 = \phi_V\, \boldsymbol{\omega} - (-\phi_U\, \boldsymbol{\omega}) \\ &= (\phi_V + \phi_U)\, \boldsymbol{\omega} = \boldsymbol{\omega}. \end{aligned}$$

so $\boldsymbol{\omega} \in \operatorname{Image}(\beta^k)$ as required.

Thus, we have an exact sequence (5.2.9) for each k. Letting d denote the exterior differentiation operator on all of the spaces of forms ($\Lambda^k(X)$, $\Lambda^k(U)$, $\Lambda^k(V)$ or $\Lambda^k(U \cap V)$ for any k) we define

$$d \oplus d : \Lambda^k(U) \oplus \Lambda^k(V) \longrightarrow \Lambda^{k+1}(U) \oplus \Lambda^{k+1}(V)$$

by $(d \oplus d)(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = (d\boldsymbol{\lambda}_1, d\boldsymbol{\lambda}_2)$ for any k and consider the diagram

$$\begin{array}{ccccccccc} & & \vdots & & \vdots & & \vdots & & \\ & & \downarrow d & & \downarrow d \oplus d & & \downarrow d & & \\ 0 & \longrightarrow & \Lambda^k(X) & \xrightarrow{\alpha^k} & \Lambda^k(U) \oplus \Lambda^k(V) & \xrightarrow{\beta^k} & \Lambda^k(U \cap V) & \longrightarrow & 0 \\ & & \downarrow d & & \downarrow d \oplus d & & \downarrow d & & \\ 0 & \longrightarrow & \Lambda^{k+1}(X) & \xrightarrow{\alpha^{k+1}} & \Lambda^{k+1}(U) \oplus \Lambda^{k+1}(V) & \xrightarrow{\beta^{k+1}} & \Lambda^{k+1}(U \cap V) & \longrightarrow & 0 \\ & & \downarrow d & & \downarrow d \oplus d & & \downarrow d & & \\ & & \vdots & & \vdots & & \vdots & & \end{array}$$

We have just shown that the diagram has exact rows. The columns, although not exact, in general, compose to zero at each stage because $d^2 = 0$.

Exercise 5.2.9 Show that each square in this diagram is commutative, i.e., that, for each k,

$$(d \oplus d) \circ \alpha^k = \alpha^k \circ d \;\text{ and }\; d \circ \beta^k = \beta^k \circ (d \oplus d). \tag{5.2.10}$$

The major result toward which we are headed (The Mayer-Vietoris Sequence) is a purely algebraic consequence of the structure of this last diagram. For this reason, and to build a structure large enough to accommodate yet another "cohomology theory" that we will require in Chapter 6, we pause momentarily to introduce a few of the basic notions of cohomological algebra.

5.3 Cochain Complexes and Their Cohomology

We recall that a principal ideal domain is a commutative ring R with unit element e in which there are no zero divisors (i.e., $ab = 0$ implies $a = 0$ or $b = 0$) and in which every ideal I is principal (i.e., $I = Ra = \{ra : r \in R\}$ for some $a \in R$). From our point of view the most important examples are fields (e.g.,) and the ring $\mathbb{Z}$ of integers. A module over R, or R-module, consists of an Abelian group C (whose operation we designate $+$) together with a "scalar multiplication" map $R \times C \to C$ which assigns to each pair $(a, x) \in R \times C$ an element ax of C satisfying $(a + b)x = ax + bx$, $a(x + y) = ax + ay$, $(ab)x = a(bx)$, and $ex = x$ for all $a, b \in R$ and $x, y \in C$. Thus, -modules are simply real vector spaces and $\mathbb{Z}$-modules are Abelian groups. All of the basic notions of linear algebra have analogues for R-modules and we will need just a few. If C_1 and C_2 are R-modules, then a map $f : C_1 \to C_2$ is said to be R-linear (or a homomorphism) if $f(ax + by) = a\,f(x) + b\,f(y)$ for all $a, b \in R$ and $x, y \in C_1$. An additive subgroup C' of the R-module C is called a submodule of C if it is closed under scalar multiplication (and therefore is itself an R-module under the same operations as C). In this case one defines the quotient module C/C' by providing the quotient group with the obvious scalar multiplication: If $[v] = v + C'$ is in C/C', then $a[v] = a(v + C') = av + C' = [av]$. Finally, if C_1 and C_2 are R-modules, their direct sum $C_1 \oplus C_2$ is obtained from the direct sum of the underlying Abelian groups by defining scalar multiplication coordinatewise: $a(x_1, x_2) = (ax_1, ax_2)$, for all $a \in R$, $x_1 \in C_1$ and $x_2 \in C_2$.

Now, a **cochain complex** C^* consists of a sequence of R-modules and homomorphisms

$$\cdots \longrightarrow C^{k-1} \xrightarrow{\delta^{k-1}} C^k \xrightarrow{\delta^k} C^{k+1} \xrightarrow{\delta^{k+1}} \cdots$$

defined for all integers k such that the image of each homomorphism is contained in the kernel of the next, i.e., $\delta^{k+1} \circ \delta^k = 0$ for each k. In the only example we have seen thus far C^k was the -module $\Lambda^k(X)$ of k-forms on a manifold X and δ^k was the exterior differentiation map d^k. The homomorphism $\delta^k : C^k \to C^{k+1}$ is called the $\boldsymbol{k^{\mathrm{th}}}$ **coboundary operator** and, when it is convenient and no confusion will result, will be denoted simply δ. Image (δ^{k-1}) is a submodule of C^k and its elements are called $\boldsymbol{k}$**-coboundaries**. ker (δ^k) is a submodule of C^k and its elements are called $\boldsymbol{k}$**-cocycles**. Since $\delta^k \circ \delta^{k-1} = 0$, Image $(\delta^{k-1}) \subseteq \ker(\delta^k)$ so we may form the quotient module

$$H^k(C^*) = \ker(\delta^k) \,/\, \text{Image}\,(\delta^{k-1}).$$

$H^k(C^*)$ is called the $\boldsymbol{k}^{\textbf{th}}$ **cohomology group** of the complex C^* (even though k^{th} cohomology *module* would be more appropriate). The elements of $H^k(C^*)$ are equivalence classes $[x]$ (called **cohomology classes**), where x is a k-cocycle and the equivalence relation is defined as follows: If $x, x' \in \ker(\delta^k)$, then $x' \sim x$ if and only if $x' = x + y$, where $y \in \text{Image}\,(\delta^{k-1})$ is a k-coboundary (we then say that x' and x are **cohomologous**).

If C_1^* and C_2^* are two cochain complexes of R-modules (with connecting homomorphisms δ_1^k and δ_2^k, respectively), then a **cochain map**

$$\alpha : C_1^* \longrightarrow C_2^*$$

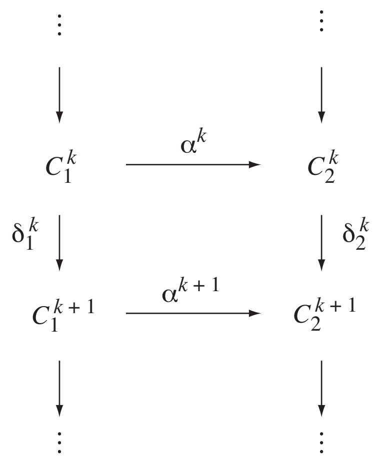
from C_1^* to C_2^* is a sequence of R-linear maps

$$\alpha^k : C_1^k \longrightarrow C_2^k, \quad k = 0, \pm 1, \ldots$$

such that

$$\delta_2^k \circ \alpha^k = \alpha^{k+1} \circ \delta_1^k$$

for each k.



Exercise 5.3.1 Show that $\alpha^k(\ker(\delta_1^k)) \subseteq \ker(\delta_2^k)$ and α^k (Image $(\delta_1^{k-1})) \subseteq$ Image (δ_2^{k-1}) for each k and conclude that α^k induces a homomorphism

$$(\alpha^k)^\# : H^k(C_1^*) \longrightarrow H^k\,(C_2^*)$$

in cohomology defined by

$$(\alpha^k)^{\#}([x]) = [\alpha^k(x)]$$

for each $x \in \ker(\delta_1^k)$.

Remark: We have seen an example of this phenomenon. If X and Y are smooth manifolds and $\Lambda^*(X)$ and $\Lambda^*(Y)$ denote their cochains of real-valued forms, then any smooth map $f : X \to Y$ induces a pullback cochain map $f^* : \Lambda^*(Y) \to \Lambda^*(X)$ which, in turn, induces the homomorphisms

$$f^{\#} : H^k_{\mathrm{de\,R}}(Y) \longrightarrow H^k_{\mathrm{de\,R}}(X), \quad k = 0, \pm 1, \ldots$$

in de Rham cohomology.

Exercise 5.3.2 Two cochain maps $\alpha : C_1^* \to C_2^*$ and $\beta : C_1^* \to C_2^*$ are said to be **algebraically homotopic** (or **cochain homotopic**) if there exists a family of homomorphisms

$$h^k : C_1^k \longrightarrow C_2^{k-1}, \qquad k = 0, \pm 1, \ldots$$

such that

$$h^{k+1} \circ \delta_2^k + \delta_1^{k-1} \circ h^k = \alpha^k - \beta^k$$

for each k (the family $h = \{ h^k \}$ of maps is then called an **algebraic homotopy**, or **cochain homotopy**, between α and β). Show that, if α and β are algebraically homotopic, then they induce the same maps in cohomology, i.e.,

$$(\alpha^k)^{\#} = (\beta^k)^{\#}$$

for each k. **Hint**: The argument is the same as for de Rham cohomology.

The composition of two cochain maps $C_1^* \to C_2^*$ and $C_2^* \to C_3^*$ is defined by composing each of the homomorphisms $C_1^k \to C_2^k$ and $C_2^k \to C_3^k$ and clearly induces homomorphisms in cohomology that are just the compositions of those induced by the two cochain maps. Denoting by 0 both the trivial R-module and the cochain complex one can form from these we will say that a sequence of cochain maps

$$0 \longrightarrow C_1^* \xrightarrow{\alpha} C_2^* \xrightarrow{\beta} C_3^* \longrightarrow 0$$

forms a **short exact sequence** if, for each k, the sequence of R-modules

$$0 \longrightarrow C_1^k \xrightarrow{\alpha^k} C_2^k \xrightarrow{\beta^k} C_3^k \longrightarrow 0$$

is exact (i.e., the image of each map *equals* the kernel of the next).

$$
\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & C_1^k & \xrightarrow{\alpha^k} & C_2^k & \xrightarrow{\beta^k} & C_3^k & \longrightarrow & 0 \\
 & & \delta_1^k \downarrow & & \delta_2^k \downarrow & & \delta_3^k \downarrow & & \\
0 & \longrightarrow & C_1^{k+1} & \xrightarrow{\alpha^{k+1}} & C_2^{k+1} & \xrightarrow{\beta^{k+1}} & C_3^{k+1} & \longrightarrow & 0 \\
 & & \delta_1^{k+1} \downarrow & & \delta_2^{k+1} \downarrow & & \delta_3^{k+1} \downarrow & & \\
0 & \longrightarrow & C_1^{k+2} & \xrightarrow{\alpha^{k+2}} & C_2^{k+2} & \xrightarrow{\beta^{k+2}} & C_3^{k+2} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & \vdots & & \vdots & & \vdots & &
\end{array}
$$

Now, α and β both induce maps in cohomology, but it need *not* be the case that the sequences

$$0 \longrightarrow H^k(C_1^*) \longrightarrow H^k(C_2^*) \longrightarrow H^k(C_3^*) \longrightarrow 0$$

are exact (we will construct an example in de Rham cohomology shortly). Thus, short exact sequences of cochain maps do not induce short exact sequences in cohomology. They do, however, induce a certain *long* exact sequence of cohomology groups which is the basis for most of our calculations of cohomology.

Theorem 5.3.1 *Let* $0 \to C_1^* \stackrel{\alpha}{\to} C_2^* \stackrel{\beta}{\to} C_3^* \to 0$ *be a short exact sequence of cochain complexes. Then there exist homomorphisms*

$$\partial^k : H^k(C_3^*) \longrightarrow H^{k+1}(C_1^*), \qquad k = 0, \pm 1, \ldots$$

such that the following sequence is exact:

$$\cdots H^{k-1}(C_3^*) \stackrel{\partial^{k-1}}{\longrightarrow} H^k(C_1^*) \stackrel{(\alpha^k)^\#}{\longrightarrow} H^k(C_2^*) \stackrel{(\beta^k)^\#}{\longrightarrow} H^k(C_3^*) \stackrel{\partial^k}{\longrightarrow} H^{k+1}(C_1^*) \longrightarrow \cdots$$

Proof: To define $\partial^k : H^k(C_3^*) \to H^{k+1}(C_1^*)$ we will make two choices and then prove that the definition is independent of those choices and has the required properties.

Any element of $H^k(C_3^*)$ is $[x]$ for some $x \in C_3^k$ with $\delta_3^k(x) = 0$. Since β^k is surjective, there exists a $y \in C_2^k$ with $\beta^k(y) = x$. Now, observe that

$$\beta^{k+1}\Big(\delta_2^k(y)\Big) = \delta_3^k\big(\beta^k(y)\big) = \delta_3^k(x) = 0$$

so

$$\delta_2^k(y) \in \ker(\beta^{k+1}) = \text{Image}(\alpha^{k+1}).$$

Since α^{k+1} is injective, there is a unique $z \in C_1^{k+1}$ such that

$$\alpha^{k+1}(z) = \delta_2^k(y).$$

We claim that $\delta_1^{k+1}(z) = 0$ so z determines a cohomology class $[z] \in H^{k+1}(C_1^*)$. To see this note that, since α^{k+2} is injective, it is enough to show that $\alpha^{k+1}(\delta_1^{k+1}(z)) = 0$. But

$$\alpha^{k+2}\Big(\delta_1^{k+1}(z)\Big) = \delta_2^{k+1}(\alpha^{k+1}(z)) = \delta_2^{k+1}\Big(\delta_2^k(y)\Big) = 0.$$

Thus, we may define

$$\partial^k([x]) = [z].$$

Now we check that this definition of $\partial^k([x])$ does not depend on either of the choices we have made.

1. The representative x of $[x] \in H^k(C_3^*)$.
2. The $y \in C_2^k$ with $\beta^k(y) = x$.

We deal with the second of these first. Suppose then that $y, y' \in C_2^k$ with $\beta^k(y) = \beta^k(y') = x$. Then $\delta_2^k(y)$ and $\delta_2^k(y')$ are both in Image (α^{k+1}) (proved above) so there exist unique elements z and z' of C_1^{k+1} with $\alpha^{k+1}(z) = \delta_2^k(y)$ and $\alpha^{k+1}(z') = \delta_2^k(y')$. We must show that $z - z' = \delta_1^k(w)$ for some $w \in C_1^k$ so that $[z'] = [z]$. But $\beta^k(y - y') = \beta^k(y) - \beta^k(y') = x - x = 0$ implies $y - y' \in \ker(\beta^k) = \text{Image}(\alpha^k)$ so there is a unique $w \in C_1^k$ with $\alpha^k(w) = y - y'$. Now compute

$$\begin{aligned}\alpha^{k+1}\Big(\delta_1^k(w)\Big) &= \delta_2^k(\alpha^k(w)) = \delta_2^k(y - y') \\ &= \delta_2^k(y) - \delta_2^k(y') = \alpha^{k+1}(z) - \alpha^{k+1}(z') \\ &= \alpha^{k+1}(z - z').\end{aligned}$$

But α^{k+1} is injective so $z - z' = \delta_1^k(w)$ as required.

Exercise 5.3.3 Use what has just been proved to show that the map $x \to [z]$ from the cocycles in C_3^k to $H^{k+1}(C_1^*)$ is well-defined and R-linear.

Next we prove #1. Thus, let $x, x' \in C_3^k$ with $\delta_3^k(x) = \delta_3^k(x') = 0$ and $x - x' = \delta_3^{k-1}(w)$ for some $w \in C_3^{k-1}$. Let $w = \beta^{k-1}(v)$ for $v \in C_2^{k-1}$. Then

$$x - x' = \delta_3^{k-1}(w) = \delta_3^{k-1}(\beta^{k-1}(v)) = \beta^k\Big(\delta_2^{k-1}(v)\Big)$$

so $x - x'$ is the image of $y = \delta_2^{k-1}(v) \in C_2^k$ under β^k. Now, $\delta_2^k(y) = \delta_2^k(\delta_2^{k-1}(v)) = 0$ so the unique element of C_1^{k+1} which α^{k+1} maps onto $\delta_2^k(y)$

is 0, i.e., $x - x' \to [0] \in H^{k+1}(C_1^*)$. But, by Exercise 5.3.3, $x - x' \to [x - x'] = [x] - [x']$ so $[x] = [x']$ as required.

We have shown then that the map ∂^k is well-defined and, by Exercise 5.3.3, it is R-linear. All that remains is to show that the long sequence of cohomology groups in the theorem is exact at $H^k(C_1^*)$, $H^k(C_2^*)$ and $H^k(C_3^*)$ for each k. These are all similar so we will prove the last and leave the other two as exercises. Thus, we show that $\text{Image}\,((\beta^k)^\#) \subseteq \ker(\partial^k)$ and $\ker(\partial^k) \subseteq \text{Image}\,((\beta^k)^\#)$. Let $[y] \in H^k(C_2^*)$ and consider $(\beta^k)^\#([y]) \in \text{Image}\,((\beta^k)^\#)$. We must show that $\partial^k((\beta^k)^\#([y])) = [0]$. But $\partial^k((\beta^k)^\#([y])) = \partial^k([\beta^k(y)]) = [z]$, where $\alpha^{k+1}(z) = \delta_2^k(y)$. Now, y is a cocycle in C_2^k so $\delta_2^k(y) = 0$ and therefore $\alpha^{k+1}(z) = 0$. But α^{k+1} is injective so $z = 0$ and the result follows. For the reverse containment we begin with $[x] \in H^k(C_3^*)$ for which $\partial^k([x]) = [0]$. We must show that $[x] = (\beta^k)^\#([y'])$ for some $[y'] \in H^k(C_2^*)$, i.e., $[x] = [\beta^k(y')]$ for some $y' \in C_2^k$ with $\delta_2^k(y') = 0$. Now, $\partial^k([x]) = [0]$ implies $[z] = 0$, where $\alpha^{k+1}(z) = \delta_2^k(y)$ and $\beta^k(y) = x$. But then z is a coboundary so there exists a $u \in C_1^k$ with $\delta_1^k(u) = z$. Thus,

$$\delta_2^k(y) = \alpha^{k+1}(z) = \alpha^{k+1}\Big(\delta_1^k(u)\Big) = \delta_2^k(\alpha^k(u))$$

and so

$$\delta_2^k(y - \alpha^k(u)) = 0.$$

Let $y' = y - \alpha^k(u) \in C_2^k$. Then $\delta_2^k(y') = 0$ and $\beta^k(y') = \beta^k(y - \alpha^k(u)) = \beta^k(y) - \beta^k(\alpha^k(U)) = \beta^k(y) = x$ as required.

Exercise 5.3.4 Give analogous arguments to establish exactness at $H^k(C_1^*)$ and $H^k(C_2^*)$. ∎

With one last bit of algebraic machinery we will be in a position to calculate some de Rham cohomology groups. Let C_1^* and C_2^* be two cochain complexes with connecting homomorphisms δ_1^k and δ_2^k, respectively. The **direct sum** of C_1^* and C_2^* is the cochain complex denoted $C_1^* \oplus C_2^*$ whose R-modules are $C_1^k \oplus C_2^k$ and whose connecting homomorphisms

$$\delta^k = \delta_1^k \oplus \delta_2^k : C_1^k \oplus C_2^k \longrightarrow C_1^{k+1} \oplus C_2^{k+1}$$

are defined by

$$\delta^k(x_1, x_2) = \Big(\delta_1^k \oplus \delta_2^k\Big)(x_1,\ x_2) = \Big(\delta_1^k(x_1),\ \delta_2^k(x_2)\Big).$$

Exercise 5.3.5 Show that $C_1^* \oplus C_2^*$ is, indeed, a cochain complex and that, for each k,

$$H^k(C_1^* \oplus C_2^*) \cong H^k(C_1^*) \oplus H^k(C_2^*).$$

5.4 The Mayer-Vietoris Sequence

We now apply Theorem 5.3.1 to the following situation in de Rham cohomology: Let X be a smooth manifold with $X = U \cup V$, where U and V are open submanifolds of X. We have cochain complexes $\Lambda^*(X)$, $\Lambda^*(U)$, $\Lambda^*(V)$, and $\Lambda^*(U \cap V)$ consisting of the -modules of real-valued k-forms with exterior differentiation as the connecting homomorphisms. In addition, we form the direct sum complex $\Lambda^*(U) \oplus \Lambda^*(V)$ and the short exact sequence

$$0 \longrightarrow \Lambda^*(X) \xrightarrow{\alpha} \Lambda^*(U) \oplus \Lambda^*(V) \xrightarrow{\beta} \Lambda^*(U \cap V) \longrightarrow 0$$

as in Section 5.2 (see page 317). Theorem 5.3.1 then provides homomorphisms

$$\partial^k : H^k_{de\,R}(U \cap V) \longrightarrow H^{k+1}_{de\,R}(X)$$

such that the following long sequence (eventually ending with zeros) is exact:

$$\begin{aligned} 0 \longrightarrow H^0_{de\,R}(X) &\xrightarrow{(\alpha^0)^{\#}} H^0_{de\,R}(U) \oplus H^0_{de\,R}(V) \xrightarrow{(\beta^0)^{\#}} H^0_{de\,R}(U \cap V) \\ &\xrightarrow{\partial^0} \cdots H^{k-1}_{de\,R}(U \cap V) \xrightarrow{\partial^{k-1}} H^k_{de\,R}(X) \xrightarrow{(\alpha^k)^{\#}} H^k_{de\,R}(U) \oplus H^k_{de\,R}(V) \\ &\xrightarrow{(\beta^k)^{\#}} H^k_{de\,R}(U \cap V) \xrightarrow{\partial^k} H^{k+1}_{de\,R}(X) \longrightarrow \cdots . \end{aligned}$$

This is called the **Mayer-Vietoris sequence** and much of what remains of this chapter will consist of an enumeration of some of its consequences. Begin by noting that if $U \cap V = \emptyset$, then each $H^k_{de\,R}(U \cap V)$ is trivial so we have short exact sequences

$$0 \longrightarrow H^k_{de\,R}(X) \longrightarrow H^k_{de\,R}(U) \oplus H^k_{de\,R}(V) \longrightarrow 0$$

from which it follows that

$$H^k_{de\,R}(X) \cong H^k_{de\,R}(U) \oplus H^k_{de\,R}(V) \qquad (U \cap V = \emptyset).$$

Thus, in computing the cohomology of a manifold X one can consider each connected component separately.

Recall that an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in \mathcal{A}}$ of a manifold X is said to be simple if any finite intersection $U_{\alpha_0} \cap \cdots \cap U_{\alpha_j}$ of its elements is either empty or diffeomorphic to ^{n} (where $n = \dim X$) and that every smooth manifold has a simple cover (see Section 3.3). A manifold is said to be of finite type if it admits a finite simple cover. Any compact manifold is of finite type, but many noncompact manifolds are as well, e.g., ^{n}. We show now that any such manifold has finite dimensional cohomology.

Theorem 5.4.1 *Let X be a smooth manifold of finite type. Then each $H^k_{de\,R}(X)$ is a finite dimensional real vector space.*

Proof: The proof is by induction on the size of the simple cover. If X has a simple cover consisting of just one open set, then it is diffeomorphic to ^{n} ($n = \dim X$) so the result follows from Theorem 5.1.1. Assume now that the result has been established for manifolds admitting simple covers consisting of N open sets and suppose that X has a simple cover $\{U_1, \ldots, U_N, U_{N+1}\}$ consisting of $N+1$ open sets. Let $U = U_1 \cup \cdots \cup U_N$ and $V = U_{N+1}$. Then $U \cap V$ is an open submanifold of X and $\{U_1 \cap U_{N+1}, \ldots, U_N \cap U_{N+1}\}$ is a simple cover of it. By the induction hypothesis, all of the cohomology groups of U and $U \cap V$ are finite dimensional. Obviously, the same is true of $V = U_{N+1} \cong \mathbb{R}^n$. Now, $X = U \cup V$ so, for each k, the Mayer-Vietoris sequence gives the exact sequence

$$\cdots \longrightarrow H^{k-1}_{de\,R}(U \cap V) \xrightarrow{\partial} H^k_{de\,R}(X) \xrightarrow{\alpha} H^k_{de\,R}(U) \oplus H^k_{de\,R}(V) \longrightarrow \cdots$$

Since $H^k_{de\,R}(U) \oplus H^k_{de\,R}(V)$ is finite dimensional, so is Image (α). Since $H^k_{de\,R}(U \cap V)$ is finite dimensional, Image (∂) = ker (α) is finite dimensional. Since dim $(H^k_{de\,R}(X))$ = dim $(\ker(\alpha))$ + dim (Image (α)), it follows that $H^k_{de\,R}(X)$ is finite dimensional as well. ∎

Next we intend to calculate all of the cohomology groups of the spheres S^n, $n \geq 0$. The Mayer-Vietoris sequence again provides an inductive approach so we begin at the beginning. Recall that S^0 is the 2-point discrete subspace $\{-1, 1\}$ of $\mathbb{R}$.

Exercise 5.4.1 Show that $H^0(S^0) \cong \mathbb{R} \oplus \mathbb{R}$ and $H^k(S^0) = 0$ for all $k \geq 1$.

Remark: The de Rham cohomology "groups" are actually vector spaces so it might, perhaps, be more reasonable to write $H^0(S^0) \cong \mathbb{R}^2$. We will, however, bow to traditions that evolved in more general cohomology theo- ries and retain the direct sum notation.

We have already calculated the cohomology of S^1((5.1.1)), but let's do it again with Mayer-Vietoris. Write $S^1 = U_N \cup U_S$, where U_N is S^1 minus the south pole $(0, -1)$ and U_S is S^1 minus the north pole $(0, 1)$. Both U_N and U_S are diffeomorphic to $\mathbb{R}$ (via a stereographic projection), while $U_N \cap U_S$ is the disjoint union of two copies of $\mathbb{R}$. Since $H^k_{de\,R}(S^1) \cong 0$ for $k \geq 2$, the Mayer-Vietoris sequence gives

$$\begin{aligned} 0 \longrightarrow H^0_{de\,R}(S^1) \longrightarrow H^0_{de\,R}(U_N) \oplus H^0_{de\,R}(U_S) \longrightarrow H^0_{de\,R}(U_N \cap U_S) \longrightarrow \\ H^1_{de\,R}(S^1) \longrightarrow H^1_{de\,R}(U_N) \oplus H^1_{de\,R}(U_S) \longrightarrow H^1_{de\,R}(U_N \cap U_S) \longrightarrow 0 \end{aligned}$$

i.e.,

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} \longrightarrow H^1(S^1) \longrightarrow 0 \longrightarrow 0 \longrightarrow 0$$

which we collapse to the exact sequence

$$0 \longrightarrow \mathbb{R} \xrightarrow{\alpha} \mathbb{R} \oplus \mathbb{R} \xrightarrow{\beta} \mathbb{R} \oplus \mathbb{R} \xrightarrow{\partial} H^1(S^1) \longrightarrow 0.$$

Now, α is injective so dim (Image (α)) $= 1 = \dim(\ker(\beta))$. Moreover, $\dim(\mathbb{R} \oplus \mathbb{R}) =$ dim (Image (β)) $+ \dim(\ker(\beta))$ implies dim (Image (β)) $=1 =$ dim $(\ker(\partial))$ and from this it follows in the same way that dim (Image (∂)) $=$ 1. But ∂ is surjective so dim $(H^1(S^1)) = 1$, i.e., $H^1(S^1) \cong$, as expected.

Remark: DeRham cohomology is computationally much simpler than singular homology and the reason is apparent in this first example. DeRham cohomology "groups" are actually vector spaces and a vector space is completely determined by a single number (its dimension). Singular homology groups really are just (Abelian) groups and these are not so easy to characterize.

The sort of dimension counting employed in this last example occurs often enough that it is worth the effort to prove a lemma that will allow us to evade it in the future.

Lemma 5.4.2 *Let*

$$0 \longrightarrow \mathcal{V}_1 \longrightarrow \mathcal{V}_2 \longrightarrow \mathcal{V}_3 \longrightarrow \cdots \longrightarrow \mathcal{V}_{k-1} \longrightarrow \mathcal{V}_k \longrightarrow 0$$

be an exact sequence of finite dimensional vector spaces. Then

$$\dim \mathcal{V}_1 - \dim \mathcal{V}_2 + \dim \mathcal{V}_3 - \cdots + (-1)^{k-1} \dim \mathcal{V}_k = 0.$$

Proof: The proof is by induction on k. If $k = 1$ and $0 \longrightarrow \mathcal{V}_1 \longrightarrow 0$ is exact, then $\mathcal{V}_1$ is trivial so $\dim \mathcal{V}_1 = 0$. Now assume the result for sequences of $k-1$ vector spaces and consider the exact sequence

$$0 \longrightarrow \mathcal{V}_1 \xrightarrow{\alpha} \mathcal{V}_2 \xrightarrow{\beta} \mathcal{V}_3 \longrightarrow \cdots \longrightarrow \mathcal{V}_{k-1} \longrightarrow \mathcal{V}_k \longrightarrow 0.$$

Since ker $(\beta) =$ Image (α), $\beta : \mathcal{V}_2 \longrightarrow \mathcal{V}_3$ induces a map

$$\hat{\beta} : \mathcal{V}_2/\text{Image } (\alpha) \longrightarrow \mathcal{V}_3$$

defined by $\hat{\beta}(v_2 +$ Image $(\alpha)) = \beta(v_2)$ for each $v_2 \in \mathcal{V}_2$. Observe that $\hat{\beta}$ is injective since $\hat{\beta}(v_2 +$ Image $(\alpha)) = \hat{\beta}(v_2' +$ Image $(\alpha))$ implies $\beta(v_2) = \beta(v_2')$ so $\beta(v_2 - v_2') = 0$ and therefore $v_2 - v_2' \in$ ker $(\beta) =$ Image (α). This then implies that $v_2 +$ Image $(\alpha) = v_2' +$ Image (α) as required. Thus,

$$0 \longrightarrow \mathcal{V}_2/\text{Image } (\alpha) \xrightarrow{\hat{\beta}} \mathcal{V}_3$$

is exact. Since Image $(\hat{\beta})$ = Image (β) it follows that, in fact,

$$0 \longrightarrow \mathcal{V}_2/\text{Image}\ (\alpha) \xrightarrow{\hat{\beta}} \mathcal{V}_3 \longrightarrow \cdots \longrightarrow \mathcal{V}_{k-1} \longrightarrow \mathcal{V}_k \longrightarrow 0$$

is exact. The induction hypothesis therefore gives

$$\begin{aligned} 0 &= \dim(\mathcal{V}_2/\text{Image}\,(\alpha)) - \dim \mathcal{V}_3 + \cdots + (-1)^k \dim \mathcal{V}_k \\ &= \dim \mathcal{V}_2 - \dim(\text{Image}\,(\alpha)) - \dim \mathcal{V}_3 + \cdots + (-1)^k \dim \mathcal{V}_k \\ &= \dim \mathcal{V}_2 - \dim \mathcal{V}_1 - \dim \mathcal{V}_3 + \ \cdots\ + (-1)^k \dim \mathcal{V}_k \end{aligned}$$

since α is one-to-one. Thus

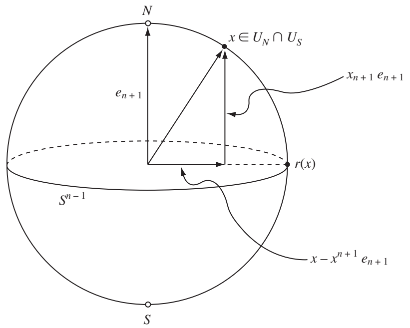
$$0 = \dim \mathcal{V}_1 - \dim \mathcal{V}_2 + \dim \mathcal{V}_3 - \ \cdots\ + (-1)^{k-1} \dim \mathcal{V}_k$$

as required. ∎

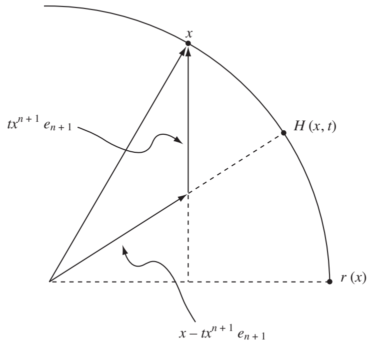
Now we will use the Mayer-Vietoris sequence to calculate the cohomology of S^n. The argument is an induction on the dimension of the sphere and is based on the following observation: S^n can be written as the union $S^n = U_N \cup U_S$ of two copies of $\mathbb{R}^n$ for which $U_N \cap U_S$ is S^n minus two points (the south and north poles). We know the cohomology of U_N and U_S. Furthermore, there is a smooth deformation retraction of $U_N \cap U_S$ onto the equator S^{n-1} in S^n (Exercise 5.4.2) so the induction hypothesis will give us the cohomology of $U_N \cap U_S$ as well. Mayer-Vietoris will then put all of this together into the cohomology of S^n.

Exercise 5.4.2 Let $n \geq 2$ be an integer and write $S^n = U_N \cup U_S$, where U_N is S^n minus the south pole $(0, \ldots, 0, -1)$ and U_S is S^n minus the north pole $(0, \ldots, 0, 1)$. Let $\{e_1, \ldots, e_n, e_{n+1}\}$ be the standard basis for $\mathbb{R}^{n+1} \supseteq S^n$ and let S^{n-1} be the equatorial $(n-1)$-sphere in S^n, i.e., $S^{n-1} = \{(x^1, \ldots, x^n, x^{n+1}) \in S^n : x^{n+1} = 0\}$. Define $r : U_N \cap U_S \longrightarrow S^{n-1}$ by

$$r(x) = \frac{x - x^{n+1} e_{n+1}}{|x - x^{n+1} e_{n+1}|}$$



Show that r is a smooth deformation retraction of $U_N \cap U_S$ onto S^{n-1}. **Hint**: For the required homotopy, consider the following figure, where $0 \le t \le 1$.



Now, to carry out an induction one needs an induction hypothesis. Our only information thus far concerns S^1:

$$H^k_{de\,R}(S^1) \cong \begin{cases} \mathbb{R}, & \text{if } k = 0, 1 \\ 0, & \text{if } k \neq 0,\ k \neq 1 \end{cases} .$$

To check that this is indicative of the general result (and to get a bit more practice with Mayer-Vietoris) we will do S^2 as well. Of course, $H^k_{de\,R}(S^2) \cong 0$ for $k \ge 3$ so the Mayer-Vietoris sequence gives

$$\begin{aligned} 0 \longrightarrow\ & H^0_{de\,R}(S^2) \longrightarrow H^0_{de\,R}(U_N) \oplus H^0_{de\,R}(U_S) \longrightarrow H^0_{de\,R}(U_N \cap U_S) \longrightarrow \\ & H^1_{de\,R}(S^2) \longrightarrow H^1_{de\,R}(U_N) \oplus H^1_{de\,R}(U_S) \longrightarrow H^1_{de\,R}(U_N \cap U_S) \longrightarrow \\ & H^2_{de\,R}(S^2) \longrightarrow H^2_{de\,R}(U_N) \oplus H^2_{de\,R}(U_S) \longrightarrow H^2_{de\,R}(U_N \cap U_S) \longrightarrow 0. \end{aligned}$$

Now, $H^0_{de\,R}(U_N) \oplus H^0_{de\,R}(U_S) = \mathbb{R} \oplus \mathbb{R}$, but the remaining direct sums are trivial. Furthermore, Exercise 5.4.2 and Corollary 5.2.4 imply that $U_N \cap U_S$ has the same cohomology as S^1. Thus, our sequence becomes

$$\begin{array}{rcccccc} 0 \longrightarrow & \mathbb{R} & \longrightarrow & \mathbb{R} \oplus \mathbb{R} & \longrightarrow \mathbb{R} & \longrightarrow & \\ & H^1_{de\,R}(S^2) & \longrightarrow & 0 & \longrightarrow \mathbb{R} & \longrightarrow & \\ & H^2_{de\,R}(S^2) & \longrightarrow & 0 & \longrightarrow 0 & \longrightarrow & 0. \end{array}$$

This "splits" into two exact sequences:

$$0 \longrightarrow \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} \longrightarrow \mathbb{R} \longrightarrow H^1_{de\,R}(S^2) \longrightarrow 0$$

and

$$0 \longrightarrow \mathbb{R} \longrightarrow H^2_{de\,R}(S^2) \longrightarrow 0.$$

Applying Lemma 5.4.2 to each gives

$$1 - 2 + 1 - \dim\,(H^1_{de\,R}(S^2)) = 0$$

and

$$1 - \dim\,(H^2_{de\,R}(S^2)) = 0$$

so

$$H^1_{de\,R}(S^2) \cong 0 \quad \text{and} \quad H^2_{de\,R}(S^2) \cong \mathbb{R}.$$

We conclude then that

$$H^k_{de\,R}(S^2) \cong \begin{cases} \mathbb{R}, & \text{if } k = 0, 2 \\ 0, & \text{if } k \neq 0,\ k \neq 2 \end{cases}.$$

Remark: Notice that, for $k = 1$, the sequence of forms and the corresponding cohomology sequence are

$$0 \longrightarrow \Lambda^1(S^2) \longrightarrow \Lambda^1(U_N) \oplus \Lambda^1(U_S) \longrightarrow \Lambda^1(U_N \cap U_S) \longrightarrow 0$$

which we know is exact and

$$0 \longrightarrow 0 \longrightarrow 0 \oplus 0 \longrightarrow \mathbb{R} \longrightarrow 0$$

which certainly is not. This provides the example promised just before Theorem 5.3.1.

Exercise 5.4.3 Prove, by induction on $n \geq 1$, that

$$H^k_{de\,R}(S^n) \cong \begin{cases} \mathbb{R}, & \text{if} \quad k = 0, n \\ 0, & \text{if} \quad k \neq 0,\ k \neq n \end{cases}.$$

Remark: Since $H^n_{de\,R}(S^n)$ has dimension one for every $n \geq 1$ it is generated by the cohomology class of any nonexact n-form on S^n, i.e., any n-form that does not integrate to zero over S^n (see Section 5.1). We found in Section 4.6 that the standard volume form on S^n is of this type.

With this we are now in a position to generalize Lemma 5.1.3.

Theorem 5.4.3 *An n-form $\boldsymbol{\nu}$ on S^n is exact if and only if $\int_{S^n} \boldsymbol{\nu} = 0$.*

Proof: All that remains is to show that, if $\int_{S^n} \boldsymbol{\nu} = 0$, then $\boldsymbol{\nu}$ is exact. Since $\boldsymbol{\nu}$ is necessarily closed, it determines a cohomology class $[\boldsymbol{\nu}]$ in $H^n_{de\,R}(S^n)$. But $\dim H^n_{de\,R}(S^n) = 1$ so $[\boldsymbol{\nu}] = \alpha[\boldsymbol{\omega}]$, where $\boldsymbol{\omega}$ is any nonexact n-form on S^n and α is some real number. Now, $[\boldsymbol{\nu}] = [\alpha\boldsymbol{\omega}]$ implies $[\boldsymbol{\nu} - \alpha\boldsymbol{\omega}] = [0]$ so $\boldsymbol{\nu} - \alpha\boldsymbol{\omega}$ is exact and therefore integrates to zero over S^n. Thus,

$$0 = \int_{S^n} (\boldsymbol{\nu} - \alpha\boldsymbol{\omega}) = \int_{S^n} \boldsymbol{\nu} - \alpha \int_{S^n} \boldsymbol{\omega}$$

so

$$\int_{S^n} \boldsymbol{\nu} = \alpha \int_{S^n} \boldsymbol{\omega}.$$

If $\int_{S^n} \boldsymbol{\nu} = 0$, then, since $\int_{S^n} \boldsymbol{\omega} \neq 0$, we must have $\alpha = 0$ and so $[\boldsymbol{\nu}] = [0]$, i.e., $\boldsymbol{\nu}$ is exact. ■

Remark: We will eventually show that the same is true for any compact, orientable manifold.

Exercise 5.4.4 For $[\boldsymbol{\omega}] \in H^n_{de\,R}(S^n)$, define the integral of $[\boldsymbol{\omega}]$ over S^n by

$$\int_{S^n} [\boldsymbol{\omega}] = \int_{S^n} \boldsymbol{\omega}.$$

Show that this is well-defined (i.e., independent of the representative $\boldsymbol{\omega}$ of $[\boldsymbol{\omega}]$) and that the map

$$\int_{S^n} : H^n_{de\,R}(S^n) \longrightarrow \mathbb{R}$$

thus defined is an isomorphism.

Remark: This too will generalize to arbitrary compact, connected, orientable manifolds.

Exercise 5.4.5 Let 0 denote the origin $(0, \ldots, 0)$ in $\mathbb{R}^{n+1}$, $n \geq 1$. Show that

$$H^k_{de\,R}(\mathbb{R}^{n+1} - \{0\}) \cong \begin{cases} \mathbb{R}, & \text{if} \quad k = 0,\ n \\ 0, & \text{if} \quad k \neq 0,\ k \neq n \end{cases}.$$

Hint: Exercise 5.4.3.

Exercise 5.4.6 Let X denote the subspace of $\mathbb{R}^{n+1}$, $n \geq 2$, obtained by deleting the "x^{n+1}-axis" (i.e., the set of all points of the form $(0, \ldots, 0, t)$ for $t \in \mathbb{R}$). Show that

$$H^k_{de\,R}(X) \cong \begin{cases} \mathbb{R}, & \text{if} \quad k = 0, n-1 \\ 0, & \text{if} \quad k \neq 0,\ k \neq n-1 \end{cases}.$$

Hint: Corollary 5.2.2.

One can use the cohomology groups we have already calculated to prove some rather nontrivial results concerning the topology of Euclidean spaces. We illustrate by proving that $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic if and only if $n = m$.

Remark: Note that the corresponding assertion in the linear category (i.e., $\mathbb{R}^n$ and $\mathbb{R}^m$ are isomorphic if and only if $n = m$) is obvious and that this renders the analogous result in the smooth category (i.e., $\mathbb{R}^n$ and $\mathbb{R}^m$ are diffeomorphic if and only if $n = m$) equally obvious since diffeomorphisms have derivatives that are isomorphisms. The game here is to get from the smooth to the topological category. This was done using singular homology in Section 3.4 of [**N4**], but the computations required to do it were substantially more involved.

In Section 3.2 we proved that, if U and V are open submanifolds of some Euclidean spaces, then any smooth map $h : U \longrightarrow V$ is (continuously) homotopic to a smooth map $f : U \longrightarrow V$ and that two smooth maps f, $f' : U \longrightarrow V$ that are continuously homotopic are also smoothly homotopic. We use these facts to prove that continuous maps from U to V induce linear maps from $H^k_{de\,R}(V)$ to $H^k_{de\,R}(U)$ for every k.

Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open sets and $h : U \longrightarrow V$ a continuous map. Select a smooth map $f : U \longrightarrow V$ that is continuously homotopic to h. Then for each k, f induces a linear map $f^\#$ from $H^k_{de\,R}(Y)$ to $H^k_{de\,R}(X)$. If $f' : U \longrightarrow V$ is any other smooth map continuously homotopic to h, then f and f' are continuously homotopic and, consequently, are also smoothly homotopic. Corollary 5.2.3 then implies that f and f' induce the same linear maps in cohomology. Thus, we may unambiguously define

$$h^\# : H^k_{de\,R}(Y) \longrightarrow H^k_{de\,R}(X)$$

by

$$h^\# = f^\#,$$

where f is *any* smooth map homotopic to h. We ask the reader to establish all of the usual properties.

Exercise 5.4.7 Let U, V and W be open sets in Euclidean spaces and k any integer. Prove each of the following.

1. If $h_0,\ h_1 : U \longrightarrow V$ are homotopic continuous maps, then
$$h_0^\# = h_1^\# : H^k_{de\,R}(V) \longrightarrow H^k_{de\,R}(U).$$
2. If $h : U \longrightarrow V$ and $g : V \longrightarrow W$ are continuous maps, then
$$(g \circ h)^\# = h^\# \circ g^\# : H^k_{de\,R}(W) \longrightarrow H^k_{de\,R}(U).$$
3. If the continuous map $h : U \longrightarrow V$ is a homotopy equivalence, then $h^\# : H^k_{de\,R}(V) \longrightarrow H^k_{de\,R}(U)$ is an isomorphism.

It follows, in particular, from Exercise 5.4.7 (3), that if $h : U \longrightarrow V$ is a homeomorphism, then $h^\#$: $H^k_{de\,R}(V) \longrightarrow H^k_{de\,R}(U)$ is an isomorphism for each k. Now, suppose $\mathbb{R}^n$ and $\mathbb{R}^m$ are Euclidean spaces. Certainly, if $n = m$, then $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic. Suppose, conversely, that there is a homeomorphism $H : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. By composing with a translation if necessary we may assume that H carries the origin $0 \in \mathbb{R}^n$ onto the origin $0 \in \mathbb{R}^m$. Consequently,

$$h = H|_{\mathbb{R}^n - \{0\}} : \mathbb{R}^n - \{0\} \longrightarrow \mathbb{R}^m - \{0\}$$

is a homeomorphism. But then

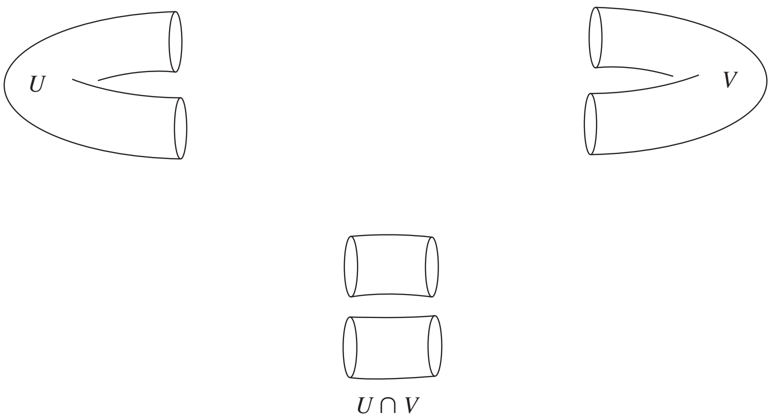
$$h^{\#} : H^k_{de\,R}(\mathbb{R}^m - \{0\}) \longrightarrow H^k_{de\,R}(\mathbb{R}^n - \{0\})$$

is an isomorphism for each k so, in particular, this is true when $k = m - 1$. Now, if $m = 1$, $H^0_{de\,R}(\mathbb{R}^n - \{0\}) \cong \mathbb{R} \oplus \mathbb{R}$ so we must have $n = 1$ as well ($n \geq 2$ implies $H^0_{de\,R}(\mathbb{R}^n - \{0\}) \cong \mathbb{R}$ since $\mathbb{R}^n - \{0\}$ is connected). If $m \geq 2$, then Exercise 5.4.5 implies that $H^{m-1}_{de\,R}(\mathbb{R}^m - \{0\}) \cong \mathbb{R}$ so $H^{m-1}_{deR}(\mathbb{R}^n - \{0\}) \cong \mathbb{R}$ as well. This is impossible if $n = 1$ and, if $n \geq 2$, Exercise 5.4.5 again implies that we must have $m - 1 = n - 1$ so $m = n$. Thus we have proved the topological invariance of dimension for Euclidean spaces.

Theorem 5.4.4 *$\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic if and only if $n = m$.*

Remark: Many other classical theorems of Euclidean topology (Brouwer Fixed Point Theorem, Jordan-Brouwer Separation Theorem, Invariance of Domain, etc.) are accessible to similar techniques. Those interested in seeing some of this are referred to Chapter 7 of [**MT**]. Similar applications were treated from the perspective of homology theory in Section 3.4 of [**N4**].

Exercise 5.4.8 Compute the de Rham cohomology of the torus $T = S^1 \times S^1$. **Hint**: Mayer-Vietoris with the open sets U and V shown below.



Exercise 5.4.9 Let $p_1, \ldots, p_l$ be l distinct points in $\mathbb{R}^n$, $n \geq 2$. Show that

$$H^k_{de\,R}(\mathbb{R}^n - \{p_1, \ldots, p_l\}) \cong \begin{cases} \mathbb{R}^l, & \text{if } k = n-1 \\ \mathbb{R}, & \text{if } k = 0 \\ 0, & \text{if } k \neq 0,\ n-1 \end{cases}.$$

Exercise 5.4.10 Show that if X and Y are smooth manifolds with Y smoothly contractible, then $H^k_{de\,R}(X \times Y) \cong H^k_{de\,R}(X)$ for every k.

Remark: The cohomology of products in general is described by the *Künneth formula* (see Chapter 5, Section 6, of [**GHVI**]).

Exercise 5.4.11 Show that the de Rham cohomology of a smooth principal bundle does not, in general, coincide with that of its base manifold.

Remark: For the cohomology of principal bundles, consult the *Leray-Hirsch Theorem* (Theorem 5.11, [**BT**]).

5.5 The Cohomology of Compact, Orientable Manifolds

Our objective here is to generalize a number of results proved in the last section for spheres. Specifically, we will show that if X is any compact, connected, orientable n-manifold, then $H^n_{de\,R}(X) \cong \mathbb{R}$. This result will be used in the next section to define the degree of a map between two such manifolds. We begin with a sequence of lemmas. The first deals with the following situation. Any $\boldsymbol{\omega} \in \Lambda^n(\mathbb{R}^n)$ is exact. Suppose $\boldsymbol{\omega}$ also has compact support. When will there exist an $\boldsymbol{\eta} \in \Lambda^{n-1}(\mathbb{R}^n)$, also with compact support, such that $\boldsymbol{\omega} = d\boldsymbol{\eta}$? The lemma asserts that the obvious necessary condition is also sufficient.

Lemma 5.5.1 *Let $\boldsymbol{\omega} \in \Lambda^n(\mathbb{R}^n)$ have compact support. Then there exists a compactly supported $\boldsymbol{\eta} \in \Lambda^{n-1}(\mathbb{R}^n)$ with $d\boldsymbol{\eta} = \boldsymbol{\omega}$ if and only if*

$$\int_{\mathbb{R}^n} \boldsymbol{\omega} \; = 0.$$

Proof: Assume first that such an $\boldsymbol{\eta}$ exists. Then $d\boldsymbol{\eta}$ is also compactly supported and therefore integrable on $\mathbb{R}^n$ so Stokes' Theorem gives

$$\int_{\mathbb{R}^n} \boldsymbol{\omega} = \int_{\mathbb{R}^n} d\boldsymbol{\eta} = 0.$$

Conversely, suppose $\int_{\mathbb{R}^n} \boldsymbol{\omega} = 0$. Let $\varphi_S : U_S \longrightarrow \mathbb{R}^n$ be the stereographic projection from the north pole N of S^n. This is an orientation preserving diffeomorphism from $U_S = S^n - \{N\}$ onto $\mathbb{R}^n$ (Exercise 5.10.14, [**N4**]). Moreover, $\boldsymbol{\omega}' = \varphi_S^* \boldsymbol{\omega}$ is an n-form on U_S whose compact support is contained in the complement of the intersection U of S^n with some open ball about N in $\mathbb{R}^{n+1}$. We may, via a bump function (Lemma 3.1.3) regard $\boldsymbol{\omega}'$ as defined on all of S^n. Moreover,

$$\int_{S^n} \boldsymbol{\omega}' = \int_{S^n - \{N\}} \boldsymbol{\omega}' = \int_{S^n - \{N\}} \varphi_S^* \boldsymbol{\omega} = \int_{\mathbb{R}^n} \boldsymbol{\omega} = 0.$$

By Theorem 5.4.3, $\boldsymbol{\omega}'$ is exact, i.e., there exists a $\boldsymbol{\xi} \in \Lambda^{n-1}(S^n)$ such that $\boldsymbol{\omega}' = d\boldsymbol{\xi}$. Moreover, $d\,\boldsymbol{\xi}$ vanishes on U and U is contractible so there exists an $(n-2)$-form $\boldsymbol{\nu}$ on U with $\boldsymbol{\xi} = d\boldsymbol{\nu}$ on U. With a bump function that is 1 on a smaller neighborhood $V \subseteq U$ of N in S^n we can regard $\boldsymbol{\nu}$ as defined on all of S^n with $d\boldsymbol{\nu} = \boldsymbol{\xi}$ on V and $d\boldsymbol{\nu} = 0$ outside U. Thus, $\boldsymbol{\xi} - d\boldsymbol{\nu} \in \Lambda^{n-1}(S^n)$ vanishes on V so $\operatorname{supp}(\boldsymbol{\xi} - d\boldsymbol{\nu}) \subseteq S^n - V$. Moreover, $d(\boldsymbol{\xi} - d\boldsymbol{\nu}) = d\boldsymbol{\xi} = \boldsymbol{\omega}'$ on S^n. Denoting the restriction of $\boldsymbol{\xi} - d\boldsymbol{\nu}$ to $S^n - \{N\}$ by $\boldsymbol{\xi} - d\boldsymbol{\nu}$ also we let $\boldsymbol{\eta} = (\varphi_S^{-1})^*(\boldsymbol{\xi} - d\boldsymbol{\nu})$. The support of $\boldsymbol{\eta}$ is contained in $\varphi_S(S^n - V)$ so it is bounded and therefore compact.

Exercise 5.5.1 Complete the proof by showing that $d\boldsymbol{\eta} = \boldsymbol{\omega}$. ■

Exercise 5.5.2 Use Lemma 5.5.1 to show that there exists a compactly supported n-form $\boldsymbol{\omega}'$ on $\mathbb{R}^n$ with

$$\int_{\mathbb{R}^n} \boldsymbol{\omega}' = 1.$$

Lemma 5.5.2 *Let X be an orientable n-manifold and U an open subset of X that is diffeomorphic to $\mathbb{R}^n$. Then there exists an n-form $\boldsymbol{\omega}$ on X with compact support contained in U such that*

$$\int_X \boldsymbol{\omega} = 1.$$

Proof: Let ϕ be an orientation preserving diffeomorphism of U into $\mathbb{R}^n$ and let $\boldsymbol{\omega}'$ be the n-form on $\mathbb{R}^n$ described in Exercise 5.5.2. Then $\phi^*\boldsymbol{\omega}'$ is an n-form on U with support $\operatorname{supp}(\phi^*\boldsymbol{\omega}') = \phi^{-1}(\operatorname{supp}\boldsymbol{\omega}')$ and this is compact. With a bump function that is 1 on $\operatorname{supp}(\phi^*\boldsymbol{\omega}')$ and 0 outside U we may regard $\boldsymbol{\omega} = \phi^*\boldsymbol{\omega}'$ as defined on all of X. Moreover,

$$\int_X \boldsymbol{\omega} = \int_U \boldsymbol{\omega} = \int_U \phi^*\boldsymbol{\omega}' = \int_{\mathbb{R}^n} \boldsymbol{\omega}' = 1. \quad ■$$

Lemma 5.5.3 *Let X be an orientable n-manifold, U an open subset of X diffeomorphic to $\mathbb{R}^n$ and $\boldsymbol{\omega}$ an n-form on X with compact support contained in U and with $\int_X \boldsymbol{\omega} = 1$. If $\boldsymbol{\xi}$ is any other n-form on X with compact support contained in U, then there exists a $c \in \mathbb{R}$ such that $\boldsymbol{\xi}$ is cohomologous to $c\boldsymbol{\omega}$ on X.*

Proof: Since $\boldsymbol{\omega}$ and $\boldsymbol{\xi}$ have supports contained in U we will use the same symbols for their restrictions to U. Let ϕ be an orientation preserving diffeomorphism of U onto $\mathbb{R}^n$ and set

$$c = \int_X \boldsymbol{\xi} = \int_U \boldsymbol{\xi} = \int_{\mathbb{R}^n} (\phi^{-1})^*\boldsymbol{\xi}.$$

Note that

$$\int_{\mathbb{R}^n} (\phi^{-1})^*\boldsymbol{\omega} = \int_U \boldsymbol{\omega} = \int_X \boldsymbol{\omega} = 1$$

so

$$\int_{\mathbb{R}^n} (\phi^{-1})^*(c\boldsymbol{\omega}) = c.$$

Thus,

$$\int_{\mathbb{R}^n} (\phi^{-1})^*(\boldsymbol{\xi} - c\,\boldsymbol{\omega}\ ) = 0.$$

Since $(\phi^{-1})^*(\boldsymbol{\xi} - c\,\boldsymbol{\omega}\ )$ has compact support, Lemma 5.5.1 implies that there exists a compactly supported $\boldsymbol{\eta} \in \Lambda^{n-1}(\mathbb{R}^n)$ such that

$$(\phi^{-1})^*(\boldsymbol{\xi} - c\,\boldsymbol{\omega}\ ) = d\boldsymbol{\eta}$$

on $\mathbb{R}^n$. Thus,

$$\boldsymbol{\xi} - c\,\boldsymbol{\omega}\ = \phi^*(d\boldsymbol{\eta}) = d(\phi^*\boldsymbol{\eta})$$

on U. Now, $\phi^*\boldsymbol{\eta}$ has compact support contained in U so, with a bump function that is 1 on $\operatorname{supp}(\phi^*\boldsymbol{\eta})$ and 0 outside U, we may regard it as an $(n-1)$-form $\boldsymbol{\nu}$ on all of X. Then

$$\boldsymbol{\xi} - c\,\boldsymbol{\omega}\ = d\boldsymbol{\nu}$$

on all of X (because everything is zero outside of U) so $\boldsymbol{\xi}$ and $c\,\boldsymbol{\omega}$ are cohomologous on X. ∎

Theorem 5.5.4 *Let X be a compact. connected, orientable n-manifold. Then*

$$H^n_{\text{de }R}(X) \cong \mathbb{R}.$$

Proof: Select an open set U in X diffeomorphic to $\mathbb{R}^n$ and an $\boldsymbol{\omega} \in \Lambda^n(X)$ with compact support contained in U and such that $\int_X \boldsymbol{\omega} = 1$ (Lemma 5.5.2). We will show for any other n-form $\boldsymbol{\xi}$ on X (which necessarily has compact support since X is compact) there exists a $c \in \mathbb{R}$ such that $\boldsymbol{\xi}$ is cohomologous to $c\,\boldsymbol{\omega}$. This will give $\dim H^n_{de\,R}(X) \leq 1$ and the reverse inequality is Lemma 5.1.2. If $\boldsymbol{\xi}$ happens to have its support also contained in U, then our result follows from Lemma 5.5.3. In general, this need not be the case, of course. However, we claim that we may assume, without loss of generality, that $\operatorname{supp}\,\boldsymbol{\xi}$ is contained in *some* open set $V \subseteq X$ that is diffeomorphic to $\mathbb{R}^n$. To see this, cover X with finitely many coordinate neighborhoods $U_1, \ldots, U_k$, each diffeomorphic to $\mathbb{R}^n$, and choose a family of functions $\phi_1, \ldots, \phi_k$ for this cover of the type guaranteed by Corollary 3.1.5. Then $\boldsymbol{\xi} = \phi_1\,\boldsymbol{\xi} + \cdots + \phi_k\,\boldsymbol{\xi}$ and each $\phi_i\,\boldsymbol{\xi}$ has support contained in U_i. It will suffice to find $c_i \in \mathbb{R}$ and $\boldsymbol{\eta}_i \in \Lambda^{n-1}(X)$ such that $\phi_i\,\boldsymbol{\xi} = c_i\,\boldsymbol{\omega} + d\boldsymbol{\eta}_i$ since then $\boldsymbol{\xi} = (c_1 + \cdots + c_k)\boldsymbol{\omega} + d(\boldsymbol{\eta}_1 + \cdots + \boldsymbol{\eta}_k)$.

Thus, we assume $\operatorname{supp}\boldsymbol{\xi} \subseteq V$, where V is an open set in X that is diffeomorphic to $\mathbb{R}^n$.

Exercise 5.5.3 Show that we may select a sequence of open sets

$$U = V_1,\ V_2, \ldots, V_{k-1},\ V_k = V$$

in X, each diffeomorphic to $\mathbb{R}^n$, with $V_i \cap V_{i+1} \neq \emptyset$ for $i = 1, 2, \ldots, k-1$. **Hint**: X is connected and therefore pathwise connected (Corollary 1.5.6, [**N4**]).

Inside $V_i \cap V_{i+1}$ select an open set U_i diffeomorphic to $\mathbb{R}^n$ and, by Lemma 5.5.2, an n-form $\boldsymbol{\omega}_i$ on X with compact support contained in U_i and such that $\int_X \boldsymbol{\omega}_i = 1$. Since supp $\boldsymbol{\omega}_1 \subseteq U_1 \subseteq V_1 \cap V_2 \subseteq U$, Lemma 5.5.3 gives a real number c_1 such that $\boldsymbol{\omega}_1$ is cohomologous to $c_1\, \boldsymbol{\omega}$ (we will write $\boldsymbol{\omega}_1 \sim c_1\, \boldsymbol{\omega}$). But supp $\boldsymbol{\omega}_1 \subseteq V_2$ and supp $\boldsymbol{\omega}_2 \subseteq V_2$ so the same lemma provides a $c_2 \in \mathbb{R}$ for which $\boldsymbol{\omega}_2 \sim c_2 \boldsymbol{\omega}_1$ and therefore $\boldsymbol{\omega}_2 \sim (c_1 c_2)\boldsymbol{\omega}$. Continuing inductively we obtain

$$\boldsymbol{\omega}_{k-1} \sim (c_1 c_2 \cdots c_{k-1})\boldsymbol{\omega}.$$

But then supp $\boldsymbol{\xi} \subseteq V = V_k$ and supp $\boldsymbol{\omega}_k \subseteq V_k$ implies that there exists a $c_k \in \mathbb{R}$ such that $\boldsymbol{\xi} \sim c_k \boldsymbol{\omega}_{k-1}$ and therefore

$$\boldsymbol{\xi} \sim (c_1 c_2 \cdots c_{k-1} c_k)\boldsymbol{\omega}$$

as required. ∎

Corollary 5.5.5 *Let X be a compact, orrientable n-manifold. Then an n-form $\boldsymbol{\omega}$ on X is exact if and only if $\int_X \boldsymbol{\omega} = 0$.*

Exercise 5.5.4 Prove Corollary 5.5.5. **Hint**: See the proof of Theorem 5.4.3. ∎

Corollary 5.5.6 *Let X be a compact, orientable n-manifold. Then*

$$\int_X : H^n_{de\,R}(X) \longrightarrow \mathbb{R}$$

is an isomorphism.

Exercise 5.5.5 Prove Corollary 5.5.6.

Exercise 5.5.6 Suppose X is a compact, connected orientable n-manifold, where $n = 2k$ for some $k \geq 1$. Define $Q_X : H^k_{de\,R}(X) \times H^k_{de\,R}(X) \longrightarrow \mathbb{R}$ by

$$Q_X([\boldsymbol{\xi}],\ [\boldsymbol{\eta}]) = \int_X \boldsymbol{\xi} \wedge \boldsymbol{\eta}.$$

Show that Q_X is well-defined, bilinear and satisfies $Q_X([\boldsymbol{\eta}],[\boldsymbol{\xi}]) = (-1)^k\, Q_X([\boldsymbol{\xi}],[\boldsymbol{\eta}])$. In particular, when $n \equiv 0 \mod 4$, Q_X is symmetric. Q_X is called the **intersection form** of X. Another definition of the intersection form for 4-manifolds was introduced in Appendix B of [**N4**].

We conclude this section with a few remarks on matters of considerable importance, but which we do not intend to pursue in detail. We have shown that the 0^{th} and n^{th} deRham cohomology groups of a compact, connected, orientable n-manifold X are always isomorphic to $\mathbb{R}$. In particular, $H^0_{de\,R}(X) \cong H^n_{de\,R}(X)$. This is a special case of a more general result known as *Poincaré Duality* which asserts that, for any $k = 0, 1, \ldots, n$,

$$H^{n-k}_{de\,R}(X) \cong H^k_{de\,R}(X).$$

Remark: The Poincaré Duality Theorem is actually much more general than we have indicated here. Those interested in pursuing this should consult either Chapter 13 of [**MT**], or Chapter 11, Volume I, of [**Sp2**].

Any n-manifold X of finite type has finite dimensional cohomology (Theorem 5.4.1) so we can define, for each $k = 0, 1, \ldots, n$,

$$b^k = \dim H^k_{de\,R}(X)$$

and form their alternating sum to obtain

$$b_0 - b_1 + b_2 - \cdots + (-1)^n b_n = \sum_{k=0}^{n} (-1)^k \dim H^k_{de\,R}(X).$$

Although it is not obvious, this alternating sum coincides with the Euler characteristic of X, defined in Chapter 3 of [**N4**] as the alternating sum of the ranks of the singular homology groups of X. We will have a bit more to say about the Euler characteristic in the Appendix.

5.6 The Brouwer Degree

Suppose X and Y are two compact, connected, orientable, n-dimensional manifolds and $f : X \longrightarrow Y$ is a smooth map. By Lemma 5.5.2 we may select a (necessarily closed) n-form $\boldsymbol{\omega}_0$ on Y with $\int_Y \boldsymbol{\omega}_0 = 1$. Corollary 5.5.5 implies that $\boldsymbol{\omega}_0$ is not exact and determines a nontrivial cohomology class $[\boldsymbol{\omega}_0] \in H^n_{de\,R}(Y)$. But Theorem 5.5.4 then implies that $[\boldsymbol{\omega}_0]$ generates $H^n_{de\,R}(Y)$. We call $\boldsymbol{\omega}_0$ a **normalized generator** for $H^n_{de\,R}(Y)$. Pull $\boldsymbol{\omega}_0$ back to X by f and integrate over X to obtain a real number that we will call the (**Brouwer**) **degree of** f and denote

$$\deg(f) = \int_X f^* \boldsymbol{\omega}_0. \qquad (5.6.1)$$

We must show that this definition does not depend on the choice of the normalized generator $\boldsymbol{\omega}_0$ for $H^n_{de\,R}(Y)$. Thus, suppose $\boldsymbol{\omega}_0'$ is another n-form on Y with $\int_Y \boldsymbol{\omega}_0' = 1$. Since $[\boldsymbol{\omega}_0]$ generates $H^n_{de\,R}(Y)$ there exists an $\alpha \in \mathbb{R}$ such that

$$\boldsymbol{\omega}_0' = \alpha \boldsymbol{\omega}_0 + d\boldsymbol{\eta}$$

for some $\boldsymbol{\eta} \in \Lambda^{n-1}(Y)$.

Exercise 5.6.1 Show that $\alpha = 1$.

Thus, $\boldsymbol{\omega}_0' = \boldsymbol{\omega}_0 + d\boldsymbol{\eta}$ so $f^* \boldsymbol{\omega}_0' = f^* \boldsymbol{\omega}_0 + d(f^* \boldsymbol{\omega}_0)$ and, by Stokes' Theorem,

$$\int_X f^* \boldsymbol{\omega}_0' = \int_X f^* \boldsymbol{\omega}_0$$

as required. Now consider any other (necessarily closed) n-form $\boldsymbol{\omega}$ on Y. Again, because $[\boldsymbol{\omega}_0]$ generates $H^n_{de\,R}(Y)$, there exists an $\alpha \in \mathbb{R}$ such that

$$\boldsymbol{\omega} = \alpha\boldsymbol{\omega}_0 + d\boldsymbol{\eta}$$

for some $\boldsymbol{\eta} \in \Lambda^{n-1}(Y)$.

Exercise 5.6.2 Show that $\alpha = \int_Y \boldsymbol{\omega}$.

Thus, $f^*\boldsymbol{\omega} = \alpha f^*\boldsymbol{\omega}_0 + d(f^*\boldsymbol{\eta}_0)$ so Stokes' Theorem gives

$$\int_X f^*\boldsymbol{\omega} \;= \alpha \int_X f^*\boldsymbol{\omega}_0 = \alpha\,\deg(f)$$

which we write as

$$\int_X f^*\boldsymbol{\omega} = \deg(f) \int_Y \boldsymbol{\omega}. \tag{5.6.2}$$

Notice that (5.6.1) is the special case of (5.6.2) in which $\boldsymbol{\omega}$ is a normalized generator for $H^n_{de\,R}(Y)$.

Remark: Although it is by no means apparent from the definition we have just given, we will prove shortly that $\deg(f)$ is actually an *integer*.

Exercise 5.6.3 Let X, Y and Z be compact, connected orientable n-manifolds and $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ smooth maps. Show that

$$\deg(g \circ f) = \deg(g)\deg(f).$$

Concrete calculations are greatly facilitated by the fact that deg is a homotopy invariant.

Theorem 5.6.1 *Let X and Y be compact, connected orientable n-manifolds and suppose $f, g : X \longrightarrow Y$ are two smooth maps that are smoothly homotopic. Then*

$$\deg(f) = \deg(g).$$

Proof: According to Corollary 5.2.3, f and g induce the same maps in cohomology:

$$f^\# = g^\# : H^n_{de\,R}(Y) \longrightarrow H^n_{de\,R}(X).$$

Thus, given any normalized generator $\boldsymbol{\omega}_0$ for $H^n_{de\,R}(Y)$ we have $f^\#([\boldsymbol{\omega}_0]) = g^\#([\boldsymbol{\omega}_0])$ so $[f^*\boldsymbol{\omega}_0] = [g^*\boldsymbol{\omega}_0]$ and, consequently,

$$f^*\boldsymbol{\omega}_0 = g^*\boldsymbol{\omega}_0 + d\boldsymbol{\eta}$$

for some $\boldsymbol{\eta} \in \Lambda^{n-1}(X)$. Stokes' Theorem then gives

$$\int_X f^*\boldsymbol{\omega}_0 = \int_X g^*\boldsymbol{\omega}_0,$$

i.e., $\deg(f) = \deg(g)$. ■

Since a constant map from X to Y obviously has degree 0 we obtain the following consequence of Theorem 5.6.1.

Corollary 5.6.2 *Let X and Y be compact, connected orientable n-manifolds. Then any smoothly nullhomotopic map from X to Y has degree zero.*

Corollary 5.6.3 *Let X be a compact, connected orientable manifold. Then $id_X : X \longrightarrow X$ is not smoothly nullhomotopic (so X is not smoothly contractible).*

Exercise 5.6.4 Prove Corollary 5.6.3. ■

We now consider the special case in which the range of our map is a sphere. Specifically, we let X be a compact, connected, orientable n-manifold and

$$f : X \longrightarrow S^n$$

a smooth map. We claim first that if f is *not* surjective, then it must be smoothly nullhomotopic and therefore have degree zero. To see this we proceed as follows: Since f is not surjective we may assume, without loss of generality, that $f(X)$ does not contain the north pole N of S^n (just choose the coordinate system in $\mathbb{R}^{n+1}$ properly). Consequently, f maps into the domain of the stereographic projection φ_S and we can define a smooth map

$$\varphi_S \circ f : X \longrightarrow \mathbb{R}^n.$$

Exercise 5.6.5 Show that $\varphi_S \circ f$ is smoothly nullhomotopic. **Hint**: Consider the map $H(x, t) = (1 - t)(\varphi_S(f(x)))$.

Exercise 5.6.6 Show that f is smoothly nullhomotopic. **Hint**: Consider $\varphi_S^{-1} \circ H$, where H is as in Exercise 5.6.5.

Theorem 5.6.4 *Let X be a compact, connected orientable n-manifold and $f : X \longrightarrow S^n$ a nonsurjective smooth map. Then f is smoothly nullhomotopic and so $\deg(f) = 0$.*

Exercise 5.6.7 Let X be a compact, connected orientable manifold and $f : X \longrightarrow X$ a diffeomorphism of X onto itself. Show that $\deg(f)$ is 1 if f is orientation preserving and -1 if f is orientation reversing.

Exercise 5.6.8 Any element A of $O(n + 1)$ determines a linear map of $\mathbb{R}^{n+1}$ onto itself which, when restricted to the unit sphere, gives a diffeomorphism of S^n onto S^n. Show that the degree of this map is $\det A$. Conclude that the degree of the **antipodal map** $A : S^n \longrightarrow S^n$ given by $A(p) = -p$ is $(-1)^{n+1}$.

Remark: In Section 3.4 of **[N4]** there is a definition of the Brouwer degree of a continuous map from S^n to S^n based on the fact that the $\mathrm{n^{th}}$ singular homology of S^n is isomorphic to $\mathbb{Z}$. In this case the degree is an integer by definition. Our current definition in the smooth category does not obviously result in an integer; it *does* result in an integer, but just not obviously so. We will prove this now. What we will not prove (but is true nonetheless) is that, for smooth maps from S^n to S^n, the two definitions give the same integer.

We conclude our discussion of the degree of a map by describing another way to calculate it and, in the process, showing that it is always an integer. Thus, we consider two compact, connected, oriented n-manifolds X and Y and a smooth map $f : X \longrightarrow Y$. Let $q \in Y$ be a regular value of f.

Remark: According to Theorem 3 – 14 of [**Sp1**], the set of critical values of a smooth map $g : A \longrightarrow \mathbb{R}^n$, A open in $\mathbb{R}^n$, has measure zero in $\mathbb{R}^n$. In particular, regular values always exist. Applying this to any coordinate expression for f we find that it too must have regular values.

Then $f^{-1}(q)$ is a compact submanifold of X of dimension $n - n = 0$, i.e., a finite set of points $\{p_1, \ldots, p_k\}$ (perhaps $\emptyset$). If $f^{-1}(p) = \{p_1, \ldots, p_k\} \neq \emptyset$, define, for each $i = 1, \ldots, k$, the **sign of** f **at** p_i by

$$\operatorname{sign}(f,\ p_i) = \begin{cases} 1 & \text{if} \quad f_{*p_i} \quad \text{is orientation preserving} \\ -1 & \text{if} \quad f_{*p_i} \quad \text{is orientation reversing} \end{cases}$$

(note that $f_{*p_i} : T_{p_i}(X) \longrightarrow T_{f(\mathrm{p}_i)}(Y)$ is an isomorphism). We claim that

$$\deg(f) = \begin{cases} \displaystyle\sum_{i=1}^{k} \operatorname{sign}\,(f, p_i), & \text{if } f^{-1}(q) = \{p_1, \ldots, p_k\} \neq \emptyset \\ \\ 0, & \text{if } f^{-1}(q) = \emptyset \end{cases} . \tag{5.6.3}$$

In particular, $\deg(f)$ is an integer (and is clearly zero whenever f is not surjective since any $q \in Y - f(X)$ is a regular value). Observe also that, although it is not obvious at the moment that the integer on the right-hand side of (5.6.3) is independent of the choice of q, this will follow once (5.6.3) is proved.

To prove (5.6.3) we first suppose $f^{-1}(q) = \emptyset$. Since X is compact, $f(X)$ is closed in Y so $Y - f(X)$ is an open set. $Y - f(X)$ is nonempty (since it contains q) and so it contains an open set U diffeomorphic to $\mathbb{R}^n$. Lemma 5.5.2 provides an n-form $\boldsymbol{\omega}_0$ on Y with support contained in U such that $\int_Y \boldsymbol{\omega}_0 = 1$. In particular, $\boldsymbol{\omega}_0$ is zero on $f(X)$ so $f^*\boldsymbol{\omega}_0$ is the zero n-form on X. Thus,

$$\deg(f) = \int_X f^*\boldsymbol{\omega}_0 = 0.$$

Now suppose $f^{-1}(q) = \{p_1, \ldots, p_k\}$. Choose coordinate neighborhoods $U_1, \ldots, U_k$ in X such that $p_i \in U_i$, $i = 1, \ldots, k$, $U_i \cap U_j = \emptyset$ if $i \neq j$

and such that f is a diffeomorphism on each U_i (since q is a regular value and $\dim X = \dim Y$, the Inverse Function Theorem applies at each p_i). We now wish to find a coordinate neighborhood V of q in Y such that $f^{-1}(V) = U_1 \cup \cdots \cup U_k$. First choose some compact neighborhood W of q in Y and consider

$$W' = f^{-1}(W) - (U_1 \cup \cdots \cup U_k).$$

This is closed in X and therefore compact. Consequently, $f(W')$ is a closed set in Y that does not contain q. Select a coordinate neighborhood V of q with $V \subseteq W - f(W')$. Then $f^{-1}(V) \subseteq U_1 \cup \cdots \cup U_k$. Now, if necessary, replace each U_i by $U_i \cap f^{-1}(V)$ (still a coordinate neighborhood) so that

$$f^{-1}(V) = U_1 \cup \cdots \cup U_k.$$

Since V must contain an open set diffeomorphic to $\mathbb{R}^n$, Lemma 5.5.2 gives an n-form $\boldsymbol{\omega}_0$ on Y with support contained in V and $\int_Y \boldsymbol{\omega}_0 = 1$. Moreover, $\operatorname{supp}(f^*\boldsymbol{\omega}_0) \subseteq U_1 \cup \cdots \cup U_k$ so, since this union is disjoint,

$$\int_X f^*\boldsymbol{\omega}_0 = \sum_{i=1}^{k} \int_{U_i} f^*\boldsymbol{\omega}_0.$$

But f is a diffeomorphism on each U_i so

$$\begin{aligned} \int_{U_i} f^*\boldsymbol{\omega}_0 &= \pm \int_{f(U_i)} \boldsymbol{\omega}_0 \\ &= \pm \int_V \boldsymbol{\omega}_0 = \pm \int_Y \boldsymbol{\omega}_0 = \pm 1, \end{aligned}$$

where the plus sign is chosen if f is orientation preserving on U_i and the minus sign is chosen if f is orientation reversing on U_i. Thus,

$$\begin{aligned} \int_{U_i} f^*\boldsymbol{\omega}_0 &= \begin{cases} 1 & \text{if} \quad f_{*p_i} \quad \text{is orientation preserving} \\ -1 & \text{if} \quad f_{*p_i} \quad \text{is orientation reversing} \end{cases} \\ &= \operatorname{sign}(f, p_i) \end{aligned}$$

so the result follows. ∎

The formula (5.6.3) for $\deg(f)$ provides some intuitive insight into the geometrical significance of the Brouwer degree that is not apparent from our definition (5.6.1). For any regular value q of f, $f^{-1}(q)$ is a finite set of points, but the number of elements in it (i.e., the number of times q is "covered" by f) generally depends on q. Thus, f generally does not cover all of its regular values the same number of times. However, if we "count" the preimages of a regular value properly (with multiplicity 1 when f is orientation preserving there and with multiplicity -1 when f is orientation reversing there) the result is the same for every regular value. This counting with multiplicity simply cancels two preimages if f maps them onto q in "opposite directions" (picture this for maps from S^1 to S^1). Thus, one thinks of $\deg(f)$ as the "net" number of times f covers each of its regular values.

Exercise 5.6.9 Identify S^1 with the set of complex numbers z of modulus 1 and let n be a positive integer. Define $f : S^1 \longrightarrow S^1$ and $g : S^1 \longrightarrow S^1$ by $f(z) = z^n$ and $g(z) = \bar{z}^n$. Show that $\deg(f) = n$ and $\deg(g) = -n$.

Remark: Identifying S^3 with the set of quaternions q of modulus 1 one can similarly define, for any positive integer n, $f : S^3 \longrightarrow S^3$ and $g : S^3 \longrightarrow S^3$ by $f(q) = q^n$ and $g(q) = \bar{q}^{\,n}$. In Section 6.4 we will compute the degrees of these maps and find that $\deg(f) = n$ and $\deg(g) = -n$.

The Brouwer degree is a homotopy invariant of maps between compact, connected, orientable n-manifolds. We will conclude this chapter with a very brief look at another homotopy invariant for maps from S^{2n-1} to S^n, $n \geq 2$, which has played a significant role in the development of modern topology.

5.7 The Hopf Invariant

Consider a smooth map $f : S^{2n-1} \longrightarrow S^n$, $n \geq 2$, and let $\boldsymbol{\omega}_0$ be a normalized generator for $H^n_{de\,R}(S^n)$. Then $f^*\boldsymbol{\omega}_0$ is an n-form on S^{2n-1}. But $H^n_{de\,R}(S^{2n-1}) \cong 0$ so there exists an $(n-1)$-form $\boldsymbol{\omega}$ on S^{2n-1} such that $d\boldsymbol{\omega} = f^*\boldsymbol{\omega}_0$. We define the **Hopf invariant** of f by

$$H(f) = \int_{S^{2n-1}} \boldsymbol{\omega} \wedge d\boldsymbol{\omega}.$$

To see that the definition does not depend on the choice of $\boldsymbol{\omega}$, suppose $\boldsymbol{\omega}'$ is another $(n-1)$-form on S^{2n-1} with $d\boldsymbol{\omega}' = f^*\boldsymbol{\omega}_0$. Then $d(\boldsymbol{\omega} - \boldsymbol{\omega}') = 0$ so

$$\begin{aligned}\int_{S^{2n-1}} \boldsymbol{\omega} \wedge d\boldsymbol{\omega} - \int_{S^{2n-1}} \boldsymbol{\omega}' \wedge d\boldsymbol{\omega}' &= \int_{S^{2n-1}} (\boldsymbol{\omega} - \boldsymbol{\omega}') \wedge d\boldsymbol{\omega} \\ &= \int_{S^{2n-1}} d\left((\boldsymbol{\omega} - \boldsymbol{\omega}') \wedge \boldsymbol{\omega}\right) = 0\end{aligned}$$

by Stokes' Theorem.

Exercise 5.7.1 Show that if n is odd, then $H(f) = 0$ for any smooth map $f : S^{2n-1} \longrightarrow S^n$. **Hint**: Compute $d(\boldsymbol{\omega} \wedge \boldsymbol{\omega})$ for any even dimensional form $\boldsymbol{\omega}$.

We show now that the Hopf invariant is a smooth homotopy invariant.

Theorem 5.7.1 *Let $f_0, f_1 : S^{2n-1} \longrightarrow S^n, n \geq 2,$ be smoothly homotopic maps. Then $H(f_0) = H(f_1)$.*

Proof: Let $F : S^{2n-1} \times (0-\epsilon,\ 1+\epsilon) \longrightarrow S^n$, be a smooth map with $F(x,0) = f_0(x)$ and $F(x,1) = f_1(x)$ for all $x \in S^{2n-1}$. Also let

$$\iota_0 : S_0 = S^{2n-1} \times \{0\} \hookrightarrow S^{2n-1} \times (0-\epsilon,\ 1+\epsilon)$$

and

$$\iota_1 : S_1 = S^{2n-1} \times \{1\} \hookrightarrow S^{2n-1} \times (0-\epsilon,\ 1+\epsilon)$$

be the inclusion maps. Identifying S_0 and S_1 with S^{2n-1} we may also identify $F \circ \iota_0$ with f_0 and $F \circ \iota_1$ with f_1, respectively. Now let $\boldsymbol{\omega}_0$ be a normalized generator for $H^n_{de\,R}(S^n)$. Then $F^*\boldsymbol{\omega}_0$ is an n-form on $S^{2n-1} \times (0-\epsilon, 1+\epsilon)$. By Theorem 5.2.6, $H^n_{de\,R}(S^{2n-1} \times (0-\epsilon, 1+\epsilon)) \cong H^n_{de\,R}(S^{2n-1})$ and, since $n \geq 2$, this is trivial. Consequently, there is an $(n-1)$-form $\boldsymbol{\eta}$ on $S^{2n-1} \times (0-\epsilon, 1+\epsilon)$ such that

$$F^*\boldsymbol{\omega}_0 = d\boldsymbol{\eta}.$$

Let $\iota_0^*\boldsymbol{\eta} = \boldsymbol{\omega}$ and $\iota_1^*\boldsymbol{\eta} = \boldsymbol{\omega}'$. Notice that

$$f_0^*\boldsymbol{\omega}_0 = (F \circ \iota_0)^*\boldsymbol{\omega}_0 = \iota_0^*(F^*\boldsymbol{\omega}_0) = \iota_0^*(d\boldsymbol{\eta}) = d(\iota_0^*\boldsymbol{\eta}) = d\boldsymbol{\omega}$$

and, similarly,

$$f_1^*\boldsymbol{\omega}_0 = d\boldsymbol{\omega}'.$$

Moreover,

$$\boldsymbol{\omega} \wedge d\boldsymbol{\omega} = (\iota_0^*\boldsymbol{\eta}) \wedge d(\iota_0^*\boldsymbol{\eta}) = (\iota_0^*\boldsymbol{\eta}) \wedge \iota_0^*(d\boldsymbol{\eta}) = \iota_0^*(\boldsymbol{\eta} \wedge d\boldsymbol{\eta})$$

and, similarly,

$$\boldsymbol{\omega}' \wedge d\boldsymbol{\omega}' = \iota_1^*(\boldsymbol{\eta} \wedge d\boldsymbol{\eta}).$$

We must show that

$$\int_{S_o} \boldsymbol{\omega} \wedge d\boldsymbol{\omega} = \int_{S_1} \boldsymbol{\omega}' \wedge d\boldsymbol{\omega}',$$

i.e., that

$$\int_{S_0} \iota_0^*(\boldsymbol{\eta} \wedge d\boldsymbol{\eta}) = \int_{S_1} \iota_1^*(\boldsymbol{\eta} \wedge d\boldsymbol{\eta}). \tag{5.7.4}$$

Now, $S^{2n-1} \times [0,1]$ is a domain with smooth boundary in $S^{2n-1} \times (0-\epsilon,\ 1+\epsilon)$ and its boundary is $\partial(S^{2n-1} \times [0,1]) = S_0 \cup S_1$. Note also that

$$d(\boldsymbol{\eta} \wedge d\boldsymbol{\eta}) = d\boldsymbol{\eta} \wedge d\boldsymbol{\eta} = F^*\boldsymbol{\omega}_0 \wedge F^*\boldsymbol{\omega}_0 = F^*(\boldsymbol{\omega}_0 \wedge \boldsymbol{\omega}_0) = 0$$

because $\boldsymbol{\omega}_0 \wedge \boldsymbol{\omega}_0$ is a $2n$-form on the n-dimensional manifold S^n.

Exercise 5.7.2 Prove (5.7.1) by applying Stokes' Theorem to the integral over $S^{2n-1} \times [0,1]$ of the restriction of $d(\boldsymbol{\eta} \wedge d\boldsymbol{\eta})$. **Hint**: S_0 and S_1 acquire opposite orientations from $S^{2n-1} \times [0,1]$. ∎

Exercise 5.7.3 Show that a nullhomotopic map $f : S^{2n-1} \longrightarrow S^n,\ n \geq 2$, has Hopf invariant zero.

Remark: As was the case for the Brouwer degree there is another, more geometrical way of calculating the Hopf invariant which shows that $H(f)$ is actually an integer. This involves the *linking number* of the preimages $f^{-1}(p)$ and $f^{-1}(q)$ of any two distinct regular values of f (see Chapter III, Section 17, of [**BT**]). Still other approaches to the Hopf invariant are described in [**Huse**].

We conclude with a few exercises that will lead the reader through the calculation of the Hopf invariant for the projection map

$$\mathcal{P}_1 : S^3 \longrightarrow S^2$$

of the complex Hopf bundle. The result will be $H(\mathcal{P}_1) = 1$ so that, in particular, Exercise 5.7.3 then implies that $\mathcal{P}_1$ is not nullhomotopic. This was, in fact, the first example of a homotopically nontrivial map of one sphere onto another of smaller dimension and came as quite a shock in the 1930s when Hopf constructed it. Recall that if S^3 is identified with the subset of $\ ^2$ consisting of those $(z_1,\ z_2)$ for which $|z_1|^2 + |z_2|^2 = 1$ and S^2 is identified with a subset of $\mathbb{R}^3$, then

$$\mathcal{P}_1(z_1,\ z_2) = \left(z^1\ \bar{z}^2 + \bar{z}^1\ z^2, -\boldsymbol{i}z^1\ \bar{z}^2 + i\bar{z}^1\ z^2,\ |z^1|^2 - |z^2|^2\right).$$

For the calculation of the Hopf invariant we will take the normalized generator $\boldsymbol{\omega}_0$ of $H^n_{de\,R}(S^2)$ to be the standard volume form for S^2 divided by the "volume" 4π of S^2 (see Section 4.6). Recall that, if $x^1,\ x^2$ and x^3 are standard coordinates on $\mathbb{R}^3$ and $\iota : S^2 \hookrightarrow \mathbb{R}^3$ is the inclusion, then

$$\boldsymbol{\omega}_0 = \frac{1}{4\pi}\iota^*(x^1dx^2 \wedge dx^3 - x^2dx^1 \wedge dx^3 + x^3dx^1 \wedge dx^2).$$

Exercise 5.7.4 Denote the standard coordinates on $\mathbb{R}^4$ by $y^1,\ y^2,\ y^3,\ y^4$ and define $p : \mathbb{R}^4 \longrightarrow \mathbb{R}^3$ by

$$p\ (y^1,\ y^2,\ y^3,\ y^4) = \Big(2y^1y^3 + 2y^2y^4, 2y^2y^3 - 2y^1y^4,$$
$$\left(y^1\right)^2 + \left(y^2\right)^2 - \left(y^3\right)^2 - \left(y^4\right)^2\Big).$$

Identify S^3 with a subset of $\mathbb{R}^4$ via $(z^1, z^2) = (y^1 + \boldsymbol{i}y^2,\ y^3 + \boldsymbol{i}y^4) \longrightarrow (y^1,\ y^2,\ y^3,\ y^4)$ and let $\iota' : S^3 \hookrightarrow \mathbb{R}^4$ be the inclusion. Show that $p \circ \iota' = \iota \circ \mathcal{P}_1$ and therefore

$$\mathcal{P}_1^*\boldsymbol{\omega}_0 = \frac{1}{4\pi}(\iota')^*\left(p^*\left(x^1dx^2 \wedge dx^3 - x^2dx^1 \wedge dx^3 + x^3dx^1 \wedge dx^2\right)\right).$$

Exercise 5.7.5 Show that

$$\mathcal{P}_1^* \boldsymbol{\omega}_0 = -\frac{1}{\pi}(\iota')^*(dy^1 \wedge dy^2 + dy^3 \wedge dy^4).$$

Exercise 5.7.6 Show that $dy^1 \wedge dy^2 + dy^3 \wedge dy^4 = d(y^1 dy^2 + y^3 dy^4)$. Then define $\boldsymbol{\omega}$ by $\boldsymbol{\omega} = -\frac{1}{\pi}(\iota')^*(y^1 dy^2 + y^3 dy^4)$ and show that $d\boldsymbol{\omega} = P_1^* \boldsymbol{\omega}_0$.

Exercise 5.7.7 Show that $\boldsymbol{\omega} \wedge d\boldsymbol{\omega} = \dfrac{1}{\pi^2}(\iota')^*(y^1 dy^2 \wedge dy^3 \wedge dy^4 + y^3 dy^1 \wedge dy^2 \wedge dy^4)$.

Exercise 5.7.8 Compute

$$H(\mathcal{P}_1) = \int_{S^3} \boldsymbol{\omega} \wedge d\boldsymbol{\omega} = 1.$$

6
Characteristic Classes

6.1 Motivation

We have had a number of previous, albeit informal encounters with the notion of a characteristic class for a principal bundle (in Section 2.2, 2.4 and 2.5) and are now in possession of sufficient machinery to take up the subject in earnest. Before plunging into the thick of the battle, however, we would like to make clear the strategy we propose to adopt. This is perhaps best achieved by going through what might be regarded as the "trivial" case in detail and then indicating the modifications required when matters are not so simple.

We begin then by considering a principal $U(1)$-bundle $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ over a manifold X and let $\boldsymbol{\omega}$ denote a connection on the bundle. Thus, one might have in mind an electromagnetic field on spacetime. Since $U(1)$ is Abelian, the curvature is given by $\boldsymbol{\Omega} = d\boldsymbol{\omega}$. For any local cross-section $s : V \longrightarrow \mathcal{P}^{-1}(V)$ we write, as usual, $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{\omega} = -\boldsymbol{i}\boldsymbol{A}$ and $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega} = -\boldsymbol{i}\boldsymbol{F}$ for the local gauge potential and field strength. Again because $U(1)$ is Abelian, the field strengths $\boldsymbol{\mathcal{F}}$ for various cross-sections s agree on the intersections of their domains and thereby determine a globally defined $u(1)$-valued 2-form on X that we also denote by $\boldsymbol{\mathcal{F}}$. This $\boldsymbol{\mathcal{F}}$ is $-\boldsymbol{i}\boldsymbol{F}$ for a globally defined real-valued 2-form $\boldsymbol{F}$ on X. Since $d\boldsymbol{\mathcal{F}} = 0$, $\boldsymbol{F}$ is closed and therefore $\frac{1}{2\pi}\boldsymbol{F}$ determines an element of $H^2_{\mathrm{de\,R}}(X;\quad)$.

Now suppose $\boldsymbol{\omega}'$ is another connection on the same bundle $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ determining global 2-forms $\boldsymbol{\mathcal{F}}'$ and $\boldsymbol{F}'$ in the same way. It follows that $\boldsymbol{\omega} - \boldsymbol{\omega}' = \boldsymbol{\tau}$ is in $\Lambda^1_{ad}(P, u(1))$. Thus, $d\boldsymbol{\omega} - d\boldsymbol{\omega}' = d\boldsymbol{\tau}$ so

$$\boldsymbol{\Omega} - \boldsymbol{\Omega}' = d\boldsymbol{\tau}.$$

Since $\boldsymbol{\tau}$ is tensorial of type ad and $U(1)$ is Abelian, $\sigma_g^*\boldsymbol{\tau} = g^{-1}\cdot\boldsymbol{\tau} = g\boldsymbol{\tau}g^{-1} = \boldsymbol{\tau}$ for every $g \in U(1)$. But then Lemma 4.5.1. implies that $\boldsymbol{\tau}$ projects to X, i.e., that there exists a unique $u(1)$-valued 1-form $\bar{\boldsymbol{\tau}}$ on X such that $\boldsymbol{\tau} = \mathcal{P}^*\bar{\boldsymbol{\tau}}$. Since (in the Abelian case) curvatures also project to X, $\boldsymbol{\Omega} = \mathcal{P}^*\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\Omega}' = \mathcal{P}^*\boldsymbol{\mathcal{F}}'$ and therefore $\mathcal{P}^*\boldsymbol{\mathcal{F}} - \mathcal{P}^*\boldsymbol{\mathcal{F}}' = d(\mathcal{P}^*\bar{\boldsymbol{\tau}})$ so

$$\mathcal{P}^*(\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{F}}') = \mathcal{P}^*(d\bar{\boldsymbol{\tau}}).$$

But projections to X are unique when they exist so we must have

$$\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{F}}' = d\bar{\boldsymbol{\tau}}.$$

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0_6,

In particular, $\frac{1}{2\pi}\boldsymbol{F}$ and $\frac{1}{2\pi}\boldsymbol{F}'$ are cohomologous and therefore determine the *same* element of $H^2_{\text{de R}}(X;\quad)$. The unique cohomology class determined in this way from any connection on $U(1) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is called the 1^{st} *Chern class* of the bundle and is denoted

$$c_1(P) = \left[\frac{1}{2\pi}\boldsymbol{F}\right] \in H^2_{\text{de R}}(X;\quad).$$

The 1^{st} Chern class of our $U(1)$-bundle is determined entirely by the bundle itself and not by the particular connection on that bundle from which it is constructed. It is true, but by no means obvious, that $c_1(P)$ is actually "characteristic" of the bundle in the sense that two principal $U(1)$-bundles over the same manifold X are equivalent if and only if their 1^{st} Chern classes are the same. Since this is easy to prove when X is a sphere S^n, we shall do so. The only interesting case is $n = 2$. The reason is that, for $n = 1, 3, 4, 5, \ldots$, every $U(1)$-bundle over S^n is trivial (for $n = 1$, see Exercise 4.4.11 of [**N4**]; for $n = 3, 4, 5, \ldots$, use the Classification Theorem for bundles over spheres and Corollary 2.5.11 of [**N4**] which asserts that $\pi_n(U(1))$ is trivial for $n \geq 2$). Since a trivial bundle admits a flat connection (Exercise 6.2.12, [**N4**]), its 1^{st} Chern class must be zero so we have the (admittedly, rather silly) result that two $U(1)$-bundles over S^n, $n = 1, 3, 4, 5, \ldots$, are equivalent if and only if their 1^{st} Chern classes are the same. The $n = 2$ case takes just slightly more work.

We know all of the principal $U(1)$-bundles over S^2. They are in one-to-one correspondence with the elements of $\pi_1(U(1)) \cong \mathbb{Z}$ (Classification Theorem). One can describe them explicitly in terms of transition functions as follows: Trivialize the bundle over U_N and U_S and consider the map $g_{SN} : U_N \cap U_S \longrightarrow U(1)$ given by $g_{SN}(\varphi, \theta) = e^{-n\theta \boldsymbol{i}}$, where $n \in \mathbb{Z}$ and (φ, θ) are standard spherical coordinates on S^2. This map determines a principal $U(1)$-bundle over S^2 for each n and two such bundles are equivalent if and only if they correspond to the same n. There are explicit descriptions of all of these bundles in Section 2.2 where we have also written out a connection (Dirac Monopole) on each. These, in fact, exhaust all of the (equivalence classes of) principal $U(1)$-bundles over S^2. The reason is that any such bundle is determined up to equivalence by the homotopy class of its characteristic map $g_{SN} \,|\, S^1$ (Theorem 4.4.2, [**N4**]) and every map from S^1 to S^1 is homotopic to $e^{-n\theta \boldsymbol{i}}$ for some n (see the proof of Theorem 2.4.4, [**N4**]). To show that these bundles are also determined up to equivalence by their 1^{st} Chern classes we will show that the integer n can be retrieved from $c_1(P)$ just by integrating it over S^2.

Any representative of the cohomology class $c_1(P)$ is a smooth 2-form on the compact manifold S^2 and so is integrable. Furthermore, all such representatives have the same integral over S^2 by Stokes' Theorem. Since we know a connection on the bundle with transition function $g_{SN}(\varphi, \theta) = e^{-n\theta \boldsymbol{i}}$ we could simply compute the corresponding representative and integrate it. It is much more instructive, however, to proceed indirectly. Thus, we assume only that we have some connection on the bundle with local gauge potentials $\boldsymbol{\mathcal{A}}_N$ and $\boldsymbol{\mathcal{A}}_S$ on U_N and U_S, respectively. Then, on $U_N \cap U_S$,

$$\boldsymbol{\mathcal{A}}_N = g_{SN}^{-1}\,\boldsymbol{\mathcal{A}}_S g_{SN} + g_{SN}^{-1}\,dg_{SN}$$

so

$$\boldsymbol{\mathcal{A}}_N = \boldsymbol{\mathcal{A}}_S - \boldsymbol{i}\,n\,d\,\theta.$$

The corresponding field strengths are denoted $\boldsymbol{\mathcal{F}}_N$ and $\boldsymbol{\mathcal{F}}_S$, respectively. To integrate $\frac{1}{2\pi}\boldsymbol{F} = \frac{\boldsymbol{i}}{2\pi}\boldsymbol{\mathcal{F}}$ over S^2 we will write $S^2 = S^2_+ \cup S^2_-$ as the union of the upper $(z \geq 0)$ and lower $(z \leq 0)$ hemispheres and denote by S^1 the equatorial $(z = 0)$ circle with $\iota : S^1 \hookrightarrow S^2$ the inclusion map. Provide S^1 with the orientation it inherits from the standard orientation of S^2_+ ("counterclockwise") and note that it receives the opposite orientation from S^2_-. Applying Stokes' Theorem twice then gives

$$\begin{aligned}
\int_{S^2} \frac{\boldsymbol{i}}{2\pi}\boldsymbol{\mathcal{F}} &= \frac{\boldsymbol{i}}{2\pi}\int_{S^2_+} \boldsymbol{\mathcal{F}}_N + \frac{\boldsymbol{i}}{2\pi}\int_{S^2_-} \boldsymbol{\mathcal{F}}_S \\
&= \frac{\boldsymbol{i}}{2\pi}\int_{S^1} \iota^*\boldsymbol{\mathcal{A}}_N - \frac{\boldsymbol{i}}{2\pi}\int_{S^1} \iota^*\boldsymbol{\mathcal{A}}_S \\
&= \frac{\boldsymbol{i}}{2\pi}\int_{S^1} \iota^*(\boldsymbol{\mathcal{A}}_S - \boldsymbol{i}\,n\,d\,\theta) - \frac{\boldsymbol{i}}{2\pi}\int_{S^1} \iota^*\boldsymbol{\mathcal{A}}_S \\
&= \frac{\boldsymbol{i}}{2\pi}\int_{S^1} \iota^*\boldsymbol{\mathcal{A}}_S + \frac{n}{2\pi}\int_{S^1} \iota^* d\,\theta - \frac{\boldsymbol{i}}{2\pi}\int_{S^1} \iota^*\boldsymbol{\mathcal{A}}_S \\
&= \frac{n}{2\pi}\int_{S^1} d(\theta \circ \iota).
\end{aligned}$$

The essential feature of the calculation to this point is that all references to the connection itself have dropped out and we are left with an integral that involves only the transition function ($g_{SN}^{-1}\,dg_{SN} = -\boldsymbol{i}\,n\,d\,\theta$). This is as it should be, of course, since the Chern class (and therefore its integral) depends only on the bundle and not on the connection. Calculating the remaining integral is easy and, in fact, was done explicitly in Section 4.6. The result was simply the length ("volume") of S^1:

$$\int_{S^1} d(\theta \circ \iota) = 2\pi.$$

Thus

$$\int_{S^2} c_1(P) = \frac{n}{2\pi}(2\pi) = n$$

as promised.

This then concludes our discussion of the "trivial" case and the issue before us now is whether or not it is possible to push through an analogous program for bundles with other structure groups. On the surface, the prospects do not appear to be good since virtually every stage in our construction depended on the commutativity of $U(1)$. It is not even altogether clear how to get the ball rolling since the very existence of the 2-form on X whose cohomology class contained the topological information about the bundle required that local field strengths agree on the intersections of their domains. Since, in general, such local field strengths are related by $\boldsymbol{\mathcal{F}}^g = g^{-1}\boldsymbol{\mathcal{F}}g$, this simply is not true when the structure group is non-Abelian. The key idea here is to evade this difficulty by considering, not the field strengths themselves, but various functions of the field strength that cannot tell the difference between $\boldsymbol{\mathcal{F}}$ and $g^{-1}\boldsymbol{\mathcal{F}}g$. Identifying all of these objects with matrices there are a number of obvious choices. For example, since the trace of a matrix is invariant under conjugation, $\operatorname{trace}(g^{-1}\boldsymbol{\mathcal{F}}g) = \operatorname{trace}\boldsymbol{\mathcal{F}}$ and the 2-forms defined locally on X by taking the trace of the local field strengths will agree on the intersections of their domains and therefore determine a globally defined 2-form $\operatorname{trace}\boldsymbol{\mathcal{F}}$ on X. Perhaps $\operatorname{trace}\boldsymbol{\mathcal{F}}$ is closed and so determines a cohomology class. Perhaps this cohomology class is independent of the connection from which $\operatorname{trace}\boldsymbol{\mathcal{F}}$ was constructed. Perhaps the same is true for other *ad*-invariant functions of $\boldsymbol{\mathcal{F}}$ such as $\operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}})$, or $\det \boldsymbol{\mathcal{F}}$. Perhaps this is all a bit too much to ask. We shall see.

6.2 Algebraic Preliminaries

We begin by considering a finite dimensional real vector space $\mathcal{V}$ (typically, this will be a Lie algebra for some matrix Lie group). If $k \geq 1$ is an integer, then a map $\tilde{f} : \mathcal{V} \times \overset{k}{\cdots} \times \mathcal{V} \longrightarrow \quad$ is ***k*-multilinear** if it is linear in each variable separately and **symmetric** if $\tilde{f}(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = \tilde{f}(v_1, \ldots, v_k)$ for all $v_1, \ldots, v_k$ in $\mathcal{V}$ and all permutations $\sigma \in S_k$. The real and imaginary parts of such an $\tilde{f}$ are therefore just symmetric k-tensors on $\mathcal{V}$. The set $S^k(\mathcal{V})$ of all such maps has the obvious pointwise structure of a complex vector space and, setting $S^0(\mathcal{V}) = \quad$, we define

$$S(\mathcal{V}) = \bigoplus_{k=0}^{\infty} S^k(\mathcal{V}),$$

(see the Remark on page 302). If $\tilde{f} \in S^k(\mathcal{V})$ and $\tilde{g} \in S^l(\mathcal{V})$ we define $\tilde{f} \odot \tilde{g}$ by

$$\begin{aligned} &\left(\tilde{f} \odot \tilde{g}\right)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_l) \\ &= \frac{1}{(k+l)!} \sum_{\tau \in S_{k+l}} \tilde{f}\left(v_{\tau(1)}, \ldots, v_{\tau(k)}\right) \tilde{g}\left(v_{\tau(k+1)}, \ldots, v_{\tau(k+l)}\right). \end{aligned} \tag{6.2.1}$$

Exercise 6.2.1 Show that $\tilde{f} \odot \tilde{g} \in S^{k+l}(\mathcal{V})$. Extend $\odot$ to $S(\mathcal{V})$ and show that, with this operation, $S(\mathcal{V})$ has the structure of a commutative algebra with identity.

If $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{V}$ and $\{e^1, \ldots, e^n\}$ is the dual basis for $\mathcal{V}^*$, then for any $\tilde{f} \in S^k(\mathcal{V})$, there exist unique complex numbers $a_{i_1 \ldots i_k}$, $i_1, \ldots, i_k = 1, \ldots, n$, symmetric in $i_1, \ldots, i_k$, such that

$$\tilde{f} = a_{i_1 \cdots i_k}\, e^{i_1} \otimes \cdots \otimes e^{i_k} \qquad \text{(summation convention).}$$

Writing $v_1 = x_1^{i_1} e_{i_1}, \ldots, v_k = x_k^{i_k} e_{i_k}$ we therefore have

$$\tilde{f}(v_1, \ldots, v_k) = a_{i_1 \cdots i_k}\, x_1^{i_1} \cdots x_k^{i_k}.$$

Now define $f \colon \mathcal{V} \longrightarrow$ by $f(v) = \tilde{f}(v, \ldots, v)$ so that, if $v = x^i e_i$,

$$f(v) = a_{i_1 \cdots i_k}\, x^{i_1} \cdots x^{i_k}.$$

Thus, f is a homogeneous polynomial of degree k in the components of v. For convenience we write

$$f = a_{i_1 \cdots i_k} e^{i_1} \cdots e^{i_k},$$

where the function on the right-hand side is defined by

$$\Big(a_{i_1 \cdots i_k} e^{i_1} \cdots e^{i_k}\Big)(v) = a_{i_1 \cdots i_k} e^{i_1}(v) \cdots e^{i_k}(v) = a_{i_1 \cdots i_k} x^{i_1} \cdots x^{i_k}.$$

More generally, let us say that a map $f : \mathcal{V} \longrightarrow$ is a **homogeneous polynomial of degree** k **on** $\mathcal{V}$ if it can be expressed as a homogeneous polynomial of degree k in $e^1, \ldots, e^n$ in the sense that there exist $a_{i_1 \cdots i_k} \in$, $i_1, \ldots, i_k = 1, \ldots, n$, symmetric in $i_1, \ldots, i_k$, such that $f = a_{i_1 \cdots i_k} e^{i_1} \cdots e^{i_k}$.

Exercise 6.2.2 Show that this definition does not depend on the choice of basis.

The collection of all such homogeneous polynomials of degree k on $\mathcal{V}$, with its obvious pointwise linear structure, is denoted $P^k(\mathcal{V})$. Setting $P^0(\mathcal{V}) =$ we define

$$P(\mathcal{V}) = \bigoplus_{k=0}^{\infty} P^k(\mathcal{V}).$$

Exercise 6.2.3 Show that, if $f \in P^k(\mathcal{V})$ and $g \in P^l(\mathcal{V})$, then the product fg defined by $(fg)(v) = f(v)g(v)$ is in $P^{k+l}(\mathcal{V})$. Extend this product structure to $P(\mathcal{V})$ and show that $P(\mathcal{V})$ thereby becomes a commutative algebra with identity.

We show next that the two algebras $S(\mathcal{V})$ and $P(\mathcal{V})$ are, in fact, isomorphic. Begin by fixing a $k \geq 1$. Then any $\tilde{f} \in S^k(\mathcal{V})$ gives rise to an $f \in P^k(\mathcal{V})$ defined by $f(v) = \tilde{f}(v, \ldots, v)$. If $\{e_1, \ldots, e_n\}$ is any basis for $\mathcal{V}$ and $\{e^1, \ldots, e^n\}$ is the dual basis for $\mathcal{V}^*$, then this assignment is given by

$$a_{i_1\cdots i_k} e^{i_1} \otimes \cdots \otimes e^{i_k} \longrightarrow a_{i_1\cdots i_k} e^{i_1} \cdots e^{i_k}$$

and is clearly linear and surjective. To prove that this is an isomorphism we must show that $\tilde{f}$ is uniquely determined by f and this involves the familiar process of *polarization*. There are several useful ways to view this process and we will begin by illustrating in the $k=2$ case. Thus, suppose we are given an $f \in P^2(\mathcal{V})$. To show that the $\tilde{f} \in S^2(\mathcal{V})$ for which $f(v) = \tilde{f}(v,v)$ is unique we will prove that, in fact,

$$\tilde{f}(v,w) = \frac{1}{2}\Big[f(v+w) - f(v) - f(w)\Big]$$

for all $v, w \in \mathcal{V}$. Indeed, one need only compute $\tilde{f}(v+w, v+w)$ using bilinearity

$$\tilde{f}(v+w, v+w) = \tilde{f}(v,v) + 2\tilde{f}(v,w) + \tilde{f}(w,w)$$

to obtain

$$f(v+w) = f(v) + 2\tilde{f}(v,w) + f(w).$$

There are analogous formulas for larger k, e.g., for $f \in P^3(\mathcal{V})$,

$$\begin{aligned}\tilde{f}(u,v,w) = \frac{1}{6}\Big[&f(u+v+w) - f(u+v) - f(u+w)\\ &- f(v+w) + f(u) + f(v) + f(w)\Big].\end{aligned}$$

Another way of arriving at $\tilde{f}$ when $k=2$ is as follows: Write $f = a_{ij}e^i e^j$ and expand $f(sv+tw)$ as a polynomial in s and t.

$$\begin{aligned}f(sv+tw) &= a_{ij}(sv^i + tw^i)(sv^j + tw^j)\\ &= (a_{ij}v^i v^j)s^2 + (2a_{ij}v^i w^j)st \ + (a_{ij}w^i w^j)t^2\\ &= (f(v))s^2 + (2\tilde{f}(v,w))st \ + (f(w))t^2\end{aligned}$$

Thus, $\tilde{f}(v,w)$ is $\frac{1}{2}$ times the coefficient of st in the expansion of $f(sv+tw)$.

Exercise 6.2.4 Show that, in the general case, $\tilde{f}(v_1, \ldots, v_k)$ is $\frac{1}{k!}$ times the coefficient of $t_1 t_2 \cdots t_k$ in the expansion of $f(t_1 v_1 + t_2 v_2 + \cdots + t_k v_k)$.

The linear isomorphisms $\tilde{f} \longrightarrow f$ of $S^k(\mathcal{V})$ onto $P^k(\mathcal{V})$, together with the identity map of $S^0(\mathcal{V})$ onto $P^0(\mathcal{V})$, gives a linear isomorphism of $S(\mathcal{V})$ onto $P(\mathcal{V})$. To show that this is, in fact, an algebra isomorphism it will suffice to prove that, if $\tilde{f} \in S^k(\mathcal{V})$ with $\tilde{f} \longrightarrow f \in P^k(\mathcal{V})$ and $\tilde{g} \in S^l(\mathcal{V})$

with $\tilde{g} \longrightarrow g \in P^l(\mathcal{V})$, then $\tilde{f} \odot \tilde{g} \in S^{k+l}(\mathcal{V})$ is sent to $fg \in P^{k+l}(\mathcal{V})$. But $\tilde{f} \odot \tilde{g}$ is sent to the element of $P^{k+l}(\mathcal{V})$ whose value at v is, by (6.2.1),

$$\begin{aligned}\left(\tilde{f} \odot \tilde{g}\right)(v,\ldots,v,v,\ldots,v) &= \frac{1}{(k+l)!}\left[(k+l)!\tilde{f}(v,\ldots,v)\tilde{g}(v,\ldots,v)\right] \\ &= f(v)g(v) \\ &= (fg)(v)\end{aligned}$$

as required.

Now, suppose that G is a Lie group and $\rho : G \longrightarrow GL(\mathcal{V})$ is a representation of G on $\mathcal{V}$. For each $k \geq 1$, let $S^k_\rho(\mathcal{V})$ and $P^k_\rho(\mathcal{V})$ denote the subspaces of $S^k(\mathcal{V})$ and $P^k(\mathcal{V})$, respectively, that are invariant under ρ, i.e., satisfy

$$\tilde{f}(\rho(g)(v_1),\ldots,\rho(g)(v_k)) = \tilde{f}(v_1,\ldots,v_k)$$

and

$$f(\rho(g)(v)) = f(v)$$

for all $g \in G$ and all $v, v_1, \ldots, v_k$ in $\mathcal{V}$. The isomorphism $\tilde{f} \longrightarrow f$ clearly carries $S^k_\rho(\mathcal{V})$ onto $P^k_\rho(\mathcal{V})$. Furthermore, each of the products we have defined preserves ρ-invariance so the subspaces $S_\rho(\mathcal{V}) = \bigoplus_{k=0}^{\infty} S^k_\rho(\mathcal{V})$ and $P_\rho(\mathcal{V}) = \bigoplus_{k=0}^{\infty} P^k_\rho(\mathcal{V})$ of $S(\mathcal{V})$ and $P(\mathcal{V})$, respectively, are isomorphic subalgebras (as usual, the $k = 0$ summand is just).

The case of most immediate interest to us is that in which G is a matrix Lie group, $\mathcal{V}$ is its Lie algebra $\mathcal{G}$ and ρ is the adjoint representation $ad : G \longrightarrow GL(\mathcal{G})$. Then $S_{ad}(\mathcal{G})$ and $P_{ad}(\mathcal{G})$ are, respectively, the algebras of symmetric multilinear functions and polynomials that are invariant under conjugation by elements of G. It is customary to write $I^k(G)$ for $S^k_{ad}(\mathcal{G})$ and $I(G) = \bigoplus_{k=0}^{\infty} I^k(G)$ for $S_{ad}(G)$

Remark: The switch from $\mathcal{G}$ to G in this notation is not accidental. Two Lie groups with the same Lie algebra need not have the same *ad*-invariant multilinear forms since a given form can be invariant under conjugation by the elements of one of the groups, but not the other.

We will now construct a number of examples. In each case, G will denote an arbitrary matrix Lie group and its Lie algebra $\mathcal{G}$ will also be identified with a set of matrices. Let $k \geq 1$ be an integer. The **symmetrized trace** is the multilinear map

$$\text{symtr} : \mathcal{G} \times \overset{k}{\cdots} \times \mathcal{G} \longrightarrow$$

defined by

$$\text{symtr}(A_1,\ldots,A_k) = \frac{1}{k!}\sum_{\sigma \in S_k} \text{trace}\Big(A_{\sigma(1)}\cdots A_{\sigma(k)}\Big). \tag{6.2.2}$$

This is clearly symmetric and is also invariant under conjugation by elements of G since

$$\begin{aligned}\operatorname{trace}\Big(\Big(g^{-1}A_{\sigma(1)}g\Big)\cdots\Big(g^{-1}A_{\sigma(k)}g\Big)\Big) &= \operatorname{trace}\Big(g^{-1}\Big(A_{\sigma(1)}\cdots A_{\sigma(k)}\Big)g\Big)\\ &= \operatorname{trace}\Big(A_{\sigma(1)}\cdots A_{\sigma(k)}\Big).\end{aligned}$$

Thus, symtr is in $I^k(G)$. The corresponding *ad*-invariant polynomial of degree k on $\mathcal{G}$ is

$$\operatorname{symtr}(A,\ldots,A) = \operatorname{trace}(A^k). \tag{6.2.3}$$

In the case of the symmetrized trace we defined the invariant multilinear form and, from it, an invariant polynomial. For the remaining examples we will define the polynomials directly (and, when necessary, retrieve the multilinear forms by polarization). Let λ denote a parameter. For any fixed $A \in \mathcal{G}$ we consider the determinant

$$\det\Big(\lambda I \ + \frac{\boldsymbol{i}}{2\pi}A\Big)$$

(the reason for the $\boldsymbol{i}$ will emerge shortly; the reason for the 2 π, a bit later). When expanded this determinant is a polynomial in λ of degree n if the dimension of G is n. The coefficients of this polynomial depend on A and we write them as follows:

$$\det\Big(\lambda I + \frac{\boldsymbol{i}}{2\pi}A\Big) = f_0(A)\lambda^n + f_1(A)\lambda^{n-1} + \cdots + f_{n-1}(A)\lambda + f_n(A). \tag{6.2.4}$$

The functions $f_k(A)$ are invariant under conjugation since

$$\begin{aligned}\det\Big(\lambda I + \frac{\boldsymbol{i}}{2\pi}(g^{-1}Ag)\Big) &= \det\Big(g^{-1}(\lambda I\ )g + g^{-1}\Big(\frac{\boldsymbol{i}}{2\pi}A\Big)\,g\Big)\\ &= \det\Big(g^{-1}\Big(\lambda I \ + \frac{\boldsymbol{i}}{2\pi}A\Big)\,g\Big)\\ &= \det\Big(\lambda I + \frac{\boldsymbol{i}}{2\pi}A\Big).\end{aligned}$$

Notice that $f_0(A) = 1$ for any A. To obtain convenient expressions for the remaining $f_k(A)$ and show that they are, in fact, homogeneous polynomials we recall a few basic facts concerning the elementary symmetric functions.

Let denote a field of characteristic zero (e.g., or) and denote by $[x_1,\ldots,x_n]$ the algebra of polynomials in the n variables $x_1,\ldots,x_n$ with coefficients in . For any $P \in \ [x_1,\ldots,x_n]$ and $\sigma \in S_n$ define ${}^\sigma P$ by ${}^\sigma P(x_1,\ldots,x_n) = P(x_{\sigma(1)},\ldots,x_{\sigma(n)})$. Then P is **symmetric** if ${}^\sigma P = P$ for every $\sigma \in S_n$. The **elementary symmetric polynomials** $S_0, S_1, \ldots, S_n$ are the elements of $\ [x_1,\ldots,x_n]$ defined by

$$
\begin{aligned}
S_0(x_1, x_2, \ldots, x_n) &= 1 \\
S_1(x_1, x_2, \ldots, x_n) &= x_1 + x_2 + \cdots + x_n \\
S_2(x_1, x_2, \ldots, x_n) &= x_1x_2 + x_1x_3 + \cdots + x_1x_n + \\
&\qquad x_2x_3 + \cdots + x_2x_n + \\
&\qquad\qquad \cdots + x_{n-1}x_n \\
S_3(x_1, x_2, \ldots, x_n) &= x_1x_2x_3 + x_1x_2x_4 + \cdots + x_1x_2x_n + \\
&\quad x_1x_3x_4 + \cdots + x_1x_3x_n + \cdots + x_{n-2}x_{n-1}x_n \\
&\vdots \\
S_k(x_1, x_2, \ldots, x_n) &= x_1x_2\cdots x_k + \cdots + x_{n-k+1}x_{n-k+2}\cdots x_n \\
&\vdots \\
S_n(x_1, x_2, \ldots, x_n) &= x_1x_2\cdots x_n.
\end{aligned}
$$

Remark: One writes down $S_k(x_1, x_2, \ldots, x_n)$ as follows: From $\{x_1, x_2, \ldots, x_n\}$ select any k distinct elements and form their product (one can always write the product with the subscripts ascending). Form the sum of all such products.

Another way of describing the same functions is as follows: Let $[x_1, \ldots, x_n, z]$ denote the algebra of polynomials over in the $n+1$ variables $x_1, \ldots, x_n, z$. In $[x_1, \ldots, x_n, z]$ form the product

$$\prod_{1\le i\le n} (z + x_i) = (z + x_1)(z + x_2)\cdots(z + x_n)$$

and write it as a polynomial in z with coefficients in $[x_1, \ldots, x_n]$. The result is

$$
\begin{aligned}
\prod_{1\le i\le n} (z + x_i) &= \sum_{k=0}^{n} S_k(x_1, \ldots, x_n) z^{n-k} \\
&= S_0(x_1, \ldots, x_n)z^n + S_1(x_1, \ldots, x_n)z^{n-1} \qquad (6.2.5) \\
&\quad + \cdots + S_{n-1}(x_1, \ldots, x_n)z + S_n(x_1, \ldots, x_n).
\end{aligned}
$$

The so-called **Fundamental Theorem on Symmetric Polynomials** consists of two assertions:

I. The elementary symmetric polynomials $S_1, \ldots, S_n$ are algebraically independent in $[x_1, \ldots, x_n]$, i.e., for every nonzero $f \in [x_1, \ldots, x_n]$ the polynomial $f(S_1, \ldots, S_n)$ is nonzero ($S_1, \ldots, S_n$ do *not* satisfy any polynomial equation with coefficients in).

II. Every symmetric polynomial P in $[x_1, \ldots, x_n]$ can be written as a unique polynomial in $S_1, \ldots, S_n$, i.e., there exists a unique $f_P \in [x_1, \ldots, x_n]$ such that $P(x_1, \ldots, x_n) = f_P(S_1, \ldots, S_n)$. For example, when $n = 3$, $P(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$ is symmetric and one verifies that

$$\begin{aligned} x_1^3 + x_2^3 + x_3^3 = (x_1 + x_2 + x_3)^3 & \\ & - 3(x_1 + x_2 + x_3)(x_1x_2 + x_1x_3 + x_2x_3) \\ & + 3x_1x_2x_3 \\ = S_1^3 - 3S_1S_2 + 3S_3 & \end{aligned}$$

so that $f_P(x_1, x_2, x_3) = x_1^3 - 3x_1x_2 + 3x_3$.

Essentially any book on Algebra should contain a proof, e.g., see Section 5.2.2 of [**Fang**]. The bottom line here is that the subalgebra of $[x_1, \ldots, x_n]$ consisting of the symmetric polynomials is just $[S_1, \ldots, S_n]$ and $\{S_1, \ldots, S_n\}$ is a minimal generating set for this subalgebra.

Now let us return to (6.2.4). We wish to find simple formulas for the functions $f_k(A)$. Begin by fixing an A. We have shown that f_k is invariant under conjugation by any nonsingular matrix so its value at A is the same as its value at $g^{-1}Ag$ for any nonsingular g. Now, one can always find a nonsingular matrix g for which $g^{-1}Ag$ is triangular (see, e.g., Corollary 2, Chapter X, of [**Lang**]). Select such a g and let $a_1, \ldots, a_n$ be the diagonal entries of $g^{-1}Ag$. Then

$$\begin{aligned} \det\left(\lambda I + \frac{\boldsymbol{i}}{2\pi}A\right) &= \det\left(\lambda I + \frac{\boldsymbol{i}}{2\pi}(g^{-1}Ag)\right) = \prod_{i=1}^{n}\left(\lambda + \frac{\boldsymbol{i}}{2\pi}a_i\right) \\ &= S_0\left(\frac{\boldsymbol{i}}{2\pi}a_1, \ldots, \frac{\boldsymbol{i}}{2\pi}a_n\right)\lambda^n \\ &\quad + S_1\left(\frac{\boldsymbol{i}}{2\pi}a_1, \ldots, \frac{\boldsymbol{i}}{2\pi}a_n\right)\lambda^{n-1} + \cdots \\ &\quad + S_{n-1}\left(\frac{\boldsymbol{i}}{2\pi}a_1, \ldots, \frac{\boldsymbol{i}}{2\pi}a_n\right)\lambda \\ &\quad + S_n\left(\frac{\boldsymbol{i}}{2\pi}a_1, \ldots, \frac{\boldsymbol{i}}{2\pi}a_n\right) \\ &= \lambda^n + \left(\frac{\boldsymbol{i}}{2\pi}\right)S_1(a_1, \ldots, a_n)\lambda^{n-1} + \cdots \\ &\quad + \left(\frac{\boldsymbol{i}}{2\pi}\right)^{n-1} S_{n-1}(a_1, \ldots, a_n) \\ &\quad + \left(\frac{\boldsymbol{i}}{2\pi}\right)^{n} S_n(a_1, \ldots, a_n). \end{aligned}$$

From (6.2.4) we therefore obtain

$$f_k(A) = \left(\frac{\boldsymbol{i}}{2\pi}\right)^k S_k(a_1, \ldots, a_n), \quad k = 0, 1, \ldots, n, \tag{6.2.6}$$

where $a_1, \ldots, a_n$ are the diagonal entries in a triangular matrix similar to A. We show now that one need not actually find a triangular matrix similar to A because all of the $S_k(a_1, \ldots, a_n)$ can be expressed directly in terms of A. Of course, $S_0(a_1, \ldots, a_n) = 1$ so

$$f_0(A) = 1. \tag{6.2.7}$$

Next, observe that $S_1(a_1, \ldots, a_n) = a_1 + \cdots + a_n = \mathrm{trace}(g^{-1}Ag) = \mathrm{trace}\, A$ so

$$f_1(A) = \frac{\boldsymbol{i}}{2\pi}\ \mathrm{trace}\, A. \tag{6.2.8}$$

For $f_2(A)$ we note that

$$\begin{aligned} S_2(a_1, \ldots, a_n) &= a_1a_2 + \cdots + a_1a_n + a_2a_3 + \cdots + a_2a_n + \cdots + a_{n-1}a_n \\ &= \frac{1}{2}\Big[(a_1 + \cdots + a_n)^2 - (a_1^2 + \cdots + a_n^2)\Big]. \end{aligned}$$

Again, $a_1 + \cdots + a_n = \mathrm{trace}\, A$. Furthermore,

$$\begin{aligned} a_1^2 + \cdots + a_n^2 &= \mathrm{trace}\Big(\left(g^{-1}Ag\right)^2\Big) = \mathrm{trace}\big((g^{-1}Ag)(g^{-1}Ag)\big) \\ &= \mathrm{trace}(g^{-1}(A^2)g) = \mathrm{trace}\,(A^2) \end{aligned}$$

so we have

$$S_2(a_1, \ldots, a_n) = \frac{1}{2}\Big[(\mathrm{trace}\, A)^2 - \mathrm{trace}(A^2)\Big]$$

and therefore

$$f_2(A) = -\frac{1}{8\pi^2}\Big[(\mathrm{trace}\, A)^2 - \mathrm{trace}(A^2)\Big]. \tag{6.2.9}$$

Exercise 6.2.5 Show that

$$S_3(a_1, \ldots, a_n) = \frac{1}{6}\Big[(\mathrm{trace}\, A)^3 - 3\ \mathrm{trace}(A^2)\mathrm{trace}\, A + 2\ \mathrm{trace}(A^3)\Big]$$

and conclude that

$$\begin{aligned} f_3(A) = -\frac{\boldsymbol{i}}{48\pi^3}\Big[&(\mathrm{trace}\, A)^3 \\ &-3\ \mathrm{trace}(A^2)\ \mathrm{trace}\, A + 2\ \mathrm{trace}(A^3)\Big]. \end{aligned} \tag{6.2.10}$$

Exercise 6.2.6 Show that

$$f_n(A) = \left(\frac{\boldsymbol{i}}{2\pi}\right)^n \det A. \tag{6.2.11}$$

There are similar formulas for all of the $f_k(A)$, but these will suffice for our purposes.

At this point we are going to restrict our attention to the two specific Lie groups that are of most interest to us, i.e., $U(n)$ and $SU(n)$. Observe that, for either of these, an element A of the Lie algebra is a skew-Hermitian matrix so $\boldsymbol{i}A$ is Hermitian and therefore has n distinct real eigenvalues. Thus, A has n distinct pure imaginary eigenvalues $a_1, \ldots, a_n$. Recalling that each term in $S_k(a_1, \ldots, a_n)$ is a product of k distinct elements of $\{a_1, \ldots, a_n\}$ we have

$$f_k(A) = \left(\frac{\boldsymbol{i}}{2\pi}\right)^k S_k(a_1, \ldots, a_n) = \frac{(-1)^k}{(2\pi)^k} S_k(\boldsymbol{i}a_1, \ldots, \boldsymbol{i}a_n).$$

Since $\boldsymbol{i}a_1, \ldots, \boldsymbol{i}a_n$ are real we conclude that f_k is real-valued on either $u(n)$ or $su(n)$. Furthermore, since the elements of $su(n)$ are skew-Hermitian and *tracefree*, (6.2.8) implies that $f_1(A) = 0$ for all A in $su(2)$.

Our next objective is a complete description of the algebra $I(U(n))$ of symmetric, *ad*-invariant multilinear forms on $u(n)$. More precisely, we will show that if $\tilde{f}_1, \ldots, \tilde{f}_n$ are the elements of $I(U(n))$ corresponding (by polarization) to the polynomials $f_1, \ldots, f_n$ on $u(n)$ described above, then the subalgebra of $I(U(n))$ generated by $\{\tilde{f}_1, \ldots, \tilde{f}_n\}$ is, in fact, all of $I(U(n))$. Since $\tilde{f} \longrightarrow f$ is an isomorphism of $I(U(n))$ onto the algebra of *ad*-invariant polynomials on $u(n)$ it will suffice to show that this latter algebra is generated by $\{f_1, \ldots, f_n\}$. Begin by letting G' denote the subgroup of $U(n)$ consisting of the diagonal matrices in $U(n)$. This is clearly also a submanifold and so a Lie group. The Lie algebra $\mathcal{G}'$ of G' is the subalgebra of $u(n)$ consisting of all

$$\operatorname{diag}(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^n) = \begin{pmatrix} \boldsymbol{i}\xi^1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{i}\xi^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \boldsymbol{i}\xi^n \end{pmatrix}$$

where $\xi^1, \ldots, \xi^n$ are real. We claim first that

$$u(n) = \{g^{-1}A'g : A' \in \mathcal{G}',\ g \in U(n)\}. \tag{6.2.12}$$

Since $\mathcal{G}' \subseteq u(n)$ it is clear that any $g^{-1}A'g$ with $A' \in \mathcal{G}'$ and $g \in U(n)$ is in $u(n)$. For the reverse containment we let $A \in u(n)$. Then A is skew-Hermitain so $-\boldsymbol{i}A$ is Hermitian. According to the Spectral Theorem

(Chapter XI, **[Lang]**), there exists a $g \in U(n)$ such that $g(-\boldsymbol{i}A)g^{-1}$ is diagonal with real entries, say,

$$g(-\boldsymbol{i}A)g^{-1} = \mathrm{diag}(\xi^1, \ldots, \xi^n).$$

Thus,

$$-\boldsymbol{i}A = g^{-1}(\mathrm{diag}(\xi^1, \ldots, \xi^n))g$$

so

$$A = g^{-1}\mathrm{diag}(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^n)g$$

as required.

Now consider the restriction map $R : I(U(n)) \longrightarrow I(G')$, defined as follows: If $\tilde{f}$: $u(n) \times \cdots \times u(n) \longrightarrow$ is in $I(U(n))$, then $R\tilde{f} : \mathcal{G}' \times \cdots \times \mathcal{G}' \longrightarrow$ is given by $(R\tilde{f})(A_1', \ldots, A_k') = \tilde{f}(A_1', \ldots, A_k')$ for all $A_1', \ldots, A_k'$ in $\mathcal{G}' \subseteq u(n)$. Also let

$$N = \{g \in U(n) : g^{-1}A'g \in \mathcal{G}' \quad \text{for all } A' \in \mathcal{G}'\} \supseteq G'$$

and

$$I_N(G') = \Big\{\tilde{f} \in I(G') : \tilde{f}(g^{-1}A_1'g, \ldots, g^{-1}A_k'g) = \tilde{f}(A_1', \ldots, A_k') \quad \text{for all } A_1', \ldots, A_k' \text{ in } \mathcal{G}' \quad \text{and all } g \in N\Big\}.$$

We claim that R carries $I(U(n))$ isomorphically into $I_N(G')$. Notice that R certainly carries $I(U(n))$ into $I_N(G')$ and is an algebra homomorphism. To show that R is injective we show that its kernel is trivial. Thus, suppose $\tilde{f} \in I(U(n))$ and $R\tilde{f}$ is the zero element of $I_N(G')$, i.e., $(R\tilde{f})(A_1', \ldots, A_k') = 0$ for all $A_1', \ldots, A_k'$ in $\mathcal{G}'$. In particular, $(R\tilde{f})(A', \ldots, A') = 0$ for all A' in $\mathcal{G}'$. By (6.2.12), every $A \in u(n)$ can be written as $g^{-1}A'g$ for some $g \in U(n)$ and some $A' \in \mathcal{G}'$. Thus, $\tilde{f}(A, \ldots, A) = \tilde{f}(g^{-1}A'g, \ldots, g^{-1}A'g) = \tilde{f}(A', \ldots, A') = (R\tilde{f})(A', \ldots, A') = 0$ for all $A \in u(n)$. Consequently, the polynomial on $u(n)$ corresponding to $\tilde{f}$ is identically zero. Since this correspondence is an isomorphism, $\tilde{f}$ is identically zero and the kernel of R is trivial.

To prove that R maps onto $I_N(G')$ we note first that every element of $I_N(G')$ gives rise to a polynomial on $\mathcal{G}'$ that is invariant under conjugation by elements of N. Identifying $\mathcal{G}'$ with the set of all diag $(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^n)$ with $\xi^1, \ldots, \xi^n$ real, we claim that such a polynomial must be a symmetric function of $\xi^1, \ldots, \xi^n$. To prove this it will suffice to show that, for every pair (i, j) of indices with $1 \leq i < j \leq n$, there exists a $g \in N$ such that

$$\begin{aligned} g^{-1}(\mathrm{diag}(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^i, \ldots, \boldsymbol{i}\xi^j, \ldots, \boldsymbol{i}\xi^n))g \\ = \mathrm{diag}(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^j, \ldots, \boldsymbol{i}\xi^i, \ \ldots, \boldsymbol{i}\xi^n). \end{aligned} \tag{6.2.13}$$

Exercise 6.2.7 Show that the $n \times n$ matrix g having 1 in the (i, j)-slot, the (j, i)-slot and the (k, k)-slot for $k \neq i, j$ and having 0 elsewhere is in N and satisfies (6.2.13).

Now consider the elements $\tilde{f}_1, \ldots, \tilde{f}_n$ in $I(U(n))$ corresponding to the polynomials $f_1, \ldots, f_n$ on $u(n)$ defined by (6.2.6). The polynomials on $\mathcal{G}'$ corresponding to $R\tilde{f}_1, \ldots, R\tilde{f}_n$ are simply the restrictions to $\mathcal{G}'$ of $f_1, \ldots, f_n$ and these are given by

$$
\begin{aligned}
f_k(\mathrm{diag}(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^n)) &= \left(\frac{\boldsymbol{i}}{2\pi}\right)^k S_k(\boldsymbol{i}\xi^1, \ldots, \boldsymbol{i}\xi^n) \\
&= \frac{(-1)^k}{(2\pi)^k} S_k(\xi^1, \ldots, \xi^n)
\end{aligned}
$$

for $k = 1, \ldots, n$. Since every symmetric polynomial in $\xi^1, \ldots, \xi^n$ is a polynomial in the elementary symmetric functions $S_1(\xi^1, \ldots, \xi^k), \ldots, S_n(\xi^1, \ldots, \xi^k)$, we conclude that every element of $I(U(n))$ gives rise to a polynomial on $\mathcal{G}'$ that can be written as a polynomial in the restrictions of $f_1, \ldots, f_n$ to $\mathcal{G}'$. Thus, every element of $I_N(\mathcal{G}')$ is a polynomial in $R\tilde{f}_1, \ldots, R\tilde{f}_n$. Since R is an algebra homomorphism, a polynomial in $R\tilde{f}_1, \ldots, R\tilde{f}_n$ is the image under R of a polynomial in $\tilde{f}_1, \ldots, \tilde{f}_n$ so every element in $I_N(G')$ is the image under R of some element of $I(U(n))$. Thus,

$$
R : I(U(n)) \longrightarrow I_N(G')
$$

is an isomorphism of algebras.

Notice that we have actually proved more. Since the restrictions of $f_1, \ldots, f_n$ to $\mathcal{G}'$ are (up to constants) the elementary symmetric polynomials in $\xi^1, \ldots, \xi^n$, they are algebraically independent and therefore the same is true of the elements $R\tilde{f}_1, \ldots, R\tilde{f}_n$ in $I(G')$. Thus, $\tilde{f}_1, \ldots, \tilde{f}_n$ are algebraically independent in $I(U(n))$. Furthermore, $R\tilde{f}_1, \ldots, R\tilde{f}_n$ generate the algebra $I_N(G')$ and so $\tilde{f}_1, \ldots, \tilde{f}_n$ generate $I(U(n))$ and we have established our major result.

Theorem 6.2.1 *Let $\tilde{f}_1, \ldots, \tilde{f}_n$ be the elements of $I(U(n))$ corresponding by polarization to the ad-invariant polynomials on $u(n)$ defined by (6.2.4) (or (6.2.6)). Then $\tilde{f}_1, \ldots, \tilde{f}_n$ are algebraically independent and generate the algebra $I(U(n))$.*

The result corresponding to Theorem 6.2.1 for $SU(n)$ is almost the same except that, since $su(n)$ consists of tracefree matrices, f_1 is identically zero and so does not appear in the generating set.

Theorem 6.2.2 *Let $\tilde{f}_2, \ldots, \tilde{f}_n$ be the elements of $I(SU(n))$ corresponding by polarization to the ad-invariant polynomials on $su(n)$ defined by (6.2.4) (or (6.2.6)). Then $\tilde{f}_2, \ldots, \tilde{f}_n$ are algebraically independent and generate the algebra $I(SU(n))$.*

Exercise 6.2.8 Prove Theorem 6.2.2 by making whatever modifications are required in the arguments that led to Theorem 6.2.1.

Keep in mind that the generators of $I(U(n))$ and $I(SU(n))$ described in Theorems 6.2.1 and 6.2.2 are actually real-valued on $u(n)$ and $su(n)$, respectively. This will be of some interest when we use them to define the Chern characteristic classes in the next section.

6.3 The Chern-Weil Homomorphism

With this algebra behind us we can at last begin building characteristic classes on principal bundles. Start with a matrix Lie group G whose Lie algebra $\mathcal{G}$ is also identified with a set of matrices and a principal G-bundle

$$G \hookrightarrow P \xrightarrow{\mathcal{P}} X$$

over some manifold X. Choose some connection $\boldsymbol{\omega}$ on the bundle and denote its curvature by $\boldsymbol{\Omega}$ (one of the major tasks before us is to show that the objects we are about to construct do not depend on this choice). For every $\tilde{f} \in I^k(G)$, $k \geq 1$, we define a complex-valued $2k$-form $\tilde{f}(\boldsymbol{\Omega})$ on P by

$$\left(\tilde{f}(\boldsymbol{\Omega})\right)_p (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2k}) = \frac{1}{2^k} \sum_{\sigma \in S_{2k}} (-1)^\sigma \tilde{f}\Big(\boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}\Big), \ldots, \boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(2k-1)}, \boldsymbol{v}_{\sigma(2k)}\Big)\Big) \tag{6.3.1}$$

for all $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2k}$ in $T_p(P)$. For $k = 0$, $\tilde{f}$ is an element of and we take $\tilde{f}(\boldsymbol{\Omega})$ to be the constant 0-form whose value is $\tilde{f}$. Notice that, if $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{G}$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}^\alpha e_\alpha$, where each $\boldsymbol{\Omega}^\alpha$ is a real-valued 2-form on P, then

$$\tilde{f}(\boldsymbol{\Omega}) = \tilde{f}(e_{\alpha_1}, \ldots, e_{\alpha_k})\boldsymbol{\Omega}^{\alpha_1} \wedge \cdots \wedge \boldsymbol{\Omega}^{\alpha_k} \tag{6.3.2}$$

for $k \geq 1$. To see this we observe that

$$\begin{aligned}\left(\tilde{f}(\boldsymbol{\Omega})\right)_p (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2k}) &= \frac{1}{2^k} \sum_{\sigma \in S_{2k}} (-1)^\sigma \tilde{f}\Big(\boldsymbol{\Omega}_p^{\alpha_1}\Big(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}\Big)e_{\alpha_1}, \\ &\qquad \ldots, \boldsymbol{\Omega}_p^{\alpha_k}\Big(\boldsymbol{v}_{\sigma(2k-1)}, \boldsymbol{v}_{\sigma(2k)}\Big)e_{\alpha_k}\Big) \\ &= \frac{1}{2^k} \sum_{\sigma \in S_{2k}} (-1)^\sigma \boldsymbol{\Omega}_p^{\alpha_1}\left(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}\right) \\ &\qquad \ldots \boldsymbol{\Omega}_p^{\alpha_k}\Big(\boldsymbol{v}_{\sigma(2k-1)}, \boldsymbol{v}_{\sigma(2k)}\Big)\tilde{f}\Big(e_{\alpha_1}, \ldots, e_{\alpha_k}\Big).\end{aligned}$$

But

$$(\boldsymbol{\Omega}^{\alpha_1} \wedge \cdots \wedge \boldsymbol{\Omega}^{\alpha_k})_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2k}) = \frac{1}{2!\,2!\cdots 2!} \sum_{\sigma \in S_{2k}} (-1)^\sigma \boldsymbol{\Omega}_p^{\alpha_1}\Big(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}\Big) \ldots \boldsymbol{\Omega}_p^{\alpha_k}\Big(\boldsymbol{v}_{\sigma(2k-1)}, \boldsymbol{v}_{\sigma(2k)}\Big)$$
$$= \frac{1}{2^k} \sum_{\sigma \in S_{2k}} (-1)^\sigma \boldsymbol{\Omega}_p^{\alpha_1}\Big(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}\Big) \ldots \boldsymbol{\Omega}_p^{\alpha_k}\Big(\boldsymbol{v}_{\sigma(2k-1)}, \boldsymbol{v}_{\sigma(2k)}\Big)$$

so the result follows.

Remark: Since any $\tilde{f} \in I(G)$ is uniquely determined (via polarization) by the *ad*-invariant polynomial f on $\mathcal{G}$ corresponding to it, the form $\tilde{f}(\boldsymbol{\Omega})$ is also uniquely determined by f (and $\boldsymbol{\Omega}$, of course). Thus, for any $f \in P^k_{ad}(\mathcal{G})$, we will often write $f(\boldsymbol{\Omega})$ for the $2k$-form $\tilde{f}(\boldsymbol{\Omega})$.

Before proving the major results of this section we pause to write out a number of concrete examples. Specifically, we will compute, for $k = 1, 2$, the $2k$-form determined by $\boldsymbol{\Omega}$ and the symmetrized trace (6.2.2). Begin with $k = 1$ so that $\tilde{f}(A) = \text{symtr}(A) = \text{trace}(A)$. Then

$$\Big(\tilde{f}(\boldsymbol{\Omega})\Big)_p(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{1}{2^1} \sum_{\sigma \in S_2} (-1)^\sigma \tilde{f}\Big(\boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}\Big)\Big)$$
$$= \frac{1}{2}\Big[\tilde{f}\left(\boldsymbol{\Omega}_p(\boldsymbol{v}_1, \boldsymbol{v}_2)\right) - \tilde{f}\left(\boldsymbol{\Omega}_p(\boldsymbol{v}_2, \boldsymbol{v}_1)\right)\Big]$$
$$= \frac{1}{2}\Big[\text{trace}(\boldsymbol{\Omega}_p(\boldsymbol{v}_1, \boldsymbol{v}_2)) - \text{trace}(\boldsymbol{\Omega}_p(\boldsymbol{v}_2, \boldsymbol{v}_1))\Big]$$
$$= \frac{1}{2}\text{trace}(\boldsymbol{\Omega}_p(\boldsymbol{v}_1, \boldsymbol{v}_2) - \boldsymbol{\Omega}_p(\boldsymbol{v}_2, \boldsymbol{v}_1))$$
$$= \frac{1}{2}\text{trace}(2\boldsymbol{\Omega}_p(\boldsymbol{v}_1, \boldsymbol{v}_2))$$
$$= \text{trace}(\boldsymbol{\Omega}_p(\boldsymbol{v}_1, \boldsymbol{v}_2))$$
$$= (\text{trace} \circ \boldsymbol{\Omega})_p(\boldsymbol{v}_1, \boldsymbol{v}_2)$$

so $\tilde{f}(\boldsymbol{\Omega}) = \text{trace} \circ \boldsymbol{\Omega}$, which we write simply as

$$\tilde{f}(\boldsymbol{\Omega}) = \text{trace } \boldsymbol{\Omega} \quad \Big(\tilde{f}(A) = \text{symtr}(A) = \text{trace}(A)\Big). \tag{6.3.3}$$

When $k = 2$ the symmetrized trace is given by $\tilde{f}(A_1, A_2) = \text{symtr}(A_1, A_2) = \frac{1}{2!}\sum_{\sigma \in S_2} \text{trace}(A_{\sigma(1)}A_{\sigma(2)}) = \frac{1}{2}\,[\text{trace}(A_1A_2) + \text{trace}(A_2A_1)] = \text{trace}(A_1A_2)$.

Thus,

$$\begin{aligned}\left(\tilde{f}(\boldsymbol{\Omega})\right)_p(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3,\boldsymbol{v}_4) &= \frac{1}{2^2}\sum_{\sigma\in S_4}(-1)^\sigma \tilde{f}\Big(\boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(1)},\boldsymbol{v}_{\sigma(2)}\Big),\\ &\qquad \boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(3)},\boldsymbol{v}_{\sigma(4)}\Big)\Big)\\ &= \frac{1}{4}\sum_{\sigma\in S_4}(-1)^\sigma \mathrm{trace}\Big(\boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(1)},\boldsymbol{v}_{\sigma(2)}\Big)\cdot\\ &\qquad \boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(3)},\boldsymbol{v}_{\sigma(4)}\Big)\Big)\\ &= \mathrm{trace}\Big(\frac{1}{2!\,2!}\sum_{\sigma\in S_4}(-1)^\sigma \boldsymbol{\Omega}_p(\boldsymbol{v}_{\sigma(1)},\boldsymbol{v}_{\sigma(2)}\Big)\cdot\\ &\qquad \boldsymbol{\Omega}_p\Big(\boldsymbol{v}_{\sigma(3)},\boldsymbol{v}_{\sigma(4)}\Big)\Big)\\ &= \mathrm{trace}\Big((\boldsymbol{\Omega}\wedge\boldsymbol{\Omega})_p(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3,\boldsymbol{v}_4)\Big)\\ &= (\mathrm{trace}\circ(\boldsymbol{\Omega}\wedge\boldsymbol{\Omega}))_p(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3,\boldsymbol{v}_4)\end{aligned}$$

so $\tilde{f}(\boldsymbol{\Omega}) = \mathrm{trace}\circ(\boldsymbol{\Omega}\wedge\boldsymbol{\Omega})$. Again, we will write this as

$$\begin{gathered}\tilde{f}(\boldsymbol{\Omega}) = \mathrm{trace}(\boldsymbol{\Omega}\wedge\boldsymbol{\Omega})\\ \left(\tilde{f}(A_1,A_2) = \mathrm{symtr}(A_1,A_2) = \mathrm{trace}(A_1A_2)\right).\end{gathered} \tag{6.3.4}$$

Exercise 6.3.1 Show that the forms trace $\boldsymbol{\Omega}$ and $\mathrm{trace}(\boldsymbol{\Omega}\wedge\boldsymbol{\Omega})$ on P both project to closed forms on X. **Hint**: Consider trace $\boldsymbol{\mathcal{F}}$ and $\mathrm{trace}(\boldsymbol{\mathcal{F}}\wedge\boldsymbol{\mathcal{F}})$, where $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ is any field strength corresponding to a local cross-section s.

Our first major result is a generalization of Exercise 6.3.1.

Theorem 6.3.1 *Let $\boldsymbol{\omega}$ be a connection with curvature $\boldsymbol{\Omega}$ on the principal G-bundle $G\hookrightarrow P \xrightarrow{\mathcal{P}} X$ and let $\tilde{f}\in I^k(G)$ for some $k\geq 1$. Then the $2k$-form $\tilde{f}(\boldsymbol{\Omega})$ defined by (6.3.1) projects to a unique closed $2k$-form $\bar{f}(\boldsymbol{\Omega})$ on X (i.e., $\tilde{f}(\boldsymbol{\Omega}) = \mathcal{P}^*(\bar{f}(\boldsymbol{\Omega}))$ and $d(\bar{f}(\boldsymbol{\Omega})) = 0$).*

Proof: To show that $\tilde{f}(\boldsymbol{\Omega})$ projects to a unique $2k$-form on X we use Lemma 4.5.1. Thus, we must show

1. $\tilde{f}(\boldsymbol{\Omega})$ is zero whenever at least one of its arguments is vertical.
2. $\sigma_g^*(\tilde{f}(\boldsymbol{\Omega})) = \tilde{f}(\boldsymbol{\Omega}))$ for each $g\in G$.

The first of these is clear from (6.3.2) since $\boldsymbol{\Omega}$ vanishes when either of its arguments is vertical. To prove the second we recall that $\sigma_g^*(\boldsymbol{\Omega}) = ad_{g^{-1}}\circ\ \boldsymbol{\Omega}$

(Lemma 6.2.2, **[N4]**) and that $\tilde{f}$ satisfies $\tilde{f}(ad_{g^{-1}}(A_1), \ldots, ad_{g^{-1}}(A_k)) = \tilde{f}(A_1, \ldots, A_k)$ so

$$\begin{aligned}
&\left(\sigma_g^*\left(\tilde{f}(\boldsymbol{\Omega})\right)\right)_p (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2k}) \\
&\quad = \left(\tilde{f}(\boldsymbol{\Omega})\right)_{p \cdot g} ((\sigma_g)_{*p}(\boldsymbol{v}_1), \ldots, (\sigma_g)_{*p}(\boldsymbol{v}_{2k})) \\
&\quad = \frac{1}{2^k} \sum_{\tau \in S_{2k}} (-1)^\tau \tilde{f}\Big(\boldsymbol{\Omega}_{p \cdot g}\Big((\sigma_g)_{*p}(\boldsymbol{v}_{\tau(1)}), (\sigma_g)_{*p}(\boldsymbol{v}_{\tau(2)})\Big), \ldots, \\
&\qquad\qquad\qquad \boldsymbol{\Omega}_{p \cdot g}\Big((\sigma_g)_{*p}(\boldsymbol{v}_{\tau(2k-1)}), (\sigma_g)_{*p}(\boldsymbol{v}_{\tau(2k)})\Big)\Big) \\
&\quad = \frac{1}{2^k} \sum_{\tau \in S_{2k}} (-1)^\tau \tilde{f}\Big((\sigma_g^* \boldsymbol{\Omega})_p(\boldsymbol{v}_{\tau(1)}, \boldsymbol{v}_{\tau(2)}), \ldots, \\
&\qquad\qquad\qquad (\sigma_g^* \boldsymbol{\Omega})_p(\boldsymbol{v}_{\tau(2k-1)}, \boldsymbol{v}_{\tau(2k)})\Big) \\
&\quad = \frac{1}{2^k} \sum_{\tau \in S_{2k}} (-1)^\tau \tilde{f}\Big(ad_{g^{-1}}\Big(\boldsymbol{\Omega}_p(\boldsymbol{v}_{\tau(1)}, \boldsymbol{v}_{\tau(2)})\Big), \ldots, \\
&\qquad\qquad\qquad ad_{g^{-1}}\Big(\boldsymbol{\Omega}_p(\boldsymbol{v}_{\tau(2k-1)}, \boldsymbol{v}_{\tau(2k)})\Big)\Big) \\
&\quad = \frac{1}{2^k} \sum_{\tau \in S_{2k}} (-1)^\tau \tilde{f}\Big(\boldsymbol{\Omega}_p(\boldsymbol{v}_{\tau(1)}, \boldsymbol{v}_{\tau(2)}), \ldots, \boldsymbol{\Omega}_p(\boldsymbol{v}_{\tau(2k-1)}, \boldsymbol{v}_{\tau(2k)})\Big) \\
&\quad = \left(\tilde{f}(\boldsymbol{\Omega})\right)_p (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2k})
\end{aligned}$$

as required.

Thus, there exists a unique $2k$-form $\bar{f}(\boldsymbol{\Omega})$ on X with $\mathcal{P}^*(\bar{f}(\boldsymbol{\Omega})) = \tilde{f}(\boldsymbol{\Omega})$. We show that $\bar{f}(\boldsymbol{\Omega})$ is closed. Notice that $d(\tilde{f}(\boldsymbol{\Omega})) = d(\mathcal{P}^*(\bar{f}(\boldsymbol{\Omega}))) = \mathcal{P}^*(d(\bar{f}(\boldsymbol{\Omega})))$ so $d(\tilde{f}(\boldsymbol{\Omega}))$ projects to $d(\bar{f}(\boldsymbol{\Omega}))$. Since projections are unique when they exist, it will suffice to show that $\tilde{f}(\boldsymbol{\Omega})$ is closed. Now, because $\tilde{f}(\boldsymbol{\Omega})$ projects, Lemma 4.5.5 implies that $d(\tilde{f}(\boldsymbol{\Omega})) = d^{\boldsymbol{\omega}}(\tilde{f}(\boldsymbol{\Omega}))$ so $d(\tilde{f}(\boldsymbol{\Omega}))$ vanishes when any one of its arguments is vertical. Thus, we need only show that $d(\tilde{f}(\boldsymbol{\Omega}))$ vanishes when all of its arguments are horizontal. For this we let $\{e_\alpha\}$ be a basis for $\mathcal{G}$ and write $\boldsymbol{\Omega} = \boldsymbol{\Omega}^\alpha e_\alpha$, where each $\boldsymbol{\Omega}^\alpha$ is a real-valued 2-form on P. Then

$$\tilde{f}(\boldsymbol{\Omega}) = \tilde{f}(e_{\alpha_1}, \ldots, e_{\alpha_k}) \boldsymbol{\Omega}^{\alpha_1} \wedge \cdots \wedge \boldsymbol{\Omega}^{\alpha_k}$$

so

$$\begin{aligned}
d(\tilde{f}(\boldsymbol{\Omega})) = \tilde{f}(e_{\alpha_1}, \ldots, e_{\alpha_k}) (&d\boldsymbol{\Omega}^{\alpha_1} \wedge \boldsymbol{\Omega}^{\alpha_2} \wedge \cdots \wedge \boldsymbol{\Omega}^{\alpha_k} \\
&+ \boldsymbol{\Omega}^{\alpha_1} \wedge d\boldsymbol{\Omega}^{\alpha_2} \wedge \cdots \wedge \boldsymbol{\Omega}^{\alpha_k} \\
&+ \cdots + \boldsymbol{\Omega}^{\alpha_1} \wedge \boldsymbol{\Omega}^{\alpha_2} \wedge \cdots \wedge d\boldsymbol{\Omega}^{\alpha_k}).
\end{aligned} \tag{6.3.5}$$

Now, the Bianchi Identity (Theorem 4.5.9) gives

$$d^{\boldsymbol{\omega}}\boldsymbol{\Omega} = 0$$
$$d^{\boldsymbol{\omega}}(\boldsymbol{\Omega}^{\alpha} e_{\alpha}) = 0$$
$$(d^{\boldsymbol{\omega}}\boldsymbol{\Omega}^{\alpha}) e_{\alpha} = 0$$

so $d^{\boldsymbol{\omega}}\boldsymbol{\Omega}^{\alpha} = 0$ for $\alpha = 1, \ldots, n$. Consequently, each $d\boldsymbol{\Omega}^{\alpha}$ vanishes when all of its arguments are horizontal and, by (6.3.5), so does $d(\tilde{f}(\boldsymbol{\Omega}))$. ∎

Exercise 6.3.2 Let $\tilde{f} \in I^k(G)$ and $\tilde{g} \in I^l(G)$. Show that

$$\left(\tilde{f} \odot \tilde{g}\right)(\boldsymbol{\Omega}) = \mathcal{P}^*\left(\bar{f}(\boldsymbol{\Omega}) \wedge \bar{g}(\boldsymbol{\Omega})\right).$$

Hint: Keep in mind that the wedge product is commutative on real-valued 2-forms so that $\boldsymbol{\Omega}^{\alpha_1} \wedge \cdots \wedge \boldsymbol{\Omega}^{\alpha_{k+l}}$ is symmetric in $\alpha_1, \ldots, \alpha_{k+l}$.

Before moving on to the major result of this section we will pause to construct a few of the projections $\bar{f}(\boldsymbol{\Omega})$. Recall (from the proof of Lemma 4.5.1) that if $\boldsymbol{\varphi}$ is a k-form on P which vanishes when any of its arguments is vertical and satisfies $\sigma_g^*\boldsymbol{\varphi} = \boldsymbol{\varphi}$ for every $g \in G$, then its unique projection $\bar{\boldsymbol{\varphi}}$ to X can be described explicitly as follows: For each $x \in X$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ in $T_x(X)$

$$\bar{\boldsymbol{\varphi}}_x(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k) = \boldsymbol{\varphi}_{s(x)}(s_{*x}(\boldsymbol{w}_1), \ldots, s_{*x}(\boldsymbol{w}_k)),$$

where s is *any* local cross-section with x in its domain. Thus, $\bar{\boldsymbol{\varphi}} = s^*\boldsymbol{\varphi}$.

Now, suppose $\tilde{f}(A) = \text{symtr}(A) = \text{trace}(A)$. Then, by (6.3.3), $\tilde{f}(\boldsymbol{\Omega}) = \text{trace}\,\boldsymbol{\Omega}$. The projection $\bar{f}(\boldsymbol{\Omega})$ of $\tilde{f}(\boldsymbol{\Omega})$ to X is obtained as follows: Choose a local cross-section $s : V \longrightarrow \mathcal{P}^{-1}(V)$. Then, for any $x \in V$ and any $\boldsymbol{w}_1, \boldsymbol{w}_2 \in T_x(X)$,

$$\begin{aligned}
(\bar{f}(\boldsymbol{\Omega}))_x(\boldsymbol{w}_1, \boldsymbol{w}_2) &= \left(\tilde{f}(\boldsymbol{\Omega})\right)_{s(x)}(s_{*x}(\boldsymbol{w}_1), s_{*x}(\boldsymbol{w}_2)) \\
&= (\text{trace}\,\boldsymbol{\Omega})_{s(x)}(s_{*x}(\boldsymbol{w}_1), s_{*x}(\boldsymbol{w}_2)) \\
&= \text{trace}\Big(\boldsymbol{\Omega}_{s(x)}(s_{*x}(\boldsymbol{w}_1), s_{*x}(\boldsymbol{w}_2))\Big) \\
&= \text{trace}((s^*\boldsymbol{\Omega})_x(\boldsymbol{w}_1, \boldsymbol{w}_2)) \\
&= \text{trace}(\boldsymbol{\mathcal{F}}_x\,(\boldsymbol{w}_1, \boldsymbol{w}_2)) \\
&= (\text{trace}\,\boldsymbol{\mathcal{F}})_x\,(\boldsymbol{w}_1, \boldsymbol{w}_2),
\end{aligned}$$

where $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ is the local field strength corresponding to $\boldsymbol{\Omega}$ in gauge s. Thus,

$$f(A) = \text{trace}\,A \Longrightarrow \bar{f}(\boldsymbol{\Omega}) = \text{trace}\,\boldsymbol{\mathcal{F}}. \tag{6.3.6}$$

Remark: Again we emphasize that different cross-sections give different local field strengths, in general, but their traces agree on the intersections of their domains so trace $\boldsymbol{\mathcal{F}}$ only appears to be a locally defined object; it is, in fact, a global 2-form on X.

Exercise 6.3.3 Show that

$$f(A) = \text{trace}(A^2) \Longrightarrow \bar{f}(\boldsymbol{\Omega}) = \text{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}). \tag{6.3.7}$$

Hint: See (6.3.4).

As a final example we consider $\tilde{f} \odot \tilde{f} \in I^2(G)$, where $\tilde{f} \in I^1(G)$ is given by $\tilde{f}(A) = \text{trace}\, A$. We compute

$$\begin{aligned}\left(\tilde{f} \odot \tilde{f}\right)(A_1, A_2) &= \frac{1}{(1+1)!} \sum_{\sigma \in S_2} \tilde{f}\left(A_{\sigma(1)}\right) \tilde{f}\left(A_{\sigma(2)}\right) \\ &= \frac{1}{2}\left[\tilde{f}(A_1)\tilde{f}(A_2) + \tilde{f}(A_2)\tilde{f}(A_1)\right] \\ &= \tilde{f}(A_1)\tilde{f}(A_2) \\ &= (\text{trace}\, A_1)(\text{trace}\, A_2).\end{aligned}$$

Note that the corresponding polynomial is $(\text{trace}\, A)^2$. Furthermore, Exercise 6.3.2 implies that the projection of $(\tilde{f} \odot \tilde{f})(\boldsymbol{\Omega})$ onto X is $\bar{f}(\boldsymbol{\Omega}) \wedge \bar{f}(\boldsymbol{\Omega})$ which, by (6.3.6), is $(\text{trace}\ \boldsymbol{\mathcal{F}}) \wedge (\text{trace}\ \boldsymbol{\mathcal{F}})$. Thus,

$$f(A) = (\text{trace}\, A)^2 \Longrightarrow \bar{f}(\boldsymbol{\Omega}) = (\text{trace}\, \boldsymbol{\mathcal{F}}) \wedge (\text{trace}\ \boldsymbol{\mathcal{F}}). \tag{6.3.8}$$

Now, suppose $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is a fixed principal G-bundle and $\tilde{f}$ is a fixed element of $I^k(G)$, but that $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}_1$ are two connections on P with curvatures $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$, respectively. One computes $\tilde{f}(\boldsymbol{\Omega}_0)$ and $\tilde{f}(\boldsymbol{\Omega}_1)$ as well as their projections $\bar{f}(\boldsymbol{\Omega}_0)$ and $\bar{f}(\boldsymbol{\Omega}_1)$ to X. There is, of course, no reason to expect these projections to be the same and, indeed, they are generally not. Our next result, however, asserts that, at least in the cohomological sense, they differ only trivially, i.e., by an exact form.

Theorem 6.3.2 *Let $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}_1$ be connections on the principal G-bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ with curvatures $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$, respectively, and let $\tilde{f} \in I^k(G)$ for some $k \geq 1$. Then there exists a $(2k-1)$-form $\boldsymbol{\varphi}$ on X such that $\bar{f}(\boldsymbol{\Omega}_1) - \bar{f}(\boldsymbol{\Omega}_0) = d\boldsymbol{\varphi}$.*

Proof: Let $\boldsymbol{\alpha} = \boldsymbol{\omega}_1 - \boldsymbol{\omega}_0$ and for each $t \in [0,1], \boldsymbol{\omega}_t = \boldsymbol{\omega}_0 + t\boldsymbol{\alpha}$. Then $\boldsymbol{\alpha}$ is tensorial of type *ad*. Thus, each $t\,\boldsymbol{\alpha}$ is tensorial of type *ad* so each $\boldsymbol{\omega}_t$ is a connection on P. Let

$$\boldsymbol{\Omega}_t = d^{\boldsymbol{\omega}_t}\boldsymbol{\omega}_t = d\boldsymbol{\omega}_t + \frac{1}{2}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_t]$$

be the curvature of $\boldsymbol{\omega}_t$ and notice that

$$\begin{aligned}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_t] &= [\boldsymbol{\omega}_0 + t\boldsymbol{\alpha}, \boldsymbol{\omega}_0 + t\boldsymbol{\alpha}] \\ &= [\boldsymbol{\omega}_0, \boldsymbol{\omega}_0] + t[\boldsymbol{\alpha}, \boldsymbol{\omega}_0] + t[\boldsymbol{\omega}_0, \boldsymbol{\alpha}] + t^2[\boldsymbol{\alpha}, \boldsymbol{\alpha}]\end{aligned}$$

so that

$$\begin{aligned}\frac{d}{dt}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_t] &= [\boldsymbol{\alpha}, \boldsymbol{\omega}_0] + [\boldsymbol{\omega}_0, \boldsymbol{\alpha}] + 2t[\boldsymbol{\alpha}, \boldsymbol{\alpha}] \\ &= ([\boldsymbol{\omega}_0, \boldsymbol{\alpha}] + t[\boldsymbol{\alpha}, \boldsymbol{\alpha}]) + ([\boldsymbol{\alpha}, \boldsymbol{\omega}_0] + t[\boldsymbol{\alpha}, \boldsymbol{\alpha}]) \\ &= [\boldsymbol{\omega}_0 + t\boldsymbol{\alpha}, \boldsymbol{\alpha}\ ] + [\boldsymbol{\alpha}, \boldsymbol{\omega}_0 + t\boldsymbol{\alpha}] \\ &= [\boldsymbol{\omega}_t, \boldsymbol{\alpha}] + [\boldsymbol{\alpha}, \boldsymbol{\omega}_t] \\ &= 2[\boldsymbol{\alpha}, \boldsymbol{\omega}_t].\end{aligned}$$

Thus,

$$\begin{aligned}\frac{d}{dt}\boldsymbol{\Omega}_t &= \frac{d}{dt}\Big(d\boldsymbol{\omega}_t + \frac{1}{2}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_t]\Big) \\ &= \frac{d}{dt}\Big(d\boldsymbol{\omega}_0 + t\,d\boldsymbol{\alpha} + \frac{1}{2}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_t]\Big) \\ \frac{d}{dt}\boldsymbol{\Omega}_t &= d\boldsymbol{\alpha} + [\boldsymbol{\alpha}, \boldsymbol{\omega}_t]. \end{aligned} \tag{6.3.9}$$

Exercise 6.3.4 Let $\tilde{f}$ be an element of $I^k(G)$ and let $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k$ be $\mathcal{G}$-valued forms on P of degree $q_1, \ldots, q_k$, respectively. Define a $(q_1 + \cdots + q_k\,)$-form $\tilde{f}(\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k)$ on P by

$$\begin{aligned}&\Big(\tilde{f}(\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k)\Big)_p\Big(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{q_1+\cdots+q_k}\Big) \\ &= \frac{1}{q_1!\cdots q_k!}\sum_{\sigma\in S_{q_1+\cdots+q_k}}(-1)^\sigma \tilde{f}\Big(\boldsymbol{\varphi}_1\Big(\boldsymbol{v}_{\sigma(1)}, \ldots, \boldsymbol{v}_{\sigma(q_1)}\Big), \\ &\qquad \ldots, \boldsymbol{\varphi}_k\Big(\boldsymbol{v}_{\sigma(q_1+\cdots+q_{k-1}+1)}, \ldots, \boldsymbol{v}_{\sigma(q_1+\cdots+q_k)}\Big)\Big).\end{aligned} \tag{6.3.10}$$

Show that, if $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{G}$ and $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_1^{\alpha_1}\, e_{\alpha_1}, \ldots, \boldsymbol{\varphi}_k = \boldsymbol{\varphi}_k^{\alpha_k}\, e_{\alpha_k}$, then

$$\tilde{f}(\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k) = \tilde{f}\Big(e_{\alpha_1}, \ldots, e_{\alpha_k}\Big)\boldsymbol{\varphi}_1^{\alpha_1} \wedge \cdots \wedge \boldsymbol{\varphi}_k^{\alpha_k}. \tag{6.3.11}$$

Now consider the $(2k-1)$-form $\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)$ on P defined as in Exercise 6.3.4. If $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{G}$ with $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{\alpha_0} e_{\alpha_0}$ and $\boldsymbol{\Omega}_t = \boldsymbol{\Omega}_t^{\alpha} e_\alpha$, (6.3.11) gives

$$\begin{aligned}\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t) = \tilde{f}\Big(e_{\alpha_0}, e_{\alpha_1}, \ldots, e_{\alpha_{k-1}}\Big)\boldsymbol{\alpha}^{\alpha_0} \\ \wedge\, \boldsymbol{\Omega}_t^{\alpha_1} \wedge \cdots \wedge \boldsymbol{\Omega}_t^{\alpha_{k-1}}.\end{aligned} \tag{6.3.12}$$

Exercise 6.3.5 Show that $\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)$ projects to a unique $(2k-1)$-form on X. **Hint**: The argument is the same as for $\tilde{f}(\boldsymbol{\Omega})$ in the proof of Theorem 6.3.1.

Next define a $(2k-1)$-form $\boldsymbol{\Phi}$ on P by

$$\boldsymbol{\Phi} = k \int_0^1 \tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t) dt. \tag{6.3.13}$$

Since $\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)$ projects to X, so does $\boldsymbol{\Phi}$ (to the analogously defined integral of the projection of $\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)$). We claim that

$$d\boldsymbol{\Phi} = \tilde{f}(\boldsymbol{\Omega}_1) - \tilde{f}(\boldsymbol{\Omega}_0). \tag{6.3.14}$$

First note that this will complete the proof of the Theorem since, if $\boldsymbol{\Phi} = \mathcal{P}^* \boldsymbol{\varphi}$ it gives

$$\begin{aligned} d(\mathcal{P}^* \boldsymbol{\varphi}) &= \mathcal{P}^* \big(\bar{f}(\boldsymbol{\Omega}_1)\big) - \mathcal{P}^* \big(\bar{f}(\boldsymbol{\Omega}_0)\big) \\ \mathcal{P}^*(d\, \boldsymbol{\varphi}) &= \mathcal{P}^* \big(\bar{f}(\boldsymbol{\Omega}_1) - \bar{f}(\boldsymbol{\Omega}_0)\big) \end{aligned}$$

and therefore, since projections are unique when they exist,

$$d\boldsymbol{\varphi} = \bar{f}(\boldsymbol{\Omega}_1) - \bar{f}(\boldsymbol{\Omega}_0). \tag{6.3.15}$$

Thus, we need only prove (6.3.14). First note that $d(\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)) = d^{\boldsymbol{\omega}_t}(\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t))$ by Lemma 4.5.5. But, from (6.3.12), the product rule for $d^{\boldsymbol{\omega}_t}$ and the Bianchi Identity,

$$\begin{aligned} d^{\boldsymbol{\omega}_t} \left(\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t) \right) &= \tilde{f}(d^{\boldsymbol{\omega}_t} \boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t) \\ &= \tilde{f}(d\boldsymbol{\alpha} + [\boldsymbol{\alpha}, \boldsymbol{\omega}_t], \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t). \end{aligned}$$

Now we use (6.3.9) to compute

$$\begin{aligned} kd \left(\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t) \right) &= k\tilde{f}(d\boldsymbol{\alpha} + [\boldsymbol{\alpha}, \boldsymbol{\omega}_t], \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t) \\ &= k\tilde{f} \left(\frac{d}{dt} \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t \right) \\ &= \tilde{f} \left(\frac{d}{dt} \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t \right) + \tilde{f} \left(\boldsymbol{\Omega}_t, \frac{d}{dt} \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t \right) \\ &\qquad + \cdots + \tilde{f} \left(\boldsymbol{\Omega}_t, \boldsymbol{\Omega}_t, \ldots, \frac{d}{dt} \boldsymbol{\Omega}_t \right) \\ &= \frac{d}{dt} \tilde{f}(\boldsymbol{\Omega}_t, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t). \end{aligned}$$

Thus,

$$d\boldsymbol{\Phi} = k\int_0^1 d\left(\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)\right) dt = \int_0^1 \frac{d}{dt}\tilde{f}(\boldsymbol{\Omega}_t, \boldsymbol{\Omega}_t, \ldots, \boldsymbol{\Omega}_t)dt$$
$$= \tilde{f}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_1) - \tilde{f}(\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_0, \ldots, \boldsymbol{\Omega}_0) = \tilde{f}(\boldsymbol{\Omega}_1) - \tilde{f}(\boldsymbol{\Omega}_0)$$

as required. ■

The situation is now as follows: We are given a principal bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$. Connections always exist on this bundle (Theorem 3.1.7) so we select one. Let $\boldsymbol{\omega}$ denote this connection and $\boldsymbol{\Omega}$ its curvature. Now, let $f \in P^k_{ad}(\mathcal{G})$ be any symmetric homogeneous polynomial on the Lie algebra $\mathcal{G}$ that is invariant under conjugation by elements of G. We have associated with f a closed $2k$-form $\bar{f}(\boldsymbol{\Omega})$ on X. A different connection $\boldsymbol{\omega}'$ will give rise to a different $2k$-form $\bar{f}(\boldsymbol{\Omega}')$, but we now know that $\bar{f}(\boldsymbol{\Omega})$ and $\bar{f}(\boldsymbol{\Omega}')$ are cohomologically the same, i.e., cohomologous. Thus, $\bar{f}(\boldsymbol{\Omega})$ and $\bar{f}(\boldsymbol{\Omega}')$ determine the same (complex) de Rham cohomology class, which we denote $w(f) \in H^{2k}_{\mathrm{de\,R}}(X;\ \ )$. We have then a mapping

$$w\ : I(G) \longrightarrow H^*_{\mathrm{de\,R}}(X;\ \ )$$

which assigns to each $\tilde{f} \in I(G)$ the cohomology class $[\bar{f}(\boldsymbol{\Omega})]$, where $\boldsymbol{\Omega}$ is the curvature of *any* connection on the bundle. This map is, in fact, an algebra homomorphism since it is clearly linear and, by Exercise 6.3.2, it also satisfies

$$w\left(\tilde{f} \odot \tilde{g}\right) = w(\tilde{f}) \wedge w(\tilde{g}).$$

The image of $I(G)$ in $H^*_{\mathrm{de\,R}}(X;\ \ )$ under this homomorphism is an algebra of complex cohomology classes on X the elements of which are called **(complex) characteristic classes** for the bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$. The map w itself is called the **Chern-Weil homomorphism** of the bundle.

If $\tilde{f} \in I(G)$ happens to be real-valued, then any $\bar{f}(\boldsymbol{\Omega})$ is a real-valued closed form on X and so we can identify $w(\tilde{f}) = [\bar{f}(\boldsymbol{\Omega})]$ with an element of $H^*_{\mathrm{de\,R}}(X;\ \ )$. This is, in fact, the case for many of our most important examples. If f_k is one of the *ad*-invariant polynomials defined by (6.2.4) (or (6.2.6)), then the corresponding characteristic class

$$c_k(P) = w(f_k) = [\bar{f}_k(\boldsymbol{\Omega})] \tag{6.3.16}$$

is called the $\boldsymbol{k}^{\mathbf{th}}$ **Chern class** of the bundle $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$. For example, using (6.2.8) and (6.3.6), the 1^{st} Chern class is given by

$$c_1(P) = \frac{\boldsymbol{i}}{2\pi}[\text{trace } \boldsymbol{\mathcal{F}}]. \tag{6.3.17}$$

Similarly, from (6.2.9), (6.3.7) and (6.3.8), the 2nd Chern class is

$$c_2(P) = -\frac{1}{8\pi^2}[(\text{trace } \boldsymbol{\mathcal{F}}) \wedge (\text{trace } \boldsymbol{\mathcal{F}}) - \text{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}})]. \tag{6.3.18}$$

There are similar formulas for the remaining $c_k(P)$. However, since $\bar{f}_k(\boldsymbol{\Omega})$ is a $2k$-form on X these will reduce to zero when $k > \frac{1}{2}\dim X$. In particular, if $\dim X = 2$, then $c_1(P)$ is the only nonzero Chern class. If $\dim X = 4$, then $c_1(P)$ and $c_2(P)$ are the only nontrivial Chern classes.

Let us consider in somewhat more detail the cases $G = U(n)$ and $G = SU(n)$ that most interest us. We have already seen that, for these groups, the Chern classes are real cohomology classes.

Remark: Although we will not do so, one can prove that this is actually true for any matrix Lie group G; Chern classes are real cohomology classes. Indeed, they are (in a sense we will describe shortly) *integral* cohomology classes.

Furthermore, Theorems 6.2.1 and 6.2.2 imply that the algebra of characteristic classes for any $U(n)$- or $SU(n)$-bundle is generated by the Chern classes.

Remark: This is not true for arbitary Lie groups. However, it is known that, for any compact, semi-simple Lie group G, the algebra $I(G)$ is finitely generated so that the Chern-Weil homomorphism for any G-bundle is determined by its values on finitely many elements of $I(G)$. The images of these then generate the algebra of characteristic classes. For many of the most important Lie groups it is possible to write out simple, finite generating sets (see Chapter XII of [**KN2**]).

In the case of $SU(n)$, trace $\boldsymbol{\mathcal{F}}$ is zero so many of the formulas simplify considerably, e.g.,

$$c_1(P) = 0 \quad (G = SU(n)), \tag{6.3.19}$$

$$c_2(P) = \frac{1}{8\pi^2}[\text{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}})] \quad (G = SU(n)). \tag{6.3.20}$$

We saw in Section 6.1 that $U(1)$-bundles over S^2 are classified up to equivalence by their 1st Chern classes (note that, for $U(1)$-bundles, $\boldsymbol{\mathcal{F}}$ is a 1×1 matrix so trace $\boldsymbol{\mathcal{F}}$ is just its sole entry and the definition of $c_1(P)$ given in Section 6.1 agrees with (6.3.17)). We would like to show that $SU(2)$-bundles over S^4 are likewise characterized by their 2nd Chern classes. First we show that equivalent bundles have, not only the same Chern classes, but the same Chern-Weil homomorphisms.

Theorem 6.3.3 *Let* $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ *and* $G \hookrightarrow P' \xrightarrow{\mathcal{P}'} X$ *be equivalent principal* G*-bundles over* X*. Then the Chern-Weil homomorphisms* $I(G) \longrightarrow H^*(X; \quad)$ *they determine are equal.*

Proof: Let $\lambda : P \longrightarrow P'$ be an equivalence (i.e., a diffeomorphism of P onto P' satisfying $\lambda(p \cdot g) = \lambda(p) \cdot g$ and $P'(\lambda(p)) = \mathcal{P}(p)$ for all $p \in P$ and $g \in G$). Fix $\tilde{f} \in I^k(G), k \geq 1$. Now, select a connection $\boldsymbol{\omega}'$ on P' with curvature $\boldsymbol{\Omega}'$. Then $\boldsymbol{\omega} = \lambda^* \boldsymbol{\omega}'$ is a connection on P with curvature $\boldsymbol{\Omega} = \lambda^* \boldsymbol{\Omega}'$. If $\{e_1, \ldots, e_n\}$ is a basis for $\mathcal{G}$ and we write $\boldsymbol{\Omega}' = (\ \boldsymbol{\Omega}')^\alpha e_\alpha$, then $\tilde{f}(\boldsymbol{\Omega}') = \tilde{f}(e_{\alpha_1}, \ldots, e_{\alpha_k})(\boldsymbol{\Omega}')^{\alpha_1} \wedge \cdots \wedge (\boldsymbol{\Omega}')^{\alpha_k}$. Since pullback is linear and commutes with wedge products we find that $\tilde{f}(\boldsymbol{\Omega}) = \lambda^*(\tilde{f}(\boldsymbol{\Omega}'))$.

To show that the Chern-Weil homomorphisms agree we need only prove that $\bar{f}(\boldsymbol{\Omega})$ and $\bar{f}(\boldsymbol{\Omega}')$ determine the same cohomology class. We will, in fact, show that they are equal. Moreover, since projections to X are unique, it will suffice to prove that $\mathcal{P}^*(\bar{f}(\boldsymbol{\Omega}')) = \tilde{f}(\boldsymbol{\Omega})$. But

$$\begin{aligned}
\mathcal{P}^*(\bar{f}(\boldsymbol{\Omega}')) &= (\mathcal{P}' \circ \lambda)^*(\bar{f}(\boldsymbol{\Omega}')) \\
&= \lambda^*\Big((\mathcal{P}')^*(\bar{f}(\boldsymbol{\Omega}'))\Big) \\
&= \lambda^*\Big(\tilde{f}(\boldsymbol{\Omega}')\Big) \\
&= \tilde{f}(\boldsymbol{\Omega})
\end{aligned}$$

as required. ∎

Exercise 6.3.6 Show that the algebra of characteristic classes for a trivial bundle is the trivial subalgebra of $H^*_{\mathrm{de\,R}}(X; \quad)$. **Hint**: The product bundle $G \hookrightarrow X \times G \xrightarrow{\pi} X$ admits a flat connection (page 353, [**N4**]).

6.4 Chern Numbers

Since characteristic classes are cohomology classes on the base manifold X of the bundle, they can be integrated over submanifolds of X.

Exercise 6.4.1 Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a principal G-bundle, $\tilde{f} \in I^k(G)$, where $1 \leq k \leq \frac{1}{2} \dim X$, N a compact, oriented submanifold of X of dimension $2k$ and $\iota : N \hookrightarrow X$ the inclusion map. Show that

$$\int_N \iota^* \left(\bar{f}(\boldsymbol{\Omega})\right)$$

takes the same value for any connection on P. Show, moreover, that if N_1 and N_2 are two such submanifolds and N_1 can be deformed into N_2 in X (Exercise 4.7.10), then

$$\int_{N_1} \iota_1^*(\bar{f}(\boldsymbol{\Omega})) = \int_{N_2} \iota_2^*\left(\bar{f}(\boldsymbol{\Omega})\right).$$

Remark: These integrals are, in general, complex numbers. The integral of the corresponding cohomology class $[\bar{f}(\boldsymbol{\Omega})]$ is, of course, taken to be the integral of any of its representatives. If $\tilde{f}$ is real-valued, we have identified $[\tilde{f}(\boldsymbol{\Omega})]$ with a real cohomology class and the integral is real. This is the case, in particular, for the Chern classes (see the Remark on page 377). One can show that, in fact, the integral of a Chern class over a submanifold is always an integer and, in this sense, Chern classes are *integral* cohomology classes. We will verify this for the 2^{nd} Chern class of an $SU(2)$-bundle over S^4 shortly. When this is done, the reason for the strange normalizing constants in (6.2.6) will be clear: The $\boldsymbol{i}$ makes the Chern classes real and the 2π makes them integral.

A special case of the integrals discussed in Exercise 6.4.1 is of particular interest. Suppose that $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ is a principal G-bundle, where X is an oriented compact manifold of dimension $2n$. Then the n^{th} Chern class $c_n(P)$ is an element of $H^{2n}_{\text{de R}}(X; \;\;)$ and so can be integrated over X. The value of the integral

$$\int_X c_n(P)$$

is called the **Chern number** of the bundle. Chern numbers for equivalent bundles are, of course, the same. In Section 6.1 we computed the Chern number for each equivalence class of $U(1)$-bundles over S^2 and found that the result was precisely the integer n that appeared in the transition function $g_{SN}(\varphi, \theta) = e^{-n\theta \boldsymbol{i}}$ characterizing the bundle. Such bundles are therefore classified up to equivalence by their Chern numbers. We would now like to carry out a similar program for $SU(2)$-bundles over S^4. Specifically, we will let

$$SU(2) \hookrightarrow P \xrightarrow{\mathcal{P}} S^4$$

be some $SU(2)$-bundle over S^4 and will calculate the Chern number

$$\int_{S^4} c_2(P) = \frac{1}{8\pi^2} \int_{S^4} \text{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}). \tag{6.4.1}$$

The general plan is essentially the same as for $U(1)$-bundles over S^2 in Section 6.1. We trivialize the bundle over U_N and U_S. Choose a connection and pull it back to obtain local potentials $\boldsymbol{\mathcal{A}}_N$ and $\boldsymbol{\mathcal{A}}_S$ and local field strengths $\boldsymbol{\mathcal{F}}_N$ and $\boldsymbol{\mathcal{F}}_S$. Let S^4_+ and S^4_- be the upper and lower hemispheres of S^4, respectively, and denote by S^3 the equatorial 3-sphere $S^4_+ \cap S^4_-$ in S^4 (with the orientation it inherits from S^4_+). Then

$$\begin{aligned} &\frac{1}{8\pi^2} \int_{S^4} \text{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}) \\ &= \frac{1}{8\pi^2} \int_{S^4_+} \text{trace}(\boldsymbol{\mathcal{F}}_N \wedge \boldsymbol{\mathcal{F}}_N) + \frac{1}{8\pi^2} \int_{S^4_-} \text{trace}(\boldsymbol{\mathcal{F}}_S \wedge \boldsymbol{\mathcal{F}}_S) \end{aligned} \tag{6.4.2}$$

and we wish to apply Stokes' Theorem to each of the integrals on the right-hand side. To do so, of course, we need to express both trace$(\boldsymbol{\mathcal{F}}_N \wedge \boldsymbol{\mathcal{F}}_N)$ and trace$(\boldsymbol{\mathcal{F}}_S \wedge \boldsymbol{\mathcal{F}}_S)$ as exterior derivatives of 3-forms on U_N and U_S, respectively. This is possible, of course, since they are closed forms on discs, but we will require more explicit information.

Lemma 6.4.1 *Let $SU(2) \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be a principal $SU(2)$-bundle over X, $\boldsymbol{\omega}$ a connection on it with curvature $\boldsymbol{\Omega}, s : V \longrightarrow \mathcal{P}^{-1}(V)$ a local cross-section and $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{\omega}$ and $\boldsymbol{\mathcal{F}} = s^*\boldsymbol{\Omega}$ the local gauge potential and field strength. Then, on V,*

$$\mathrm{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}) = d\left(\mathrm{trace}\left(\boldsymbol{\mathcal{A}} \wedge d\boldsymbol{\mathcal{A}} + \frac{2}{3}\boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}}\right)\right). \tag{6.4.3}$$

Proof: Throughout the proof we will work exclusively on the trivial bundle $SU(2) \hookrightarrow \mathcal{P}^{-1}(V) \longrightarrow V$, but, for convenience, we will denote any relevant restrictions by the same symbo1, e.g., $\mathcal{P}$ is the projection on $\mathcal{P}^{-1}(V), \boldsymbol{\omega}$ is the connection on $SU(2) \hookrightarrow \mathcal{P}^{-1}(V) \xrightarrow{\mathcal{P}} V$, etc. We intend to apply (6.3.14), where $\tilde{f}$ is $\tilde{f}_2$, $\boldsymbol{\Omega}_1$ is $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ is the curvature of a connection whose existence we now ask the reader to demonstrate.

Exercise 6.4.2 Let $G \hookrightarrow P \xrightarrow{\mathcal{P}} X$ be any principal G-bundle, $s : V \longrightarrow \mathcal{P}^{-1}(V)$ a local cross-section and $\Psi : \mathcal{P}^{-1}(V) \longrightarrow V \times G$ the associated trivialization (so $s(x) = \Psi^{-1}(x, e)$ for every $x \in V$). Show that there exists a connection $\boldsymbol{\omega}_0$ on $G \hookrightarrow \mathcal{P}^{-1}(V) \longrightarrow V$ for which the local gauge potential $\boldsymbol{\mathcal{A}}_0 = s^*\boldsymbol{\omega}_0$ is identically zero. **Hint**: If $\boldsymbol{\Theta}$ is the Cartan 1-form for G and $\pi : V \times G \longrightarrow G$ is the projection, then $\pi^*\boldsymbol{\Theta}$ is a flat connection on $V \times G$ (page 353, [**N4**]). Show that $\boldsymbol{\omega}_0 = (\pi \circ \Psi)^*\boldsymbol{\Theta}$ has the required properties.

Now, (6.3.14) asserts that

$$\tilde{f}_2(\boldsymbol{\Omega}) - \tilde{f}_2(\boldsymbol{\Omega}_0) = d\boldsymbol{\Phi}, \tag{6.4.4}$$

where

$$\begin{aligned}
\boldsymbol{\Phi} &= 2\int_0^1 \tilde{f}_2(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t)dt \\
\boldsymbol{\alpha} &= \boldsymbol{\omega} - \boldsymbol{\omega}_0 \\
\boldsymbol{\omega}_t &= \boldsymbol{\omega}_0 + t\boldsymbol{\alpha} \\
\boldsymbol{\Omega}_t &= d^{\boldsymbol{\omega}_t}\boldsymbol{\omega}_t = d\boldsymbol{\omega}_0 + t\,d\boldsymbol{\alpha} + \frac{1}{2}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_t] \\
&= d\boldsymbol{\omega}_0 + t\,d\boldsymbol{\alpha} + \boldsymbol{\omega}_t \wedge \boldsymbol{\omega}_t.
\end{aligned}$$

Pulling (6.4.4) back by s gives

$$\bar{f}_2(\boldsymbol{\Omega}) - \bar{f}_2(\boldsymbol{\Omega}_0) = d(s^*\boldsymbol{\Phi})$$

from which we obtain, since $\boldsymbol{\mathcal{F}}_0 = 0$,

$$\operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}) = d(s^*\boldsymbol{\Phi}). \tag{6.4.5}$$

Now, we compute $s^*\boldsymbol{\Phi}$ as follows:

Exercise 6.4.3 Show that

$$\tilde{f}_2(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t) = \operatorname{trace}(\boldsymbol{\alpha} \wedge \boldsymbol{\Omega}_t).$$

But

$$\begin{aligned}\boldsymbol{\alpha} \wedge \boldsymbol{\Omega}_t &= \boldsymbol{\alpha} \wedge (d\boldsymbol{\omega}_0 + t\, d\boldsymbol{\alpha} + \boldsymbol{\omega}_t \wedge \boldsymbol{\omega}_t) \\ &= \boldsymbol{\alpha} \wedge d\boldsymbol{\omega}_0 + t\boldsymbol{\alpha} \wedge d\boldsymbol{\alpha} + \boldsymbol{\alpha} \wedge (\boldsymbol{\omega}_0 + t\boldsymbol{\alpha}) \wedge (\boldsymbol{\omega}_0 + t\boldsymbol{\alpha}) \\ &= \boldsymbol{\alpha} \wedge d\boldsymbol{\omega}_0 + t\boldsymbol{\alpha} \wedge d\boldsymbol{\alpha} + \boldsymbol{\alpha} \wedge \boldsymbol{\omega}_0 \wedge \boldsymbol{\omega}_0 + t\boldsymbol{\alpha} \wedge \boldsymbol{\omega}_0 \wedge \boldsymbol{\alpha} \\ &\qquad + t\boldsymbol{\alpha} \wedge \boldsymbol{\alpha} \wedge \boldsymbol{\omega}_0 + t^2\boldsymbol{\alpha} \wedge \boldsymbol{\alpha} \wedge \boldsymbol{\alpha}\end{aligned}$$

so

$$\begin{aligned}\boldsymbol{\Phi} &= 2\int_0^1 \tilde{f}_2(\boldsymbol{\alpha}, \boldsymbol{\Omega}_t)dt = 2\int_0^1 \operatorname{trace}(\boldsymbol{\alpha} \wedge \boldsymbol{\Omega}_t)dt \\ &= \operatorname{trace}\left(2\int_0^1 \boldsymbol{\alpha} \wedge \boldsymbol{\Omega}_t\, dt\right) \\ &= \operatorname{trace}\Big(2\boldsymbol{\alpha} \wedge d\boldsymbol{\omega}_0 + \boldsymbol{\alpha} \wedge d\boldsymbol{\alpha} + 2\boldsymbol{\alpha} \wedge \boldsymbol{\omega}_0 \wedge \boldsymbol{\omega}_0 + \boldsymbol{\alpha} \wedge \boldsymbol{\omega}_0 \wedge \boldsymbol{\alpha} \\ &\qquad + \boldsymbol{\alpha} \wedge \boldsymbol{\alpha} \wedge \boldsymbol{\omega}_0 + \frac{2}{3}\boldsymbol{\alpha} \wedge \boldsymbol{\alpha} \wedge \boldsymbol{\alpha}\Big).\end{aligned}$$

Noting that, by the way we have chosen $\boldsymbol{\omega}_0$, $s^*\boldsymbol{\omega}_0 = \boldsymbol{\mathcal{A}}_0 = 0$ and $s^*\boldsymbol{\alpha} = s^*(\boldsymbol{\omega} - \boldsymbol{\omega}_0) = \boldsymbol{\mathcal{A}}$ we obtain

$$s^*\boldsymbol{\Phi} = \operatorname{trace}\left(\boldsymbol{\mathcal{A}} \wedge d\boldsymbol{\mathcal{A}} + \frac{2}{3}\boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}}\right)$$

and the Lemma follows from (6.4.5). ∎

Note that, since $\boldsymbol{\mathcal{F}} = d\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}}$, (6.4.3) can be written

$$\operatorname{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}) = d\left(\operatorname{trace}\left(\boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{F}} - \frac{1}{3}\boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}} \wedge \boldsymbol{\mathcal{A}}\right)\right). \tag{6.4.6}$$

We now return to the evaluation of the integrals in (6.4.2). Applying Stokes' Theorem twice on the right-hand side and keeping in mind that the equatorial 3-sphere S^3 is assumed to have the orientation it inherits from S^4_+ we obtain

$$
\begin{aligned}
&\frac{1}{8\pi^2}\int_{S^4}\operatorname{trace}(\boldsymbol{\mathcal{F}}\wedge\boldsymbol{\mathcal{F}})\\
&\quad=\frac{1}{8\pi^2}\int_{S^3}\iota^*\left(\operatorname{trace}\left(\boldsymbol{\mathcal{A}}_N\wedge\boldsymbol{\mathcal{F}}_N-\frac{1}{3}\boldsymbol{\mathcal{A}}_N\wedge\boldsymbol{\mathcal{A}}_N\wedge\boldsymbol{\mathcal{A}}_N\right)\right)\\
&\qquad\quad-\frac{1}{8\pi^2}\int_{S^3}\iota^*\left(\operatorname{trace}\left(\boldsymbol{\mathcal{A}}_S\wedge\boldsymbol{\mathcal{F}}_S-\frac{1}{3}\boldsymbol{\mathcal{A}}_S\wedge\boldsymbol{\mathcal{A}}_S\wedge\boldsymbol{\mathcal{A}}_S\right)\right)\\
&\quad=\frac{1}{8\pi^2}\int_{S^3}\iota^*\left(\operatorname{trace}\left(\boldsymbol{\mathcal{A}}_N\wedge\boldsymbol{\mathcal{F}}_N-\frac{1}{3}\boldsymbol{\mathcal{A}}_N\wedge\boldsymbol{\mathcal{A}}_N\wedge\boldsymbol{\mathcal{A}}_N\right.\right.\\
&\qquad\quad\left.\left.-\boldsymbol{\mathcal{A}}_S\wedge\boldsymbol{\mathcal{F}}_S+\frac{1}{3}\boldsymbol{\mathcal{A}}_S\wedge\boldsymbol{\mathcal{A}}_S\wedge\boldsymbol{\mathcal{A}}_S\right)\right),
\end{aligned}
$$

where $\iota : S^3 \hookrightarrow S^4$ is the inclusion.

Exercise 6.4.4 Substitute $\boldsymbol{\mathcal{A}}_N = g_{SN}^{-1}\,\boldsymbol{\mathcal{A}}_S\,g_{SN} + g_{SN}^{-1}\,dg_{SN}$ and $\boldsymbol{\mathcal{F}}_N = g_{SN}^{-1}\,\boldsymbol{\mathcal{F}}_S g_{SN}$ into this last expression to show that

$$
\begin{aligned}
&\frac{1}{8\pi^2}\int_{S^4}\operatorname{trace}(\boldsymbol{\mathcal{F}}\wedge\boldsymbol{\mathcal{F}})\\
&\quad=\frac{1}{8\pi^2}\int_{S^3}\iota^*\Big[-\frac{1}{3}\operatorname{trace}((g_{SN}^{-1}\,dg_{SN})\wedge(g_{SN}^{-1}\,dg_{SN})\\
&\qquad\quad\wedge(g_{SN}^{-1}\,dg_{SN}))+d(\operatorname{trace}(\boldsymbol{\mathcal{A}}_S\,dg_{SN}\,g_{SN}^{-1}))\Big]
\end{aligned}
$$

and apply Stokes' Theorem to obtain

$$
\begin{aligned}
&\frac{1}{8\pi^2}\int_{S^4}\operatorname{trace}(\boldsymbol{\mathcal{F}}\wedge\boldsymbol{\mathcal{F}})\\
&\quad=-\frac{1}{24\pi^2}\int_{S^3}\iota^*\Big(\operatorname{trace}((g_{SN}^{-1}\,dg_{SN})\\
&\qquad\quad\wedge(g_{SN}^{-1}\,dg_{SN})\wedge\;(g_{SN}^{-1}\,dg_{SN}))\Big).
\end{aligned}
\tag{6.4.7}
$$

Notice that, as expected, the integral in (6.4.7), i.e., the Chern number, does not depend on the connection or the curvature, but only on the transition function g_{SN} of the bundle. To obtain a computationally more efficient formula we expand $g_{SN}^{-1}dg_{SN} = g_{SN}^{*}\boldsymbol{\Theta}$ in terms of the basis

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

for $su(2) \cong \operatorname{Im}$. Thus, let

$$
g_{SN}^{-1}\,dg_{SN}=\boldsymbol{\Theta}^1\boldsymbol{I}+\boldsymbol{\Theta}^2\boldsymbol{J}+\boldsymbol{\Theta}^3\boldsymbol{K}=\begin{pmatrix}\boldsymbol{\Theta}^1\boldsymbol{i} & \boldsymbol{\Theta}^2+\boldsymbol{\Theta}^3\boldsymbol{i}\\ -\boldsymbol{\Theta}^2+\boldsymbol{\Theta}^3\boldsymbol{i} & -\boldsymbol{\Theta}^1\boldsymbol{i}\end{pmatrix}.
$$

Exercise 6.4.5 Compute the matrix products and show that

$$\mathrm{trace}((g_{SN}^{-1}dg_{SN}) \wedge (g_{SN}^{-1}dg_{SN}) \wedge (g_{SN}^{-1}dg_{SN})) = -12\boldsymbol{\Theta}^1 \wedge \boldsymbol{\Theta}^2 \wedge \boldsymbol{\Theta}^3.$$

Thus,

$$\frac{1}{8\,\pi^2}\int_{S^4} \mathrm{trace}(\boldsymbol{\mathcal{F}} \wedge \boldsymbol{\mathcal{F}}) = \frac{1}{2\pi^2}\int_{S^3} \iota^*(\boldsymbol{\Theta}^1 \wedge \boldsymbol{\Theta}^2 \wedge \boldsymbol{\Theta}^3). \tag{6.4.8}$$

Our objective now is to show that the Chern number captures the topological type of an $SU(2)$-bundle over S^4. For this we recall that the equivalence class of such a bundle is uniquely determined by the homotopy type of $g_{SN} \mid S^3$ (this is essentially the Classification Theorem for bundles over spheres, but Lemma 4.4.1 and Theorem 4.4.2 of **[N4]** make the correspondence explicit). Our procedure then will be to select one representative from each homotopy class in $\pi_3(SU(2)) \cong \pi_3(S^3)$ and show that the bundles with these transition functions have different Chern numbers.

The notation is less cumbersome if we identify both S^3 and $SU(2)$ with the unit quaternions. Fix a base point $1 \in S^3$ and identify $\pi_3(S^3)$ with the set $[(S^3,1),(S^3,1)]$ of homotopy classes of maps $g : (S^3,1) \longrightarrow (S^3,1)$ (see page 154, **[N4]**). For each $n = 0,1,2,\ldots$ we let

$$g_n : (S^3,1) \longrightarrow (S^3,1)$$

be the restriction to S^3 of the map $\tilde{g}_n : \quad \longrightarrow \quad$ given by $\tilde{g}_n(q) = q^n$. Thus, $\iota \circ g_n = \tilde{g}_n \circ \iota$, where $\iota : S^3 \hookrightarrow \quad$ is the inclusion. We will show that the degree of g_n and the Chern number of the bundle whose $g_{SN} \mid S^3$ is g_n are both n. This is, of course, clear for $n = 0$ since the constant map g_0 has degree 0, while $dg_0 = 0$ and (6.4.7) shows that the Chern number of the corresponding bundle is also 0. Next notice that g_1 is the identity map on S^3 and so has degree 1. Furthermore, $g_1^{-1}dg_1 = q^{-1}dq$.[2pt]

Exercise 6.4.6 Show that, on ,

$$\begin{aligned} q^{-1}dq = {} & [(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2]^{-2} \\ & \cdot [(x^0dx^0 + x^1dx^1 + x^2dx^2 + x^3dx^3) \\ & + (x^0dx^1 - x^1dx^0 - x^2dx^3 + x^3dx^2)\boldsymbol{i} \\ & + (x^0dx^2 - x^2dx^0 - x^3dx^1 + x^1dx^3)\boldsymbol{j} \\ & + (x^0dx^3 - x^3dx^0 - x^1dx^2 + x^2dx^1)\boldsymbol{k}]. \end{aligned}$$

Show also that, on $S^3 \subseteq \quad$, the real part of $q^{-1}dq$ is zero and the wedge product of the $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ components is

$$x^0 dx^1 \wedge dx^2 \wedge dx^3 - x^1 dx^0 \wedge dx^2 \wedge dx^3 + \\ x^2 dx^0 \wedge dx^1 \wedge dx^3 - x^3 dx^0 \wedge dx^1 \wedge dx^2. \tag{6.4.9}$$

Now, observe that the restriction of the 3-form in (6.4.9) to S^3 is the standard volume form for S^3 (Section 4.6) so its integral over S^3 is $2\pi^2$. Thus, according to (6.4.8), the Chern number of the bundle corresponding to g_1 is $\frac{1}{2\pi^2}(2\pi^2) = 1$ which, in particular, coincides with $\deg(g_1)$.

For the general case we proceed as follows: The degree of a map $g : S^3 \longrightarrow S^3$ can be computed as

$$\deg(g) = \int_{S^3} g^* \boldsymbol{\omega},$$

where $\boldsymbol{\omega}$ is a 3-form on S^3 whose integral over S^3 is 1. Such a 3-form on S^3 is given by $\frac{1}{2\pi^2}$ times the standard volume form on S^3. According to Exercise 6.4.6, the volume form on S^3 is the restriction to S^3 of $\boldsymbol{\Theta}^1 \wedge \boldsymbol{\Theta}^2 \wedge \boldsymbol{\Theta}^3$, where $\boldsymbol{\Theta}^1 \boldsymbol{I} + \boldsymbol{\Theta}^2 \boldsymbol{J} + \boldsymbol{\Theta}^3 \boldsymbol{K}$ is the imaginary part of $q^{-1}dq$. By Exercise 6.4.5, this, in turn, is the restriction to S^3 of $-\frac{1}{12}\,\mathrm{trace}((q^{-1}dq) \wedge (q^{-1}dq) \wedge (q^{-1}dq))$. Thus,

$$\boldsymbol{\omega} = -\frac{1}{24\pi^2}\iota^*\Big(\mathrm{trace}((q^{-1}dq) \wedge (q^{-1}dq) \wedge (q^{-1}dq))\Big)$$

so

$$\begin{aligned} g^*\boldsymbol{\omega} &= -\frac{1}{24\pi^2}(\iota \circ g)^*\Big(\mathrm{trace}((q^{-1}dq) \wedge (q^{-1}dq) \wedge (q^{-1}dq))\Big) \\ &= -\frac{1}{24\pi^2}\,\mathrm{trace}\Big((\iota \circ g)^{-1}d(\iota \circ g) \wedge (\iota \circ g)^{-1}d(\iota \circ g) \\ &\qquad \wedge\ (\iota \circ g)^{-1}d(\iota \circ g)\Big). \end{aligned}$$

Now, for the map g_n defined above we have $\iota \circ g_n = \tilde{g}_n \circ \iota$ so

$$\begin{aligned} g_n^*\boldsymbol{\omega} &= -\frac{1}{24\pi^2}\,\mathrm{trace}\Big((\tilde{g}_n \circ \iota)^{-1}d(\tilde{g}_n \circ \iota) \\ &\qquad \wedge (\tilde{g}_n \circ \iota)^{-1}d(\tilde{g}_n \circ \iota) \wedge (\tilde{g}_n \circ \iota)^{-1}d(\tilde{g}_n \circ \iota)\Big) \\ &= -\frac{1}{24\pi^2}\iota^*\Big(\mathrm{trace}((\tilde{g}_n^{-1}d\tilde{g}_n) \wedge (\tilde{g}_n^{-1}d\tilde{g}_n) \wedge (\tilde{g}_n^{-1}d\tilde{g}_n))\Big). \end{aligned}$$

Integrating both sides over S^3 shows that the degree of g_n is the same as the Chern number of the corresponding bundle. To show that these are both equal to n we need the following identity.

Exercise 6.4.7 Let $g, h : S^3 \longrightarrow S^3$ and define $gh : S^3 \longrightarrow S^3$ by $(gh)(q) = g(q)h(q)$, where the product on the right-hand side is the quaternion product.

Show that

$$\begin{aligned}\operatorname{trace}((gh)^{-1}d(gh) \wedge (gh)^{-1}d(gh) \wedge (gh)^{-1}d(gh)) \\ = \operatorname{trace}(g^{-1}dg \wedge g^{-1}dg \wedge g^{-1}dg) \\ + \operatorname{trace}(h^{-1}dh \wedge h^{-1}dh \wedge h^{-1}dh) \\ + d(3 \operatorname{trace}(g^{-1}dg \wedge h^{-1}dh)).\end{aligned} \tag{6.4.10}$$

Since $g_2 = g_1 g_1, g_3 = g_2 g_1, \ldots$ and since the integral over S^3 of the last term in (6.4.10) is zero by Stokes' Theorem, we find from (6.4.7) and induction that the Chern number of the bundle corresponding to g_n (and therefore the degree of g_n) is n. All that remains of our stated objective then is to extend this to the negative integers as well. Thus, for $n = -1, -2, \ldots$ we let $g_n : (S^3, 1) \longrightarrow (S^3, 1)$ be the restriction to S^3 of the map $\tilde{g}_n : \quad -\{0\} \longrightarrow \quad -\{0\}$ given by $\tilde{g}_n(q) = q^n$.

Exercise 6.4.8 Show that $\deg(g_{-1})$ and the Chern number of the bundle corresponding to g_{-1} are both -1. **Hints**: Show that g_{-1} is an orientation reversing diffeomorphism of S^3 and so has degree -1. Next note that, on $S^3, g_{-1}^{-1} dg_{-1} = q d\bar{q}$ and perform calculations analogous to those outlined in Exercise 6.4.6.

Exercise 6.4.9 Use (6.4.10) to show that, for each $n = -1, -2, \ldots, \deg(g_n)$ and the Chern number of the bundle corresponding to g_n are both n.

For the record, we combine what we have just proved with the results of Section 6.1 to obtain the following theorem.

Theorem 6.4.2 *The Chern number for a $U(1)$-bundle over S^2 or an $SU(2)$-bundle over S^4 is an integer that uniquely determines the equivalence class of the bundle.*

Remark: In the physics literature a bundle with Chern number n is said to have **topological charge** $-n$ (think of the minus sign as simply an attempt to keep life interesting). For $U(1)$-bundles the topological charge is also called the **magnetic charge**, while for $SU(2)$-bundles it is known as the **instanton number** of the bundle. The idea here is that such a bundle is characterized topologically by the strength of the Dirac monopole or instanton that can live on it (see Sections 2.2 and 2.5, respectively).

For $U(n)$- and $SU(n)$-bundles the algebra of characteristic classes is generated by the Chern classes so, in particular, any formal linear combination of Chern classes is itself a characteristic class. For example, the sum

$$c(P) = 1 + c_1(P) + c_2(P) + \cdots$$

of all the Chern classes (necessarily finite since $c_k(P) = 0$ when $2k > \dim X$) is called the **total Chern class** of the bundle. For bundles with other structure groups one can often isolate analogous finite generating sets and build from them interesting and useful characteristic classes. Those interested in pursuing these matters should proceed to Chapter XII of [**KN2**] and, for applications to physics, to [**EGH**]. We intend now to take up a rather different means of associating with certain bundles various cohomology classes of the base manifold, one of which will tell us whether or not a spacetime manifold admits a spinor structure. The "cohomology" is not that of de Rham, however.

6.5 $_2$-Čech Cohomology for Smooth Manifolds

The cohomology theory relevant to the question of whether or not a spacetime admits spinor fields is called Čech cohomology with coefficients in the group $\mathbb{Z}_2$. This can be viewed as a (very) special case of what is called "sheaf cohomology," but in this section we take a more pedestrian approach. We construct the groups of interest explicitly from a simple cover of the manifold X and discuss only briefly the sheaf-theoretic proof that the construction is actually independent of the choice of the simple cover. We then show how the orthonormal frame bundle of a semi-Riemannian manifold X gives rise to a Čech cohomology class of X (the 1^{st} Stiefel-Whitney class) which we show represents the obstruction to orientability in the sense that X is orientable if and only if this cohomology class is trivial. When X is an oriented and time oriented spacetime, another such Čech cohomology class (the 2^{nd} Stiefel-Whitney class) is shown to represent the obstruction to the existence of a spinor structure for X. Since diffeomorphic manifolds have isomorphic Čech cohomology groups we arrive at the rather surprising conclusion that the existence of spinor structures is a topological matter and does not depend on the particular Lorentz metric defined on the manifold. Before immersing ourselves in the intricacies of a new cohomology theory, however, we will try to get some feel for what the topology of a spacetime X has to do with one's ability to define spinor fields on X (a quick review of Sections 2.4 and 3.5 might be in order here).

We consider then an oriented and time oriented spacetime X. Of course, if the topology of X is trivial ($X \cong {}^4$), then any principal bundle over X is likewise trivial so, in particular, its oriented, time oriented, orthonormal frame bundle can be identified with the product bundle

$$\mathcal{L}_+^\uparrow \hookrightarrow X \times \mathcal{L}_+^\uparrow \longrightarrow X.$$

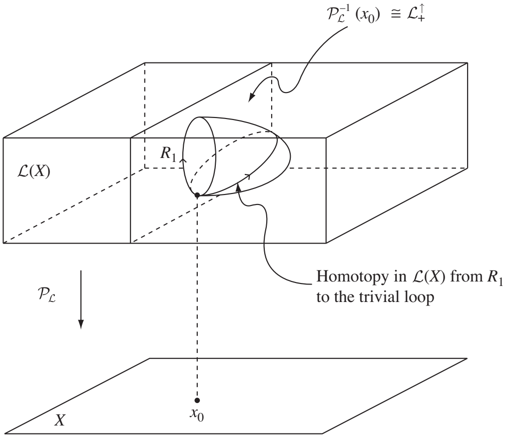
In this case one can build a spinor structure for X from the product bundle

$$SL(2, \ ) \hookrightarrow X \times SL(2, \ ) \longrightarrow X$$

by simply defining λ to be the spinor map Spin on each fiber $\{x\} \times SL(2, \ )$. Indeed, X need not be trivial for this to work. All that's required is that the frame bundle be trivial and this is the case whenever it has a global cross-section, i.e., whenever X has a global orthonormal frame field

$$s : X \longrightarrow \mathcal{L}(X) \qquad (P_{\mathcal{L}} \circ s \ = id_X).$$

Suppose now that the frame bundle $\mathcal{L}(X)$ is not trivial. It will simplify the discussion and not obscure any of the essential ideas if we assume that X is simply connected. In this case it can occur that the frame bundle $\mathcal{L}(X)$ is also simply connected (this cannot occur when the frame bundle is trivial since $\pi_1(X \times \mathcal{L}_+^{\uparrow}) \cong \pi_1(X) \oplus \pi_1(\mathcal{L}_+^{\uparrow}) \cong \pi_1(X) \oplus \mathbb{Z}_2$ which is just $\mathbb{Z}_2$ when X is simply connected). Now, fix an $x_0 \in X$ and consider the fiber $\mathcal{P}_{\mathcal{L}}^{-1}(x_0) \cong \mathcal{L}_+^{\uparrow}$ above x_0. In Section 3.4 we constructed a curve $R_1(t), 0 \leq t \leq 2\pi$, in $\mathcal{L}_+^{\uparrow}$ representing a continuous rotation of the spatial coordinate axes through 360°. We now regard R_1 as a curve in the fiber $\mathcal{P}_{\mathcal{L}}^{-1}(x_0)$. Within $\mathcal{P}_{\mathcal{L}}^{-1}(x_0), R_1$ is *not* nullhomotopic. However, if $\mathcal{L}(X)$ is simply connected, the loop $R_1(t)$, $0 \leq t \leq 2\pi$, *is* nullhomotopic in $\mathcal{L}(X)$ (see the figure).



But a Dirac spinor field (for example) must change sign under a 360° rotation at x_0 (because R_1 lifts to a path in $SL(2,\ )$ from $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ to $-\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$) and obviously does not change sign under "no rotation at all." Since the wavefunction cannot change continuously from ψ to $-\psi$ but would have to do so over the homotopy in $\mathcal{L}(X)$ from R_1 to the trivial loop, it is not possible to unambiguously define such spinor fields on X. With this as motivation we now set about finding a characteristic class, the vanishing of which prohibits this sort of topological obstruction.

Recall (Section 3.3) that an open cover $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ of a manifold X is said to be simple if any finite intersection $U_{\alpha_0} \cap \cdots \cap U_{\alpha_j}$ of its elements is either empty or diffeomorphic to $\ ^n$ (where $n = \dim X$). Smooth manifolds admit simple covers and, indeed, any open cover of a smooth manifold has a countable, locally finite, simple open refinement in which each element has compact closure. Our construction of the Čech cohomology groups begins with the selection of a locally finite, simple open cover $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ for X with each $\bar{U}_k$ compact. We first build the Čech cohomology groups $\check{H}^j(\mathcal{U};\mathbb{Z}_2)$ corresponding to this cover and then discuss how one goes about showing that the construction is actually independent of this choice. We will identify $\mathbb{Z}_2$ with the multiplicative group $\{-1,1\}$.

For each integer $j \geq 0$, a ***j*-simplex** (for $\mathcal{U}$) is an ordered $(j+1)$-tuple $\sigma = (k_0,\ldots,k_j)$ of indices for which $U_{k_0} \cap \cdots \cap U_{k_j}$ is nonempty (and therefore diffeomorphic to $\;^n$). The **support** of σ is the nonempty open set $|\sigma| = U_{k_0} \cap \cdots \cap U_{k_j}$. For $i = 0,\ldots,j$, the ***i*th-face** of the j-simplex $\sigma = (k_0,\ldots,k_j)$ is the $(j-1)$-simplex $\sigma^i = (k_0,\ldots,k_{i-1},k_{i+1},\ldots,k_j)$. The set of all j-simplexes for $j \geq 0$ is called the **nerve** of $\mathcal{U}$ and denoted $N(\mathcal{U})$. For each $j \geq 0$, a **($\mathbb{Z}_2$-Čech) *j*-cochain** is a function f which assigns to each j-simplex $\sigma = (k_0,\ldots,k_j)$ an element $f(\sigma) = f(k_0,\ldots,k_j)$ of $\mathbb{Z}_2$ and that is totally symmetric, i.e.,

$$f\Big(k_{\tau(0)},\ldots,k_{\tau(j)}\Big) = f(k_0,\ldots,k_j)$$

for all permutations $\tau \in S_{j+1}$. The set $\check{C}^j(\mathcal{U};\mathbb{Z}_2)$ of all j-cochains is an Abelian group under pointwise multiplication, i.e.,

$$(fg)(k_0,\ldots,k_j) = f(k_0,\ldots,k_j)\ g(k_0,\ldots,k_j)$$

for all j-simplexes $(k_0,\ldots,k_j)$. Notice that the identity element of this group assigns the value $1 \in \mathbb{Z}_2$ to each j-simplex and that every element of $\check{C}^j(\mathcal{U};\mathbb{Z}_2)$ is its own inverse.

Before proceeding with the construction we would like to persuade the reader that Čech cochains actually arise in nature, as it were. Suppose then that X is a semi-Riemannian manifold and consider the orthonormal frame bundle $O(m,n-m) \hookrightarrow F(X) \xrightarrow{\mathcal{P}} X$. Select a simple cover $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ for X of the sort described above. Since each U_k is diffeomorphic to $\;^n$, the bundle is trivial over U_k and so we may select an orthonormal frame field on U_k, i.e., a cross-section $s_k : U_k \longrightarrow F(X)$. Do this for each element of $\mathcal{U}$. Now, if k_0 and k_1 are two indices for which $U_{k_0} \cap U_{k_1} \neq \emptyset$, then s_{k_0} and s_{k_1} are related on $U_{k_0} \cap U_{k_1}$ by $s_{k_1} = s_{k_0} \cdot g_{k_0k_1}$, where $g_{k_0k_1}$ is the transition function. Since this transition function takes values in $O(m,n-m)$, its determinant is either 1 or -1 at each point $U_{k_0} \cap U_{k_1}$. But $U_{k_0} \cap U_{k_1}$ is diffeomorphic to $\;^n$ (and therefore connected) so $\det(g_{k_0k_1})$ is either 1 everywhere on $U_{k_0} \cap U_{k_1}$ or -1 everywhere on $U_{k_0} \cap U_{k_1}$. Thus, we may define $f(k_0,k_1) \in \mathbb{Z}_2$ by

$$f(k_0,k_1) = \det\Big(g_{k_0k_1}\Big).$$

Since $\det(g_{k_1k_0}) = \det({g_{k_0k_1}}^{-1}) = (\det(g_{k_0k_1}))^{-1} = \det(g_{k_0k_1})$,

$$f(k_1, k_0) = f(k_0, k_1)$$

and we have defined a Čech 1-cochain $f \in \check{\mathrm{C}}(\mathcal{U}; \mathbb{Z}_2)$.

It will be convenient to define $\check{C}^j(\mathcal{U}; \mathbb{Z}_2)$ for $j < 0$ to be the trivial group $\{1\}$. Now, for each j we define a **coboundary operator**

$$\delta^j : \check{C}^j(\mathcal{U}; \mathbb{Z}_2) \longrightarrow \check{C}^{j+1}(\mathcal{U}; \mathbb{Z}_2)$$

as follows: For $j < 0, \delta^j$ is the trivial homomorphism, while, for $j \geq 0$,

$$\begin{aligned}(\delta^j f)(k_0, \ldots, k_{j+1}) &= f(k_1, \ldots, k_{j+1}) f(k_0, k_2, \ldots, k_{j+1}) \\ &\qquad \cdots f(k_0, \ldots, k_j) \\ &= \prod_{i=0}^{j+1} f\left(k_0, \ldots, \hat{k}_i, \ldots, k_{j+1}\right).\end{aligned}$$

Since f is totally symmetric, so is $\delta^j f$ and therefore $\delta^j f \in \check{C}^{j+1}(\mathcal{U}; \mathbb{Z}_2)$.

Exercise 6.5.1 Show that, because $\mathbb{Z}_2$ is Abelian, each δ^j is a homomorphism.

Notice that, if $f \in \check{C}^0(\mathcal{U}; \mathbb{Z}_2)$, then

$$(\delta^0 f)(k_0, k_1) = f(k_0) f(k_1)$$

and

$$\begin{aligned}(\delta^1(\delta^0 f))(k_0, k_1, k_2) &= (\delta^0 f)(k_1, k_2)(\delta^0 f)(k_0, k_2)(\delta^0 f)(k_0, k_1) \\ &= f(k_1) f(k_2) f(k_0) f(k_2) f(k_0) f(k_1) \\ &= (f(k_0))^2 (f(k_1))^2 (f(k_2))^2 \\ &= 1.\end{aligned}$$

Thus, $\delta^1 \circ \delta^0$ is the trivial homomorphism. We claim that this is true in general.

Lemma 6.5.1 $\delta^{j+1} \circ \delta^j : \check{C}^j(\mathcal{U}; \mathbb{Z}_2) \longrightarrow \check{C}^{j+2}(\mathcal{U}; \mathbb{Z}_2)$ *is the trivial homomorphism.*

Remark: As is customary one generally drops the indices on the coboundary operators, writing them all as δ and abbreviating Lemma 6.5.1 as $\delta^2 = 1$.

Proof: For $j < 0$ the result is obvious so assume $j \geq 0$. Let $f \in \check{C}^j(\mathcal{U}; \mathbb{Z}_2)$ and let $(k_0, \ldots, k_{j+2})$ be a $(j+2)$-simplex. We must show that $((\delta^{j+1} \circ \delta^j) f)(k_0, \ldots, k_{j+2}) = 1$. But

$$
\begin{aligned}
(\delta^{j+1}(\delta^j f))(k_0,\ldots,k_{j+2}) &= \prod_{i=0}^{j+2}(\delta^j f)\left(k_0,\ldots,\hat{k}_i,\ldots,k_{j+2}\right)\\
&= (\delta^j f)(k_1,k_2,\ldots,k_{j+2})(\delta^j f)(k_0,k_2,\ldots,k_{j+2})\\
&\qquad \cdots(\delta^j f)(k_0,k_1,\ldots,k_{j+1})\\
&= \Big[f(k_2,k_3,\ldots,k_{j+2})f(k_1,k_3,\ldots,k_{j+2})\cdots f(k_1,k_2,\ldots,k_{j+1})\Big]\\
&\quad \cdot\Big[f(k_2,k_3,\ldots,k_{j+2})f(k_0,k_3,\ldots,k_{j+2})\cdots f(k_0,k_2,\ldots,k_{j+1})\Big]\cdot\cdots\\
&\quad \cdot\Big[f(k_1,k_2,\ldots,k_{j+1})f(k_0,k_2,\ldots,k_{j+1})\cdots f(k_0,k_1,\ldots,k_j)\Big].
\end{aligned}
$$

Each of these factors has precisely two of the indices $k_0,\ldots,k_{j+2}$ missing. Moreover, for any pair of indices, there are precisely two factors in which these indices are missing. Thus, any factor that takes the value -1 occurs precisely twice and the product must be 1. ■

Lemma 6.5.1 asserts that the sequence of cochain groups and coboundary operators forms a cochain complex (Section 5.3) and so we may build its cohomology theory in the usual way. A **Čech *j*-coboundary** is an element of $\check{C}^j(\mathcal{U};\mathbb{Z}_2)$ that is in the image of δ^{j-1} and the set

$$\check{B}^j(\mathcal{U};\mathbb{Z}_2) = \text{Image }\delta^{j-1}$$

of all such is a subgroup of $\check{C}^j(\mathcal{U};\mathbb{Z}_2)$. A **Čech *j*-cocycle** is an element of $\check{C}^j(\mathcal{U};\mathbb{Z}_2)$ that is in the kernel of δ^j and the set

$$\check{Z}^j(\mathcal{U};\mathbb{Z}_2) = \ker\delta^j$$

of all such is a subgroup of $\check{C}^j(\mathcal{U};\mathbb{Z}_2)$. Lemma 6.5.1 asserts that

$$\check{B}^j(\mathcal{U};\mathbb{Z}_2) \subseteq \check{Z}^j(\mathcal{U};\mathbb{Z}_2)$$

so $\check{B}^j(\mathcal{U};\mathbb{Z}_2)$ is a (normal) subgroup of $\check{Z}^j(\mathcal{U};\mathbb{Z}_2)$. The quotient group

$$\check{H}^j(\mathcal{U};\mathbb{Z}_2) = \check{Z}^j(\mathcal{U};\mathbb{Z}_2)/\check{B}^j(\mathcal{U};\mathbb{Z}_2)$$

is the **j^{th} $\mathbb{Z}_2$-Čech cohomology group** of the cover $\mathcal{U}$. The elements of $\check{H}^j(\mathcal{U};\mathbb{Z}_2)$ are equivalence classes $[f]$ of j-cocycles ($\delta^j f = 1$), where the equivalence relation is $f' \sim f$ if and only if $f'f^{-1}$ is a coboundary, i.e.,

$$f' = (\delta^{j-1}h)f$$

for some $h \in \check{C}^{j-1}(\mathcal{U};\mathbb{Z}_2)$.

Let us pause for a moment to compute a simple example. Thus, we consider again a semi-Riemannian manifold X with simple cover $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ and orthonormal frame bundle $O(m,n-m) \hookrightarrow F(X) \xrightarrow{\mathcal{P}} X$. Choose an orthonormal frame field $s_k : U_k \longrightarrow F(X)$ on each U_k and consider the 1-cochain $f(k_0, k_1) = \det(g_{k_0k_1})$ constructed on page 391. We claim that f is a 1-cocycle and therefore determines a cohomology class in $\check{H}^1(\mathcal{U}; \mathbb{Z}_2)$. To see this we let (k_0, k_1, k_2) be a 2-simplex and compute

$$\begin{aligned}(\delta^1 f)(k_0, k_1, k_2) &= f(k_1, k_2)\, f(k_0, k_2)\, f(k_0, k_1)\\ &= \det\Big(g_{k_1k_2}\Big)\, \det\Big(g_{k_0k_2}\Big)\, \det\Big(g_{k_0k_1}\Big)\\ &= \det\Big(g_{k_0k_1}\Big)\, \det\Big(g_{k_1k_2}\Big)\, \det\Big(g_{k_2k_0}\Big)\\ &= \det\Big(g_{k_0k_1}\; g_{k_1k_2}\; g_{k_2k_0}\Big)\\ &= 1\end{aligned}$$

since, by the cocycle condition satisfied by transition functions (page 32), $g_{k_0k_1}\; g_{k_1k_2}\; g_{k_2k_0}$ is the identity matrix. Thus, f determines

$$[f] \in \check{H}^1(\mathcal{U}; \mathbb{Z}_2).$$

One disturbing aspect of this last example is that the cohomology class $[f]$ is not obviously intrinsic to the topology of X as is the case, for example, for de Rham cohomology classes. Indeed, it appears to depend not only on the simple cover $\mathcal{U}$ from which it is constructed, but even on the particular orthonormal frame fields $s_k : U_k \longrightarrow F(X)$ selected on each of the elements of $\mathcal{U}$. Let us show that at least this last dependence on the cross-sections s_k is only apparent. Thus, we let $s'_k : U_k \longrightarrow F(X)$ be another orthonormal frame field on U_k for each $k = 1, 2, \ldots$. Both s_k and s'_k give rise to trivializations of the bundle defined on $\mathcal{P}^{-1}(U_k)$ and these trivializations are related by a transition function g_k defined on U_k (so $s'_k = s_k \cdot g_k$ on U_k).

Exercise 6.5.2 Let (k_0, k_1) be a 1-simplex with $s_{k_1} = s_{k_0} \cdot g_{k_0k_1}$ and $s'_{k_1} = s'_{k_0} \cdot g'_{k_0k_1}$ on $U_{k_0} \cap U_{k_1}$. Show that $g'_{k_0k_1} = g_{k_0}^{-1} g_{k_0k_1} g_{k_1}$ on $U_{k_0} \cap U_{k_1}$.

Now, defining a 1-cocycle f' just as we did for f, but using the cross-section s' rather than s gives

$$\begin{aligned}f'(k_0, k_1) = \det\Big(g_{k_0k'_1}\Big) &= \det\Big(g_{k_0}^{-1} g_{k_0k_1}\; g_{k_1}\Big)\\ &= \Big(\det\Big(g_{k_0}\Big)\, \det\Big(g_{k_1}\Big)\Big)\, f(k_0, k_1).\end{aligned}$$

Exercise 6.5.3 Show that $h(k) = \det(g_k)$ defines an element h of $\check{C}^0(\mathcal{U}; \mathbb{Z}_2)$ and that

$$(\delta^0 h)(k_0, k_1) = \det\Big(g_{k_0}\Big)\, \det\Big(g_{k_1}\Big).$$

We conclude from Exercise 6.5.3 that

$$f'(k_0, k_1) = \Big((\delta^0 h)(k_0, k_1)\Big) f(k_0, k_1),$$

i.e.,

$$f' = (\delta^0 h) f$$

so f and f' differ by a 0-coboundary. Thus, $[f'] = [f]$ and the cohomology class we constructed above does not depend on the choice of the local orthonormal frames.

Exercise 6.5.4 Show that, if $\mathcal{U}$ consists of the single open set $U_0 = X$ (so that, in particular, X is diffeomorphic to $\ ^n$), then $\check{H}^0(\mathcal{U}; \mathbb{Z}_2) \cong \mathbb{Z}_2$, but $\check{H}^j(\mathcal{U}; \mathbb{Z}_2)$ is trivial for every other j.

The issue of whether or not our construction is independent of the choice of the simple cover $\mathcal{U}$ is not so easily resolved. Indeed, the standard argument by which this independence is established employs the methods of axiomatic sheaf cohomology (for a clear and concise exposition that proceeds from the basics to the result in question see Chapter IV, Section A, and Chapter VI, Sections A, B and D of [**GR**]). We will content ourselves here with a few elementary observations. Notice first that since any two simple covers of X of the type employed in our construction have a common refinement of the same type it would suffice to prove that $\check{H}^j(\mathcal{U}; \mathbb{Z}_2) \cong \check{H}^j(\mathcal{V}; \mathbb{Z}_2)$ whenever $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ and $\mathcal{V} = \{V_l\}_{l=1,2,\ldots}$ are two such covers and $\mathcal{V}$ is a refinement of $\mathcal{U}$. In this case one can select a refining map $\rho : \{1, 2, \ldots\} \longrightarrow \{1, 2, \ldots\}$ with the property that $V_l \subseteq U_{\rho(l)}$ for each $l = 1, 2, \ldots$. Notice that ρ is generally not uniquely determined, nor is it one-to-one in general. However, any such ρ must carry a j-simplex of $\mathcal{V}$ onto a j-simplex of $\mathcal{U}$ since $V_{l_0} \cap \cdots \cap V_{l_j} \neq \emptyset$ implies $U_{\rho(l_0)} \cap \cdots \cap U_{\rho(l_j)} \neq \emptyset$. Thus, we can define, for $j \geq 0$, a homomorphism

$$\rho^j : \check{C}^j(\mathcal{U}; \mathbb{Z}_2) \longrightarrow \check{C}^j(\mathcal{V}; \mathbb{Z}_2)$$

by

$$(\rho^j f)(l_0, \ldots, l_j) = f\Big(\rho(l_0), \ldots, \rho(l_j)\Big)$$

for each $f \in \check{C}^j(\mathcal{V}; \mathbb{Z}_2)$ and every j-simplex $(l_0, \ldots, l_j)$ for $\mathcal{V}$. For $j < 0$ we take $\rho^j : \check{C}^j(\mathcal{U}; \mathbb{Z}_2) \longrightarrow \check{C}^j(\mathcal{V}; \mathbb{Z}_2)$ to be the trivial homomorphism.

Lemma 6.5.2 *For each integer j, $\rho^{j+1} \circ \delta^j = \delta^j \circ \rho^j$.*

$$\begin{array}{ccccccc}
\cdots & \longrightarrow & \check{C}^j(\mathcal{U};\mathbb{Z}_2) & \xrightarrow{\delta^j} & \check{C}^{j+1}(\mathcal{U};\mathbb{Z}_2) & \longrightarrow & \cdots \\
 & & \rho^j \big\downarrow & & \big\downarrow \rho^{j+1} & & \\
\cdots & \longrightarrow & \check{C}^j(\mathcal{V};\mathbb{Z}_2) & \xrightarrow[\delta^j]{} & \check{C}^{j+1}(\mathcal{V};\mathbb{Z}_2) & \longrightarrow & \cdots
\end{array}$$

Proof: For $j < 0$ this is trivial so assume $j \geq 0$. Let $f \in \check{C}^j(\mathcal{U};\mathbb{Z}_2)$. Then, for any $(j+1)$-simplex $(l_0, \ldots, l_{j+1})$ in $\mathcal{V}$,

$$\begin{aligned}
((\delta^j \circ \rho^j)f)(l_0, \ldots, l_{j+1}) &= \delta^j(\rho^j f)(l_0, \ldots, l_{j+1}) \\
&= \prod_{i=0}^{j+1} (\rho^j f)\left(l_0, \ldots, \hat{l}_i, \ldots, l_{j+1}\right) \\
&= \prod_{i=0}^{j+1} f\ \left(\rho(l_0), \ldots, \widehat{\rho(l_i)}, \ldots, \rho(l_{j+1})\right) \\
&= (\delta^j f)\Big(\rho(l_0), \ldots, \rho(l_{j+1})\Big) \\
&= (\rho^{j+1}(\delta^j f))(l_0, \ldots, l_{j+1}) \\
&= ((\rho^{j+1} \circ \delta^j)f)(l_0, \ldots, l_{j+1}).
\end{aligned}$$

Thus, $(\delta^j \circ \rho^j)f = (\rho^{j+1} \circ \delta^j)f$ for any $f \in \check{C}^j(\mathcal{U};\mathbb{Z}_2)$, as required. ∎

According to Lemma 6.5.2 the maps $\rho^j : \check{C}^j(\mathcal{U};\mathbb{Z}_2) \longrightarrow \check{C}^j(\mathcal{V};\mathbb{Z}_2)$ determine a cochain map and so induce maps

$$(\rho^j)^{\#} : \check{H}^j(\mathcal{U};\mathbb{Z}_2) \longrightarrow \check{H}^j(\mathcal{V};\mathbb{Z}_2)$$

in cohomology. These homomorphisms are actually independent of the choice of the refining map as the reader may wish to verify in the following exercise.

Exercise 6.5.5 Let $\tau : \{1, 2, \ldots\} \longrightarrow \{1, 2, \ldots\}$ be another map with the property that $V_l \subseteq U_{\tau(l)}$ for each $l = 1, 2, \ldots$ For each j, let $\tau^j : \check{C}^j(\mathcal{U};\mathbb{Z}_2) \longrightarrow \check{C}^j(\mathcal{V};\mathbb{Z}_2)$ be the corresponding homomorphism. Define $K^j : \check{C}^j(\mathcal{U};\mathbb{Z}_2) \longrightarrow \check{C}^{j-1}(\mathcal{V};\mathbb{Z}_2)$ for $j > 0$ by

$$(K^j f)(l_0, \ldots, l_{j-1}) = \prod_{i=0}^{j-1} f\Big(\rho(l_0), \ldots, \rho(l_i), \tau(l_i), \ldots, \tau(l_{j-1})\Big)$$

and let K^j be the trivial homomorphism for $j \leq 0$. Show that

$$(\rho^j)(\tau^j)^{-1} = (\delta^{j-1} \circ K^j)(K^{j+1} \circ \delta^j)$$

and conclude that the cochain maps $\{\rho^j\}$ and $\{\tau^j\}$ are algebraically homotopic and therefore induce the same maps in cohomology.

The next step (which we do not intend to take) would be to show that the homomorphisms $(\rho^j)^\#$ are, in fact, isomorphisms (again, we refer to [**GR**]). Granting this, we will henceforth call the groups we have constructed the **$\mathbb{Z}_2$ - Čech cohomology groups of** X and denote them $\check{H}^j(X;\mathbb{Z}_2)$. Any calculation of these groups for some specific manifold X would, of course, begin with the selection of some convenient simple cover. For example, Exercise 6.5.4 now implies that $\check{H}^0(\ ^n;\mathbb{Z}_2) \cong \mathbb{Z}_2$ and $\check{H}^j(\ ^n;\mathbb{Z}_2)$ is trivial for every other j. Our particular concern will be with isolating certain specific cohomology classes which act as "obstructions" to the existence of various desirable structures on a manifold. For example, the cohomology class in $\check{H}^1(X;\mathbb{Z}_2)$ associated with the orthonormal frame bundle $O(m,n-m) \hookrightarrow F(X) \xrightarrow{\mathcal{P}} X$ of a semi-Riemannian manifold X in the manner described above is called the **1st Stiefel-Whitney class** of X and is denoted $w_1(X)$. The next order of business is to show that the 1st Stiefel-Whitney class $w_1(X) \in \check{H}^1(X;\mathbb{Z}_2)$ is the obstruction to orientability for a semi-Riemannian manifold X.

Theorem 6.5.3 *Let X be a semi-Riemannian manifold and $w_1(X) \in \check{H}^1(X;\mathbb{Z}_2)$ its 1st Stiefel-Whitney class. Then X is orientable if and only if $w_1(X)$ is trivial.*

Proof: We select an arbitrary locally finite, simple cover $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ for X and identify $\check{H}^j(X;\mathbb{Z}_2)$ with $\check{H}^j(\mathcal{U};\mathbb{Z}_2)$ for each j. Suppose first that X is orientable. Then we may select local orthonormal frame fields $s_k : U_k \longrightarrow F(X)$ consistent with the orientation of X. In particular, the transition functions $g_{k_0k_1}$ that relate these local orthonormal frames are in $SO(m,n-m)$ and therefore satisfy $\det(g_{k_0k_1}) = 1$ for each 1-simplex (k_0,k_1). But then the 1-cocycle f defined by $f(k_0,k_1) = \det(g_{k_0k_1})$, whose cohomology class is $w_1(X)$, is identically equal to 1 so $w_1(X)$ is the identity element of $\check{H}^1(\mathcal{U};\mathbb{Z}_2)$.

Now suppose, conversely, that $w_1(X)$ is trivial. Select local orthonormal frame fields $s_k : U_k \longrightarrow F(X),\ k = 1,2,\ldots$ and let $g_{k_0k_1}$ be the corresponding transition functions. Let f be the 1-cocycle defined by $f(k_0,k_1) = \det(g_{k_0k_1})$. Since $w_1(X) = [f]$, f is a coboundary, i.e., $f = \delta^0 f_0$ for some $f_0 \in \check{C}^0(\mathcal{U};\mathbb{Z}_2)$. For each index $k, f_0(k) = \pm 1$ so we can select a $g_k \in O(m,n-m)$ with $\det(g_k) = f_0(k)$. Now define new local orthonormal frames $s'_k : U_k \longrightarrow F(X)$ by $s'_k = s_k \cdot g_k$ for $k = 1,2,\ldots$. According to Exercise 6.5.2, the transition functions $g'_{k_0k_1}$ corresponding to the new trivializations are given by

$$g'_{k_0k_1} = g_{k_0}^{-1}\, g_{k_0k_1}\, g_{k_1}$$

so

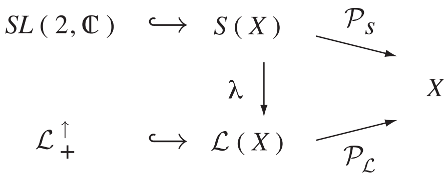
$$\begin{aligned}\det\left(g'_{k_0k_1}\right) &= \det\left(g_{k_0}\right)\,\det\left(g_{k_1}\right)\,\det\left(g_{k_0k_1}\right)\\ &= f_0(k_0)\; f_0(k_1)\; f(k_0,k_1)\\ &= (\delta_0 f_0)\;(k_0,k_1)\; f(k_0,k_1)\\ &= (f(k_0,k_1))^2\\ &= 1.\end{aligned}$$

Since $s'_{k_1} = s'_{k_0} \cdot g'_{k_0k_1}$ for every 1-simplex (k_0, k_1), the orthonormal frames s'_{k_0} and s'_{k_1} therefore determine the same orientation at each $p \in U_{k_0} \cap U_{k_1}$. Consequently, the collection of all $s'_k : U_k \longrightarrow F(X), k = 1, 2, \ldots,$ determines an orientation on X. ∎

We now restrict our attention to an oriented, time oriented spacetime X and seek a Čech cohomology class which represents the obstruction to the existence of a spinor structure for X in the same sense that the 1st Stiefel-Whitney class represents the obstruction to the existence of an orientation for a semi-Riemannian manifold. We recall (Section 2.4) that a **spinor structure** for X consists of a principal $SL(2,\ \ )$-bundle

$$SL(2,\ \ ) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X$$

and a map $\lambda : S(X) \longrightarrow \mathcal{L}(X)$ to the oriented, time oriented, orthonormal frame bundle such that

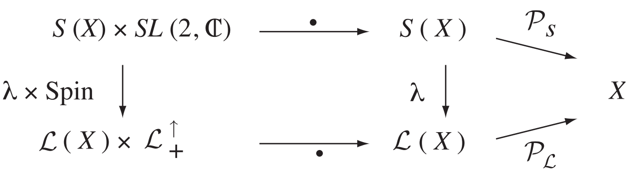


$$\mathcal{P}_{\mathcal{L}} \circ \lambda = \mathcal{P}_S$$

and

$$\lambda(p \cdot g) = \lambda(p) \cdot \mathrm{Spin}\,(g)$$

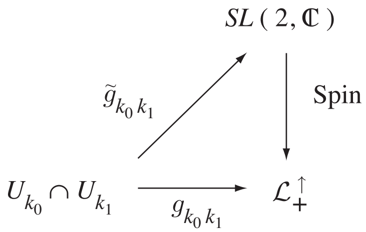
for all $p \in S(X)$ and all $g \in SL(2,\ \ )$:



Let $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ be a locally finite, simple open cover of X with each $\bar{U}_k$ compact and choose local orthonormal frame fields $s_k : U_k \longrightarrow \mathcal{L}(X)$, $k = 1, 2, \ldots$. The corresponding transition functions

$$g_{k_0 k_1} : U_{k_0} \cap \ U_{k_1} \longrightarrow \mathcal{L}_+^{\uparrow}$$

map into $\mathcal{L}_+^{\uparrow}$ for every 1-simplex (k_0, k_1). Because each $U_{k_0} \cap U_{k_1}$ is diffeomorphic to $\ ^4$ and Spin : $SL(2,\ ) \longrightarrow \mathcal{L}_+^{\uparrow}$ is a covering map, these transition functions lift to $SL(2,\ )$:



We can clearly select these lifts in such a way that

$$\tilde{g}_{k_1 k_0} = \tilde{g}_{k_0 k_1}{}^{-1}$$

for all $k_0, k_1 = 1, 2, \ldots$. Furthermore, since Spin is a homomorphism,

$$\text{Spin}\Big(\tilde{g}_{k_0 k_1}\ \tilde{g}_{k_1 k_2}\ \tilde{g}_{k_2 k_0}\Big) = g_{k_0 k_1}\ g_{k_1 k_2}\ g_{k_2 k_0} = id.$$

But the kernel of Spin is $\{\pm id\}$ so

$$\tilde{g}_{k_0 k_1}\ \tilde{g}_{k_1 k_2}\ \tilde{g}_{k_2 k_0} = \pm id$$

at each point in $U_{k_0} \cap U_{k_1} \cap U_{k_2}$. But this intersection is connected (diffeomorphic to $\ ^4$) and so this map is either id everywhere or $-id$ everywhere. Thus, we can define a map Z from 2-simplexes to $\mathbb{Z}_2$ by

$$\tilde{g}_{k_0 k_1}\ \tilde{g}_{k_1 k_2}\ \tilde{g}_{k_2 k_0} = Z(k_0, k_1, k_2)(id).$$

Exercise 6.5.6 Show that Z is a 2-cocycle and therefore determines a cohomology class $[Z] \in \check{H}^2(X; \mathbb{Z}_2)$.

Once again it would appear that the cohomology class $[Z]$ we have defined depends on not only the initial choice of local orthonormal frames s_k, but also on the lifts $\tilde{g}_{k_0 k_1}$ of the corresponding transition functions. We show now that this is not, in fact, the case. Thus, we suppose $s'_k : U_k \longrightarrow \mathcal{L}(X),\ k = 1, 2, \ldots,$ are also cross-sections. Then $s'_k = s_k \cdot g_k$ on U_k for some $g_k : U_k \longrightarrow \mathcal{L}_+^{\uparrow}$. As in Exercise 6.5.2 we let $s_{k_1} = s_{k_0} \cdot g_{k_0 k_1}$ and $s'_{k_1} = s'_{k_0} \cdot g'_{k_0 k_1}$ on $U_{k_0} \cap U_{k_1}$

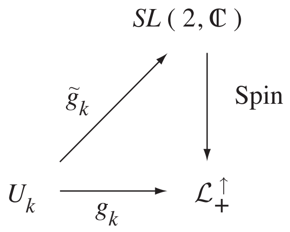
and conclude that

$$\tilde{g}'_{k_0k_1} = g_{k_0}^{-1}\ g_{k_0k_1} g_{k_1}.$$

Let $\tilde{g}_{k_0k_1}$ and $\tilde{g}'_{k_0k_1}$ be lifts and define the 2-cocycles Z and Z' as above. To show that these determine the same cohomology class we must prove that

$$Z' = (\delta^1 h)Z$$

for some 1-cochain h. Lift the maps g_k to $SL(2,\ \ )$.



Now define, for each 1-simplex (k_0, k_1), a map

$$h_{k_1k_0} : U_{k_0} \cap U_{k_1} \longrightarrow SL(2,\ \ )$$

by

$$h_{k_0k_1} = \tilde{g}_{k_0k_1}\ \tilde{g}_{k_1}\ \tilde{g}'_{k_1k_0}\ \tilde{g}_{k_0}^{-1}.$$

Since Spin is a homomorphism, $\mathrm{Spin} \circ h_{k_0k_1}$ takes the value $id \in \mathcal{L}_+^{\uparrow}$ for each x in $U_{k_0} \cap U_{k_1}$. Thus, $h_{k_0k_1}$ maps into $\{\pm id\} \subseteq SL(2,\ \ )$ and connectivity of $U_{k_0} \cap U_{k_1}$ implies that it must be constant. Thus, we may define a 1-cochain h by

$$h(k_0, k) == h_{k_0k_1}(x)$$

where x is any element of $U_{k_0} \cap U_{k_1}$.

Exercise 6.5.7 Show that $Z' = (\delta^1 h)Z$.

Thus, we have a well-defined cohomology class $[Z]$ which we will call the **2nd Stiefel-Whitney class** of X and denote $w_2(X)$. To unearth its revelance to the existence of spinor structures on X we begin with the simple observation that $w_2(X)$ is trivial if and only if there exist lifts $\tilde{g}_{k_0k_1}$ of the transition functions which satisfy

$$\tilde{g}_{k_0k_1}\ \tilde{g}_{k_1k_2}\ \tilde{g}_{k_2k_0} = id \tag{6.5.1}$$

at each point of $U_{k_0} \cap\ U_{k_1} \cap\ U_{k_2}$. Of course, if such lifts exist, $Z(k_0, k_1, k_2) = 1$ so $w_2(X) = [Z]$ is trivial. Suppose, conversely, that $w_2(X)$ is trivial. Select lifts $\tilde{g}_{k_0k_1}$ and define the 2-cochain $Z(k_0, k_1, k_2)$ as above. Since $w_2(X) = [Z]$ is trivial there exists a 1-cochain f such that $Z = \delta^1 f$, i.e.,

$$Z(k_0, k_1, k_2) = f(k_1, k_2)\ f(k_0, k_2)\ f(k_0, k_1)$$

for every 2-simplex (k_0, k_1, k_2). Define

$$\tilde{g}'_{k_0 k_1} : U_{k_0} \cap U_{k_1} \longrightarrow SL(2,\ \ )$$

by

$$\tilde{g}'_{k_0 k_1} = \tilde{g}_{k_0 k_1}\ f(k_0, k_1).$$

Then

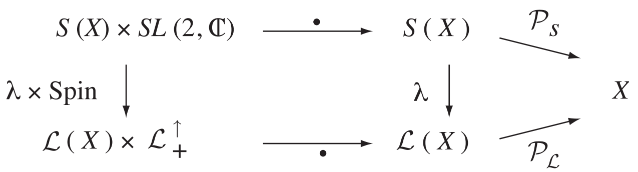
$$\begin{aligned} \tilde{g}'_{k_0 k_1}\ \tilde{g}'_{k_1 k_2}\ \tilde{g}'_{k_2 k_0} &= \tilde{g}_{k_0 k_1}\ f(k_0, k_1)\ \tilde{g}_{k_1 k_2}\ f(k_1, k_2)\ \tilde{g}_{k_2 k_0}\ f(k_2, k_0) \\ &= (Z(k_0, k_1, k_2))^2\ (id) \\ &= id \end{aligned}$$

so the primed lifts satisfy (6.5.1). Now we will prove our major result by showing that the existence of lifts satisfying (6.5.1) is equivalent to the existence of a spinor structure on X.

Theorem 6.5.4 *Let X be an oriented, time oriented spacetime. Then X admits a spinor structure if and only if the 2^{nd} Stiefel-Whitney class $w_2(X) \in \check{H}^2(X; \mathbb{Z}_2)$ of X is trivial.*

Proof: For both parts of the proof we fix at the outset a locally finite, simple open cover $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ for X.

Suppose first that a spinor structure



for X exists. The spinor bundle $SL(2,\ \ ) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X$ is trivial over each U_k so we can choose a cross-section

$$\tilde{s}_k : U_k \longrightarrow \mathcal{P}_S^{-1}(U_k) = \lambda^{-1}\left(\mathcal{P}_{\mathcal{L}}^{-1}(U_k)\right)$$

for each $k = 1, 2, \ldots$. Define

$$s_k : U_k \longrightarrow \mathcal{P}_S^{-1}(U_k)$$

by

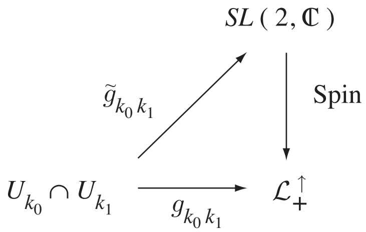
$$s_k = \lambda \circ \tilde{s}_k.$$

Then each s_k is a cross-section of the frame bundle $\mathcal{L}_+^\uparrow \hookrightarrow \mathcal{L}(X) \xrightarrow{\mathcal{P}_\mathcal{L}} X$ because $\mathcal{P}_\mathcal{L} \circ \lambda = \mathcal{P}_s$. These cross-sections give rise to trivializations and therefore transition functions and we denote the transition functions by $\tilde{g}_{k_0k_1}$ and $g_{k_0k_1}$, respectively.

Exercise 6.5.8 Show that, on $U_{k_0} \cap U_{k_1}$, $\mathrm{Spin} \circ \tilde{g}_{k_0k_1} = g_{k_0k_1}$.

Thus, the $\tilde{g}_{k_0k_1}$ are lifts of the $g_{k_0k_1}$. Since they are transition functions of a bundle, they satisfy the cocycle condition (6.5.1) and so we conclude that $w_2(X)$ is trivial.

For the converse we begin with local orthonormal frame fields $s_k : U_k \longrightarrow \mathcal{P}_\mathcal{L}^{-1}(U_k),\ k = 1, 2, \ldots$ and denote by $g_{k_0k_1} \colon U_{k_0} \cap U_{k_1} \longrightarrow \mathcal{L}_+^\uparrow$ the corresponding transition function for each 1-simplex (k_0, k_1). Our assumption now is that $w_2(X)$ is trivial and we have seen that it follows from this that there exist lifts



satisfying $\tilde{g}_{k_1k_0} = \tilde{g}_{k_0k_1}^{-1}$ and (6.5.1). The Reconstruction Theorem therefore guarantees the existence of a unique (up to equivalence) $SL(2,\ \ )$-bundle

$$SL(2,\ \ ) \hookrightarrow S(X) \xrightarrow{\mathcal{P}_S} X$$

with the $\tilde{g}_{k_0k_1}$ as transition functions. All that remains is to define a map

$$\lambda : S(X) \longrightarrow \mathcal{L}(X)$$

satisfying $P_\mathcal{L} \circ \lambda = \mathcal{P}_S$ and $\lambda(p \cdot g) = \lambda(p) \cdot \mathrm{Spin}\,(g)$. For this we will define λ above each U_k and then show that the definitions agree on any nonempty intersections.

The trivialization $\Psi_k : \mathcal{P}_\mathcal{L}^{-1}(U_k) \longrightarrow U_k \times \mathcal{L}_+^\uparrow$ determined by s_k is given by

$$\Psi_k(s_k(x) \cdot g) = (x, g)$$

for all $x \in U_k$ and $g \in \mathcal{L}_+^\uparrow$. Let $\tilde{\Psi}_k : \mathcal{P}_S^{-1}(U_k) \longrightarrow U_k \times SL(2,\ \ )$, $k = 1, 2, \ldots$, be trivializations of $S(X)$ related by the transition functions $\tilde{g}_{k_0k_1}$ and let $\tilde{s}_k : U_k \longrightarrow \mathcal{P}_S^{-1}(U_k),\ k = 1, 2, \ldots$, be the corresponding canonical cross-sections. Define

$$\lambda_k : \mathcal{P}_S^{-1}(U_k) \longrightarrow \mathcal{P}_\mathcal{L}^{-1}(U_k)$$

by

$$\lambda_k = \Psi_k^{-1} \circ (id_{U_k} \times \text{Spin}) \circ \tilde{\Psi}_k$$

for each $k = 1, 2, \ldots$.

Exercise 6.5.9 Show that $\lambda_k(\tilde{s}_k(x) \cdot g) = s_k(x) \cdot \text{Spin}(g)$ for all $x \in U_k$ and all $g \in SL(2, \ )$ and conclude that, on $\mathcal{P}_S^{-1}(U_k)$,

$$\mathcal{P}_{\mathcal{L}} \circ \lambda_k = \mathcal{P}_S$$

and

$$\lambda_k(p \cdot g) = \lambda_k(p) \cdot \text{Spin}\,(g).$$

To show that the maps λ_k determine a map $\lambda : S(X) \longrightarrow \mathcal{L}(X)$ satisfying $\mathcal{P}_{\mathcal{L}} \circ \lambda = \mathcal{P}_S$ and $\lambda(p \cdot g) = \lambda(p) \cdot \text{Spin}(g)$ it will be enough to show that they agree on any nonempty intersection

$$\mathcal{P}_S^{-1}\left(U_{k_0}\right) \cap \mathcal{P}_S^{-1}\left(U_{k_1}\right) = \mathcal{P}_S^{-1}\left(U_{k_0} \cap U_{k_1}\right).$$

But, for $x \in U_{k_0} \cap U_{k_1}$, we have

$$\lambda_{k_1}\left(\tilde{s}_{k_1}(x) \cdot g\right) = s_{k_1}(x) \cdot \text{Spin}(g)$$

$$\tilde{s}_{k_1}(x) = \tilde{s}_{k_0}(x) \cdot \tilde{g}_{k_0 k_1}(x)$$

and

$$s_{k_1}(x) = s_{k_0}(x) \cdot g_{k_0 k_1}(x)$$

so

$$\begin{aligned}
\lambda_{k_0}\left(\tilde{s}_{k_1}(x) \cdot g\right) &= \lambda_{k_0}\left(\tilde{s}_{k_0}(x) \cdot \left(\tilde{g}_{k_0 k_1}(x) g\right)\right) \\
&= s_{k_0}(x) \cdot \text{Spin}\left(\tilde{g}_{k_0 k_1}(x) g\right) \\
&= \left(s_{k_0}(x) \cdot \text{Spin}\left(\tilde{g}_{k_0 k_1}(x)\right)\right) \cdot \text{Spin}(g) \\
&= \left(s_{k_0}(x) \cdot g_{k_0 k_1}(x)\right) \cdot \text{Spin}(g) \\
&= s_{k_1}(x) \cdot \text{Spin}(g) \\
&= \lambda_{k_1}\left(\tilde{s}_{k_1}(x) \cdot g\right)
\end{aligned}$$

as required. ■

Theorem 6.5.4 has some rather obvious consequences that are of sufficient importance to be recorded officially. If X happens to be diffeomorphic to ^{4}, then $\check{H}^2(X;\mathbb{Z}_2)$ is trivial so $w_2(X)$ is trivial and X admits a spinor structure.

Corollary 6.5.5 *A time oriented spacetime diffeomorphic to* ^{4} *admits a spinor structure.*

This applies, in particular, to Minkowski spacetime and the Einstein-deSitter spacetime. The deSitter spacetime and the Einstein cylinder are diffeomorphic, not to ^{4}, but to $S^3 \times$. That they nevertheless do admit spinor structures will follow from our next result.

Corollary 6.5.6 *Let X be an oriented, time oriented spacetime and suppose that the oriented, time oriented orthonormal frame bundle* $\mathcal{L}_+^\uparrow \hookrightarrow \mathcal{L}(X) \xrightarrow{\mathcal{P}_\mathcal{L}} X$ *is trivial. Then X admits a spinor structure.*

Proof: By assumption, there exists a global cross-section $s : X \longrightarrow \mathcal{L}(X)$. Letting $\mathcal{U} = \{U_k\}_{k=1,2,\ldots}$ be a locally finite simple open cover we define $s_k : U_k \longrightarrow \mathcal{P}_\mathcal{L}^{-1}(U_k)$ by $s_k = s|U_k$. Since $s_{k_0} = s_{k_1} \cdot g_{k_0k_1}$ on $U_{k_0} \cap U_{k_1}$, all of the corresponding transition functions $g_{k_0k_1}$ are identically equal to $id \in \mathcal{L}_+^\uparrow$. These surely lift to maps $\tilde{g}_{k_0k_1} : U_{k_0} \cap U_{k_1} \longrightarrow SL(2, \;)$ that take the value $id \in SL(2, \;)$ everywhere and these satisfy (6.5.1) so $w_2(X)$ is trivial and X admits a spinor structure. ∎

Remark: Geroch [**G1**] has shown that the triviality of $\mathcal{L}_+^\uparrow \hookrightarrow \mathcal{L}(X) \xrightarrow{\mathcal{P}_\mathcal{L}} X$ actually characterizes the existence of spinor structures for oriented, time oriented (noncompact) spacetimes.

Since we have shown in Chapter 3 that both deSitter spacetime and the Einstein cylinder satisfy the hypotheses of Corollary 6.5.6., they too admit spinor structures.

Appendix

Seiberg-Witten Gauge Theory

Spin $\frac{1}{2}$-electrodynamics is a gauge theory of a $U(1)$-gauge field coupled to a spinor field and, although the connection is defined on a different $U(1)$-bundle and the spinor field is of a different type, so is Seiberg-Witten theory. The significance of the latter, however, resides in an entirely different arena and one cannot truly appreciate this significance without placing the theory in its historical context. This we will attempt to do in Section A.l. The (rather substantial) algebraic preliminaries required just to describe the basic elements of the theory are introduced in Section A.2 and then we move on to the field equations and their moduli space of solutions. Much of what we have to say lies in considerably deeper waters than the main body of the text and so we must content ourselves with something akin to an initial geographical survey of the terrain with signposts to the literature for those who wish to dive in. What we offer is really a continuation of the story begun in Appendix B of [**N4**] and we will need to assume that our reader is aquainted with this.

A.1 Donaldson Invariants and TQFT

The first incursion of gauge theory into topology was Donaldson's theorem on compact, simply connected, smooth 4-manifolds M with $b_2^+(M) = 0$ and definite intersection form. The proof is based on an analysis of the moduli space of anti-self-dual connections on the $SU(2)$-bundle $SU(2) \hookrightarrow P_1 \xrightarrow{\mathcal{P}_1} M$ over M with Chern number 1. The structure of this moduli space and Donaldson's proof of his theorem are sketched (ever so briefly) in Appendix B of [**N4**]. Much of the analysis described there can be carried out for the bundles $SU(2) \hookrightarrow P_k \xrightarrow{\mathcal{P}_k} M$ with positive Chern number $k > 0$ and without the assumption that $b_2^+(M) = 0$, although the results are rather different. We will not repeat this analysis here, but will simply ask that our reader become aquainted with Appendix B of [**N4**] and state the end result.

Throughout this discussion M will denote a compact, simply connected, oriented, smooth 4-manifold with $b_2^+(M) > 0$ (soon we will impose additional conditions on $b_2^+(M)$ and explain why). For each $k > 0$ we let $SU(2) \hookrightarrow P_k \xrightarrow{\mathcal{P}_k} M$ be the principal $SU(2)$-bundle over M with Chern number k, $\mathcal{A}(P_k)$ the (affine) space of all connection 1-forms on P_k and $\mathcal{G}(P_k)$ the gauge group of automorphisms of P_k. As usual, $\mathcal{G}(P_k)$ acts on $\mathcal{A}(P_k)$ by pullback and two connections $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \mathcal{A}(P_k)$ are gauge equivalent if there is an $f \in \mathcal{G}(P_k)$ such that $\boldsymbol{\omega}' = f^*\boldsymbol{\omega}$. The moduli space of all gauge equivalence classes $[\boldsymbol{\omega}]$ is

G.L. Naber, *Topology, Geometry and Gauge fields: Interactions*,
Applied Mathematical Sciences 141, DOI 10.1007/978-1-4419-7895-0,

denoted $\mathcal{B}(P_k) = \mathcal{A}(P_k)/\mathcal{G}(P_k)$. The stabilizer of $\boldsymbol{\omega}$ is the subgroup of $\mathcal{G}(P_k)$ consisting of those f that leave $\boldsymbol{\omega}$ fixed and always contains a copy of $\mathbb{Z}_2$. If the stabilizer is precisely $\mathbb{Z}_2$, then $\boldsymbol{\omega}$ is said to be irreducible (otherwise, it is reducible). The (open) subset of $\mathcal{A}(P_k)$ consisting of irreducible connections is denoted $\hat{\mathcal{A}}(P_k)$ and its moduli space is $\hat{\mathcal{B}}(P_k) = \hat{\mathcal{A}}(P_k)/\mathcal{G}(P_k)$.

Now we let $\boldsymbol{g}$ be a Riemannian metric on M and consider the set $\mathrm{Asd}(P_k, \boldsymbol{g})$ of all $\boldsymbol{\omega} \in \mathcal{A}(P_k)$ that are $\boldsymbol{g}$-anti-self-dual, i.e., satisfy ${}^*\boldsymbol{F}_{\boldsymbol{\omega}} = -\boldsymbol{F}_{\boldsymbol{\omega}}$, where $*$ is the corresponding Hodge star operator and $\boldsymbol{F}_{\boldsymbol{\omega}}$ is the curvature of $\boldsymbol{\omega}$ thought of as an ad P_k-valued 2-form on M. This might well be empty (when $k = 1$ and $b_2^+(M) = 0$ a theorem of Taubes prohibits this, but we are assuming $b_2^+(M) > 0$). The moduli space

$$\mathcal{M}(P_k, \boldsymbol{g}) = \mathrm{Asd}(P_k, \boldsymbol{g})/\mathcal{G}(P_k)$$

of $\boldsymbol{g}$-anti-self-dual connections on P_k could therefore also be empty. The assumption that $b_2^+(M)$ is positive has beneficial consequences as well, however. Under this assumption Donaldson proved a generic metrics theorem to the effect that there is a dense subset of the space $\mathcal{R}$ of Riemannian metrics on M with the property that for any $\boldsymbol{g}$ in this set every $\boldsymbol{g}$-anti-self-dual connection $\boldsymbol{\omega}$ on P_k is irreducible and the moduli space $\mathcal{M}(P_k, \boldsymbol{g})$ is either empty or a smooth orientable manifold of dimension

$$8k - 3\left(1 + b_2^+(M)\right)$$

(moreover, an orientation for $\mathcal{M}(P_k, \boldsymbol{g})$ is canonically determined by choosing an orientation for the vector space $H_+^2(M;\quad)$).

Remark: The restriction on $b_2^+(M)$ arises because the subset of $\mathcal{R}$ consisting of those $\boldsymbol{g}$ for which reducible $\boldsymbol{g}$-ASD connections exist is a countable union of smooth submanifolds of codimension $b_2^+(M)$. If $b_2^+(M) = 0$, then reducibles are generically unavoidable (and we saw in Appendix B of **[N4]** that this is a good, not a bad thing as it leads to the cone singularities required for the cobordism proof of Donaldson's Theorem).

With this information in hand Donaldson set about constructing a sequence $\gamma_d(M), d = 0, 1, 2, \ldots$, of differential topological invariants of M (this means that the invariants agree for two smooth 4-manifolds that are diffeomorphic, but need not agree if the manifolds are merely homeomorphic, but not diffeomorphic). The first $\gamma_0(M)$ is simply an integer, but for $d \geq 1$, $\gamma_d(M) : H_2(M) \longrightarrow \mathbb{Z}[\frac{1}{2}]$ is a function on the second homology of M with values that are integer multiples of $\frac{1}{2}$. They are called the **Donaldson invariants** of M. Their definition and, especially, their computation is enormously complex, but these computations produced remarkable insights into the structure of smooth 4-manifolds. Eventually they were superceded by invariants arising from Seiberg-Witten gauge theory, but to understand how this came about one must have at least some general, intuitive sense of how they are defined (those in search of honest definitions should proceed first to **[M2]** and then move on to [**DK**]). In the brief description that follows we assume that we have selected a generic Riemannian metric $\boldsymbol{g}$ and fixed some orientation of $H_2^+(M; \quad)$.

Notice that, by an appropriate arrangement of $b_2^+(M)$ and the Chern number k, it is entirely possible for the dimension $8k-3(1+b_2^+(M))$ of $\mathcal{M}(P_k, \boldsymbol{g})$ to come out zero. Then $\mathcal{M}(P_k, \boldsymbol{g})$ is either empty or an oriented, 0-dimensional manifold, i.e., it is a set of isolated points $[\boldsymbol{\omega}]$ each equipped with a sign ± 1 which we will denote $(-1)^{[\boldsymbol{\omega}]}$. As it happens, in this case (and *only* in this case) $\mathcal{M}(P_k, \boldsymbol{g})$ is compact, i.e., finite, so we can define

$$\gamma_0(M) = \begin{cases} \qquad 0 \qquad\qquad , & \mathcal{M}(P_k, \boldsymbol{g}) = \emptyset \\ \displaystyle\sum_{[\boldsymbol{\omega}] \in \mathcal{M}(P_k, \boldsymbol{g})} (-1)^{[\boldsymbol{\omega}]}, & \mathcal{M}(P_k, \boldsymbol{g}) \neq \emptyset \end{cases} .$$

Regrettably, this integer need not be an invariant and may, in fact, depend on the choice of generic metric $\boldsymbol{g}$. To prevent this one must ensure that not only is it possible to choose a metric for which there are no reducible ASD connections, but also that a generic variation of that metric (generic path in the space $\mathcal{R}$ of Riemannian metrics on M) does not introduce reducibles. For this one needs the set of metrics in $\mathcal{R}$ for which there are reducibles to be sufficiently "thin" and this, by the Remark above, means that $b_2^+(M)$ must be sufficiently large. One can show that

$$b_2^+(M) > 1$$

will do and so we henceforth make this assumption about M (one more assumption on $b_2^+(M)$ is forthcoming). With this we have the **0-dimensional Donaldson invariant** $\gamma_0(M) \in \mathbb{Z}$ of M.

The definition of $\gamma_d(M)$ for $d > 0$ is a very great deal more complicated. Even the briefest sketch would require substantial effort and is not really necessary for what we have in mind here. For our purposes it will suffice to have a naive, intuitive picture of the idea that lies behind the definition and some appreciation of what makes it so naive. Very roughly, the idea is that the values of $\gamma_d(M)$ will be obtained by integrating over the (finite-dimensional) moduli space certain carefully selected differential forms. To implement this idea Donaldson defined a map

$$\mu : H_2(M) \longrightarrow H^2\left(\mathcal{M}(P_k, \boldsymbol{g})\right)$$

(think of this as assigning 2-forms on $\mathcal{M}(P_k, \boldsymbol{g})$ to surfaces in M) and wedged enough of the images together to get a form with rank $\dim \mathcal{M}(P_k, \boldsymbol{g})$ that can then be integrated over $\mathcal{M}(P_k, \boldsymbol{g})$. Of course, any wedge product of 2-forms has even rank so this cannot succeed unless $\dim \mathcal{M}(P_k, \boldsymbol{g}) = 8k-3(1+b_2^+(M))$ is even, i.e., unless $b_2^+(M)$ is odd. Thus, we arrive at our final restriction on

$b_2^+(M)$. Henceforth, we assume

$$b_2^+(M) \equiv 1 \mod 2 \text{ and } b_2^+(M) > 1.$$

Remark: The definition of Donaldson's μ-map is technical and involves a number of ideas that we have not encountered (Pontryagin characteristic classes and an operation from algebraic topology called the slant product). The details can be found in [**M2**] and [**DK**]. In the special case in which k is odd a rough idea of what is behind the definition can be described as follows: Donaldson constructs a certain auxiliary $SU(2)$-bundle $SU(2) \hookrightarrow P \longrightarrow M \times \mathcal{M}(P_k, \boldsymbol{g})$ over the product $M \times \mathcal{M}(P_k, \boldsymbol{g})$. Its second Chern class $c_2(P)$ is an element of $H^4(M \times \mathcal{M}(P_k, \boldsymbol{g}))$. There is a general result on the cohomology of products (called the Künneth formula) that gives $H^4(M \times \mathcal{M}(P_k, \boldsymbol{g})) \cong \bigoplus_{i+j=4} H^i(M) \otimes H^j(\mathcal{M}(P_k, \boldsymbol{g}))$ so $c_2(P)$ has a (2, 2)-part $(\alpha, \beta) \in H^2(M) \otimes H^2(\mathcal{M}(P_k, \boldsymbol{g}))$ Then, if x is an element of $H_2(M)$ and we identify x with a smoothly embedded surface in M representing it, we can define

$$\mu(x) = \left(\int_x \alpha\right)\beta \in H^2(\mathcal{M}(P_k, \boldsymbol{g})).$$

Now we are prepared to describe our "naive" definition of $\gamma_d(M)$ when $8k - 3(1 + b_2^+(M)) > 0$. First we fix the Chern number k. Since $b_2^+(M)$ is odd we can write

$$8k - 3\left(1 + b_2^+(M)\right) = 2d_k$$

for some positive integer d_k. For any $x \in H_2(M), \mu(x)$ is in $H^2(\mathcal{M}(P_k, \boldsymbol{g}))$ so $\mu(x) \wedge \overset{d_k}{\cdots} \wedge \mu(x) \in H^{2d_k}(\mathcal{M}(P_k, \boldsymbol{g}))$ and we set

$$\gamma_{d_k}(M)(x) = \int_{\mathcal{M}(P_k, \boldsymbol{g})} \mu(x) \wedge \overset{d_k}{\cdots} \wedge \mu(x).$$

Now for any positive integer d with

$$d \equiv -\frac{3}{2}\left(1 + b_2^+(M)\right) \mod 4$$

we can select a k for which $2d = 8k - 3(1 + b_2^+(M))$ so that $d = d_k$ and we have "defined" $\gamma_d(M)$.

Remarks: There are probably no more than a thousand things wrong with this "definition" and setting it all straight requires a huge amount of technical labor. Nevertheless, morally at least it represents the correct idea and is, in fact, the way these invariants are often viewed by physicists (more on this shortly). We will offer just a brief tour of what is wrong and what it takes to make it right. Most of the fuss can be traced to the fact that, when its dimension is positive, $\mathcal{M}(P_k, \boldsymbol{g})$ is never compact and so one can generally not integrate over it. Deep analytical results of Uhlenbeck and Taubes and much labor led Donaldson to a compactification $\bar{\mathcal{M}}(P_k, \boldsymbol{g})$, called the **Uhlenbeck compactification**, of $\mathcal{M}(P_k, \boldsymbol{g})$ and an extension $\bar{\mu} : H_2(M) \longrightarrow H^2(\bar{\mathcal{M}}(P_k, \boldsymbol{g}))$ of the μ-map to it. Unfortunately, although it is compact, $\bar{\mathcal{M}}(P_k, \boldsymbol{g})$ is not a manifold and so integration no longer makes sense. Algebraic topology presents the alternative of pairing the cohomology class $\bar{\mu}(x) \wedge \cdots \wedge \bar{\mu}(x)$ with what is called the fundamental homology class $[\bar{\mathcal{M}}(P_k, \boldsymbol{g})]$ of $\bar{\mathcal{M}}(P_k, \boldsymbol{g})$. Alas, $\mathcal{M}(P_k, \boldsymbol{g})$ has a fundamental class only for sufficiently large k (the so-called "stable range" $k > \frac{3}{4}(1+b_2^+(M))$ or, equivalently, $d_k > \frac{3}{2}(1+b_2^+(M))$. For $d \equiv -\frac{3}{2}(1+b_2^+(M)) \bmod 4$ and $d > \frac{3}{2}(1+b_2^+(M))$ one can define $\gamma_d(M)$ in the manner just described. Removing the stable range condition and the required mod 4 congruence requires another μ-map

$$\mu : H_0(M) \longrightarrow H^4\Big(\mathcal{M}(P_k, \boldsymbol{g})\Big),$$

a detour around the fact that this one does not extend fully to $\bar{\mathcal{M}}(P_k, \boldsymbol{g})$ and a blow-up formula relating $\gamma_d(M)$ to $\gamma_d(M \# \overline{\mathbb{CP}}^2)$, where $M \# \overline{\mathbb{CP}}^2$ is the connected sum of M and $\overline{\mathbb{CP}}^2$ (which is obtained by deleting an open 4-ball from both M and $\overline{\mathbb{CP}}^2$ and identifying the boundary spheres). Incidentally, it is only in this last step that the values of $\gamma_d(M)$ become multiples of $\frac{1}{2}$ rather than just integers. Those inclined to find out what all of this really means are referred to [**M2**] and [**DK**].

It is certainly not our intention here to compute Donaldson invariants or use them to obtain topological information about 4-manifolds. Rather we would like to sketch how, by adopting a slightly different perspective, they lead, by way of quantum field theory, to the Seiberg-Witten theory that is our real concern here. To keep the discussion as uncluttered as possible we will focus most of our attention on $\gamma_0(M)$ and will economize on notation by writing P for $P_k, \mathcal{G}$ for $\mathcal{G}(P_k), \mathcal{A}$ for $\mathcal{A}(P_k)$, etc.

The gauge group $\mathcal{G}$ does not act freely on the space $\hat{\mathcal{A}}$ of irreducible connections since even irreducibles have a $\mathbb{Z}_2$ stabilizer. However, $\hat{\mathcal{G}} = \mathcal{G}/\mathbb{Z}_2$ does act freely on $\hat{\mathcal{A}}$ so we have an infinite-dimensional principal bundle

$$\hat{\mathcal{G}} \hookrightarrow \hat{\mathcal{A}} \longrightarrow \hat{\mathcal{B}}.$$

We build a vector bundle associated to this principal bundle as follows: We claim that there is a smooth left action of $\hat{\mathcal{G}}$ on the (infinite-dimensional) vector space $\Lambda^2_+(M, \operatorname{ad} P)$ of self-dual 2-forms on M with values in the adjoint bundle $\operatorname{ad} P$. To see this we think of $\mathcal{G}$ as the group of sections of the nonlinear adjoint bundle $\operatorname{Ad} P$ under pointwise multiplication. Since the elements of $\Lambda^2_+(M, \operatorname{ad} P)$ take values in the $su(2)$-fibers of $\operatorname{ad} P$, $\mathcal{G}$ acts on these fibers by conjugation. Moreover, conjugation takes the same value at $\pm f \in \mathcal{G}$ so this $\mathcal{G}$-action on $\Lambda^2_+(M, \operatorname{ad} P)$ descends to a $\mathcal{G}/\mathbb{Z}_2 = \hat{\mathcal{G}}$-action. Thus, we have an associated vector bundle

$$\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Lambda^2_+(M, \operatorname{ad} P),$$

the elements of which are equivalence classes $[\boldsymbol{\omega},\ \alpha] = [\boldsymbol{\omega} \cdot f,\ f^{-1} \cdot \alpha\ ]$ with $\boldsymbol{\omega} \in \hat{\mathcal{A}},\ \alpha \in \Lambda^2_+(M,\ \mathrm{ad}\, P)$ and $f \in \hat{\mathcal{G}}$.

Now recall that sections of associated vector bundles can be identified with equivariant maps from the principal bundle space to the vector space fiber. In our case we have an obvious map from $\hat{\mathcal{A}}$ to $\Lambda^2_+(M,\ \mathrm{ad}\, P)$, namely, the self-dual curvature map

$$\boldsymbol{F}^+ : \hat{\mathcal{A}} \longrightarrow \Lambda^2_+(M, \mathrm{ad}\, P)$$
$$\boldsymbol{F}^+(\boldsymbol{\omega}) = \boldsymbol{F}^+_{\boldsymbol{\omega}} = \tfrac{1}{2}\left(\boldsymbol{F}_{\boldsymbol{\omega}} + {}^*\boldsymbol{F}_{\boldsymbol{\omega}}\right).$$

Since the $\hat{\mathcal{G}}$-action on $\hat{\mathcal{A}}$ is by conjugation and curvature transforms by conjugation under a gauge transformation, $\boldsymbol{F}^+$ is equivariant:

$$\boldsymbol{F}^+(\boldsymbol{\omega} \cdot f) = \boldsymbol{F}^+_{\boldsymbol{\omega}\cdot f} = f^{-1}\boldsymbol{F}^+_{\boldsymbol{\omega}} f = f^{-1} \cdot \boldsymbol{F}^+_{\boldsymbol{\omega}} = f^{-1} \cdot \boldsymbol{F}^+(\boldsymbol{\omega}).$$

$\boldsymbol{F}^+$ can therefore be identified with the section

$$s_+ : \hat{\mathcal{B}} \longrightarrow \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Lambda^2_+(M, \mathrm{ad}\, P)$$
$$s_+([\boldsymbol{\omega}]) = [\boldsymbol{\omega},\ \boldsymbol{F}^+_{\boldsymbol{\omega}}]$$

of our vector bundle. Now notice that the moduli space $\mathcal{M}$ of anti-self-dual connections ($\boldsymbol{F}^+_{\boldsymbol{\omega}} = 0$) is precisely the zero set of the section s_+ (for generic $\boldsymbol{g}$ all such connections are irreducible). It is a general fact that the base of any smooth vector bundle is diffeomorphic to the image of any cross-section of the vector bundle (in particular, the one that picks out the zero element in each fiber). Thus, we can identify $\hat{\mathcal{B}}$ with the image of the zero section

$$s_0 : \hat{\mathcal{B}} \longrightarrow \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Lambda^2_+(M, \mathrm{ad}\, P)$$
$$s_0([\boldsymbol{\omega}]) = [\boldsymbol{\omega},\ 0].$$

We conclude that

$$\mathcal{M} = s_+\left(\hat{\mathcal{B}}\right) \cap s_0\left(\hat{\mathcal{B}}\right).$$

In the case in which $\mathcal{M}$ is 0-dimensional each point in this intersection acquires a sign ± 1 and the Donaldson invariant $\gamma_0(M)$ is the sum of these signs.

This last interpretation of $\gamma_0(M)$ is reminiscent of a very famous result in topology that we now digress momentarily to describe. We have seen (Section 3.3) that a smooth vector field $\boldsymbol{V}$ on a manifold X can be identified with a cross-section $\boldsymbol{V} : X \longrightarrow TX$ of the tangent bundle of X. The image $\boldsymbol{V}(X)$ of X is a submanifold of TX diffeomorphic to X and the same is true of the zero cross-section (vector field) $\boldsymbol{V}_0 : X \longrightarrow TX$. We will identify X with $\boldsymbol{V}_0(X)$. Then the zeros (singularities) of $\boldsymbol{V}$ are just the points of $\boldsymbol{V}(X) \cap \boldsymbol{V}_0(X)$. Now assume that X is compact and oriented. Then TX is also oriented and we can give $\boldsymbol{V}(X)$ the orientation that makes $\boldsymbol{V}$ orientation preserving. Now we proceed exactly as in the definition of the intersection form (Appendix B of [**N4**]) to define the index $\mathrm{Ind}(\boldsymbol{V})$ of $\boldsymbol{V}$: Perturb $\boldsymbol{V}$ slightly so that $\boldsymbol{V}(X)$ and $\boldsymbol{V}_0(X)$ intersect transversally and therefore (since $\dim TX = 2\dim X$) in a finite set of isolated points. Such a point p is assigned the value 1 if an oriented basis for $T_p(\boldsymbol{V}_0(TX))$ followed by an oriented basis for $T_p(\boldsymbol{V}(TX))$ gives an oriented basis for $T_p(TX)$; otherwise, it is assigned the value-1. Then $\mathrm{Ind}(\boldsymbol{V})$ is the sum of these values over all points in $\boldsymbol{V}_0(X) \cap \boldsymbol{V}(X)$. The **Poincaré-Hopf Theorem** then asserts that $\mathrm{Ind}(\boldsymbol{V})$ is equal to the Euler characteristic $\chi(X)$ of X (a proof of this can be found in [**MT**]). The point is that we have a topological invariant of X that is determined by cross-sections of a vector bundle over X and in much the same way (sum of signed points in the zero set of the section) that $\gamma_0(M)$ is determined by the section s_+ of $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Lambda^2_+(M, \mathrm{ad}\, P)$. Notice also that one can generalize the definition of $\mathrm{Ind}(\boldsymbol{V})$ to define the index of any section s of any vector bundle $E \longrightarrow X$ that shares a few properties of the tangent bundle (it should be orientable in the sense that orientations can be smoothly supplied to the fibers and the fiber dimension should be the same as $\dim X$).

To understand how physics enters these considerations (and we must understand this because that's where Seiberg-Witten came from) we will take a moment to describe another very famous result of topology/geometry that provides yet another way to calculate $\chi(X)$, this time by integration of a certain characteristic class. Since it turns out that $\chi(X)$ is always zero when $\dim X$ is odd we will restrict ourselves to (compact, oriented) manifolds X of dimension $2k$ for some $k \geq 1$. The result we need is the **Gauss-Bonnet-Chern Theorem** which asserts that

$$\chi(X) = \int_X e(X),$$

where $e(X)$ is called the **Euler class** of X. This is a characteristic class that can be defined by the Chern-Weil procedure described in Section 6.3. Briefly, the construction goes like this: Choose a Riemannian metric on X and consider the corresponding oriented, orthonormal frame bundle

$$SO(2k) \hookrightarrow F_+(X) \longrightarrow X$$

(Exercise 3.3.17). Choose a connection $\boldsymbol{\omega}$ on this bundle with curvature $\boldsymbol{\Omega}$. To apply the Chern-Weil procedure we now need a symmetric polynomial on the Lie algebra $so(2k)$ of $SO(2k)$ that is invariant under the adjoint action. The appropriate choice this time is called the **Pfaffian**

$$Pf : \ so(2k) \longrightarrow \quad .$$

Although there are much more elegant ways to define this (see [**MS**]) we will opt for a simple-minded formula in terms of the matrix entries. Let $A = (A_{ij})$

be an element of $so(2k)$ and define

$$Pf(A) = \frac{1}{2^k k!} \sum_{\sigma \in S_{2k}} (-1)^\sigma A_{\sigma(1)\sigma(2)} \cdots A_{\sigma(2k-1)\sigma(2k)},$$

where S_{2k} is the group of permutations of $\{1, \ldots 2k\}$ and $(-1)^\sigma$ denotes the sign of the permutation σ. One can check, for example, that if

$$A = \left\{ \begin{array}{cccc} 0 & \lambda_1 & & \\ -\lambda_1 & 0 & & 0 \\ & & \ddots & \\ & 0 & 0 & \lambda_k \\ & & -\lambda_k & 0 \end{array} \right\}$$

then $Pf(A) = \lambda_1 \cdots \lambda_k$ (in fact, it is always the case that $(Pf(A))^2 = \det A$). One can show that Pf is invariant under the adjoint action of $SO(2k)$ on $so(2k)$ so Chern-Weil guarantees that if we write $\mathbf{\Omega} = (\mathbf{\Omega}_{ij})$ as a matrix of 1-forms (and normalize by $-\frac{1}{2\pi}$), then

$$Pf\left(-\frac{1}{2\pi}\mathbf{\Omega}\right) = \frac{(-1)^k}{2^{2k}\pi^k k!} \sum_{\sigma \in S_{2k}} (-1)^\sigma \mathbf{\Omega}_{\sigma(1)\sigma(2)} \wedge \cdots \wedge \mathbf{\Omega}_{\sigma(2k-1)\sigma(2k)}$$

descends to a closed $2k$-form on X whose cohomology class $e(X)$ does not depend on the choice of the connection $\mathbf{\Omega}$. This is the Euler class and its integral over X is the Euler characteristic (this is proved in [**MT**]).

All of this can be generalized to an arbitrary oriented, real vector bundle $E \longrightarrow X$ of fiber dimension $2k$ over a compact, oriented manifold of dimension $2k$. Here one chooses a fiber metric (smoothly varying positive definite inner products on the fibers of E) to get an oriented, orthonormal frame bundle $SO(2k) \hookrightarrow F_+(E) \longrightarrow X$. The Euler class $e(E)$ is then defined just as above and one defines the Euler number $\chi(E)$ of the bundle to be its integral over X (this is no longer the Euler characteristic of X, of course). One can then prove an analogue of the Poincaré-Hopf Theorem that gives $\chi(E)$ as the intersection number for a cross-section of E (see [**MT**]).

What we have seen is that the Donaldson invariant $\gamma_0(M)$ is analogous to the Poincaré-Hopf version of an Euler number (although the vector bundle is infinite-dimensional and the Poincaré-Hopf Theorem itself is valid only in finite dimensions). If one were to take this analogy seriously it might suggest the possibility of an integral representation of $\gamma_0(M)$ analogous to the Gauss-Bonnet-Chern Theorem. Notice, however, that such an integral would be over the base of the vector bundle which is the infinite-dimensional moduli space $\hat{\mathcal{B}}$. Now, integrals over infinite-dimensional manifolds are notoriously difficult to make rigorous mathematical sense of, but, fortunately, this does not bother the physicists at all. As Nigel Hitchin has said, "This is such stuff as quantum field theory is made of." Indeed, it was Edward Witten [**W2**] who first produced a (formal) integral representation of $\gamma_0(M)$, not directly, but as what is called the partition function of a certain variant of supersymmetric quantum Yang-Mills theory. Indeed, this quantum field theory also yielded formal integral representations for *all* of the Donaldson invariants and eventually led to the Seiberg-Witten invariants. Athough we are not so presumptuous as to attempt any sort of exegesis of Witten's work a quick tour of a few of the ideas is the only way to see the emergence of Seiberg-Witten.

There are various approaches to the construction of a quantum field theory, but the only one that will concern us here is the so-called Feynmann path integral approach. Here one begins with a classical field theory of just the sort we discussed in Chapter 2. Thus, one is given a collection $\boldsymbol{\xi}$ of classical fields (gauge fields, i.e., connections, and matter fields, i.e., sections of vector bundles) and an action $S(\boldsymbol{\xi})$. The action has various symmetries (e.g., gauge invariance, relativistic invariance, etc.) so that the physically significant object of study is the moduli space $\mathcal{F}/S$ of fields modulo symmetrics. Real-valued functions

$$\mathcal{O} : \mathcal{F}/S \longrightarrow$$

on the moduli space are called **observables**. In this context, "quantization" is viewed as the process of assigning expectation values $\langle \mathcal{O} \rangle$ to observables $\mathcal{O}$ and, according to the rules of the game, this is accomplished by a weighted integral over the moduli space which, when the underlying manifold is Riemannian (as opposed to semi-Riemannian), is of the form

$$\langle \mathcal{O} \rangle = \int_{F/\mathcal{S}} e^{-S(\boldsymbol{\xi})/e^2} \mathcal{O}([\boldsymbol{\xi}])\, [\mathcal{D}\boldsymbol{\xi}].$$

Here e is a "coupling constant" and $[\mathcal{D}\boldsymbol{\xi}]$ represents a (generally nonexistent) measure on the infinite-dimensional moduli space. Such integrals as a rule have no precise mathematical definition, but physicists compute with them to great effect nevertheless. The end result is a **quantum field theory**; it is called a **topological quantum field theory** (**TQFT**) if for some distinguished set of observables these expectation values are independent of the choice of Riemannian metric on the underlying manifold. The integral

$$Z = \int_{F/\mathcal{S}} e^{-S(\boldsymbol{\xi})/e^2} [\mathcal{D}\boldsymbol{\xi}]$$

is called the **partition function** of the quantum field theory.

Witten [**W2**] constructed the first example of such a TQFT with the specific intention of exhibiting the Donaldson invariants as expectation values of certain observables. The classical gauge theory from which the quantum field theory was built is rather more complicated than any we have encountered and has more structure than our examples, but very briefly it looked something like this: M as usual is a compact, simply connected, oriented, smooth

4-manifold with $b_2^+(M) > 1$ and odd and we choose some generic metric $\boldsymbol{g}$ on M. The field content consists of one gauge field (connection) $\boldsymbol{\omega}$ and five matter fields $\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\chi}$ and $\boldsymbol{\psi}$ all of which are forms (of various degrees) with values in the adjoint bundle ad P. But this gauge theory of Witten's is supersymmetric which means that each matter field is classified as either "bosonic" or "fermionic" and there is an additional (super) symmetry operator that interchanges bosons and fermions. We will make no attempt to explain what this means, but will simply refer the curious to [**W1**] and record the types of the matter fields:

Bosonic	Fermionic
$\boldsymbol{\phi} \in \Lambda^0(M, \text{ad}\, P)$	$\boldsymbol{\eta} \in \Lambda^0(M, \text{ad}\, P)$
$\boldsymbol{\lambda} \in \Lambda^0(M, \text{ad}\, P)$	$\boldsymbol{\chi} \in \Lambda^1(M, \text{ad}\, P)$
	$\boldsymbol{\psi} \in \Lambda^2_+(M, \text{ad}\, P)$

A 6-tuple $\boldsymbol{\xi} = (\boldsymbol{\omega},\ \boldsymbol{\phi},\ \boldsymbol{\lambda},\ \boldsymbol{\eta},\ \boldsymbol{\chi},\ \boldsymbol{\psi})$ of such fields is called a **field configuration** for the gauge theory. Now we need an action functional. This is called the **Donaldson-Witten Action**. It is somewhat more intimidating than the others we have seen and we will make no real use of it, but since everyone should see Witten's Lagrangian once in their lives, here it is,

$$S_{\mathcal{D}W}(\boldsymbol{\xi}) = \int_M \text{trace}\Big(-\frac{1}{4}\boldsymbol{F_\omega} \wedge {}^*\boldsymbol{F_\omega} - \frac{1}{4}\boldsymbol{F_\omega} \wedge \boldsymbol{F_\omega} + \frac{1}{2}[\boldsymbol{\psi},\ \boldsymbol{\psi}]\boldsymbol{\phi} + \boldsymbol{i}d^{\boldsymbol{\omega}}\boldsymbol{\chi} \wedge \boldsymbol{\psi} - 2\boldsymbol{i}[\boldsymbol{\chi}, {}^*\boldsymbol{\chi}]\,\boldsymbol{\lambda} + \boldsymbol{i}^*(\boldsymbol{\phi}\Delta_0^{\boldsymbol{\omega}}\boldsymbol{\lambda}) + \boldsymbol{\chi} \wedge {}^*d^{\boldsymbol{\omega}}\boldsymbol{\eta}\Big)$$

Remarks: A few of the terms in the action are familiar. The first is a Yang-Mills term. The second Witten calls a topological term since it is essentially the Chern class. The rest are to be regarded as interaction terms. The path Witten followed to arrive at the field content, the symmetries and the terms in the action is described in detail in [**W2**], but it is not a path easily traversed by mathematicians. Remarkably, Atiyah and Jeffrey [**AJ**] have shown that the fields and the action all arise naturally from purely geometrical considerations by formally applying to the infinite-dimensional vector bundle $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Lambda^2_+(M, \text{ad}\, P)$ a formula for the Euler class proved (for finite-dimensional vector bundles) by Mathai and Quillen [**MQ**].

$S_{\mathcal{D}W}(\boldsymbol{\xi})$ is a perfectly well-defined mathematical object, but now we must quantize and this means integrating over the entire moduli space of field configurations. The partition function, for example, is formally written as the path integral

$$Z_{\mathcal{D}W} = \int e^{-S_{\mathcal{D}W}(\boldsymbol{\xi})/e^2}[\mathcal{D}\boldsymbol{\xi}].$$

Now, a path integral is really not an integral at all, but just a suggestive notation for a certain limit (a limit that generally does not exist). Physicists have developed elaborate techniques for dealing with such "integrals", but we will say nothing about this. Rather we will simply sketch, in very broad terms, how Witten was led to identify $Z_{\mathcal{D}W}$ with $\gamma_0(M)$ in the case in which $8k - 3(1 + b_2^+\ (M)) = 0$. The crucial observation is that $Z_{\mathcal{D}W}$, whatever it means, is a function of the coupling constant e. At the classical level, e plays no real role and can simply be regarded as a rescaling of the action, but in the quantum theory its size will determine the computability (or not) of the relevant physical quantities. This is because the usual procedure for dealing with these quantities is to do perturbation calculations and this involves series expansions in e. If e is "small" (i.e., in what is called the "weak coupling limit") such calculations are extraordinarily effective, but if e is "large" ("strong coupling") they fail completely.

Witten computed $Z_{\mathcal{D}W}$ is the weak coupling limit by performing what the physicists call the "semi-classical approximation" which he concluded, based on all of the symmetries built into the action, must, in fact, be exact.

Remark: This is an infinite-dimensional analogue of the famous Duistermaat-Heckman Theorem on exact stationary phase approximations (see [**BV**]).

From this he argued that the path integral defining $Z_{\mathcal{D}W}$ "localizes" to an integral over the moduli space $\mathcal{M}$ of anti-self-dual connections on P and, when $\dim \mathcal{M} = 0$, this integral over the 0-dimensional moduli space is just a sum that can be identified with the Donaldson invariant $\gamma_0(M)$.

Remark: Intuitively, this is not unlike the Residue Theorem which localizes a contour integral around a closed path to a sum of contributions from the singularities of the integrand inside (although there are no "symmetry" considerations here). Much closer mathematical analogues are the Equivariant Localization Theorems of Berline and Vergne (see [**BV**]) which extract and generalize the essential content of the Duistermaat-Heckman Theorem.

Witten also isolated observables in his TQFT whose expectation values formally coincide with the integrals we used in our "naive" definition of $\gamma_d(M)$ for $d > 0$. Very briefly, it goes like this: For any field configuration $\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\chi}, \boldsymbol{\psi})$ define $W = \text{trace}(\frac{1}{4}\boldsymbol{\psi} \wedge \boldsymbol{\psi} + \boldsymbol{\phi}\boldsymbol{F}_{\boldsymbol{\omega}})$. This is an ordinary 2-form on M. For each homology class $x \in H_2(M)$, thought of as an embedded surface, define

$$\mathcal{O}(x) = \int_x W.$$

Each $\mathcal{O}(x)$ maps a field configuration to a real number. Again due to all of the built-in symmetries, $\mathcal{O}(x)$ is actually defined on the moduli space and so is an observable. Witten associates with each $\mathcal{O}(x)$ a closed 2-form $\alpha(x)$ on the moduli space $\mathcal{M}$ whose cohomology class can be identified with $\mu(x)(\mu$ is the

Donaldson μ-map). Finally, if $\dim \mathcal{M} = 2d,\ d \geq 1$, he shows that $\mathcal{O}^d$, defined on $H_2(M)$ by $\mathcal{O}^d(x) = (\mathcal{O}(x))^d$, is a real-valued function on the moduli space (i.e., an observable) whose expectation value, localized to $\mathcal{M}$, is

$$\langle \mathcal{O}^d(x) \rangle = \int_{\mathcal{M}} \alpha(x) \wedge \overset{d}{\cdots} \wedge \alpha(x)$$

and so coincides with our integral description of $\gamma_d(M)$.

All of this is quite formal, of course, and the end result is not really the Donaldson invariants but only our naive description of some of them. Nevertheless, it is remarkable. Still more remarkable is the fact that this is just the beginning of the story and that the best part is yet to come. To relate this part of the story we must first say a few words about another feature of Witten's TQFT. This is a very subtle and deep type of "duality" that we cannot do justice to here, but which has its origins in the fact that, unlike most quantum fields theories, its partition function and expectation values are actually independent of the coupling constant e. This is not a mathematical theorem, of course, but a consequence of formal path integral manipulations based on the myriad symmetries of the action. Ordinarily one would expect the weak and strong coupling regimes to describe very different physical systems, but in the case at hand they must be entirely equivalent, although their mathematical descriptions would no doubt look quite different (since, for example, perturbation calculations are possible in one, but not the other). This suggests the possibility of an entirely different description of the Donaldson invariants buried in the strong coupling regime. Witten was well aware of this in 1988 when **[W2]** appeared, but could do nothing about it because no one knew how to compute anything in the strong coupling (nonperturbative) regime. And so matters stood until 1994 when Seiberg and Witten developed entirely new techniques for doing exact calculations in strong coupling for so-called $N = 2$ supersymmetric Yang-Mills theories. This done, Witten dusted off his old TQFT, applied the new techniques and uncovered the dual version of Donaldson theory. This is, of course, the Seiberg-Witten theory that we have been leading up to all along. The details of the argument leading from Donaldson-Witten to Seiberg-Witten are at the deepest levels of theoretical physics and, alas, quite beyond the powers of your poor author whose only service can be to refer the stout-hearted among his readers to **[W3]**. The end result, however, was what Witten was convinced must be a "substitute" for Donaldson theory. It contained new fields (not a single $SU(2)$-connection, but a $U(1)$-connection and a spinor field), new equations (not the anti-self-dual equations, but the Seiberg-Witten equations which are, from the point-of-view of partial differential equations, much simpler), a new moduli space and new invariants. But the physical equivalence of the quantum field theories from which they arose ("duality") left no doubt in Witten's mind that they must contain the same topological information. The story of how this conjecture was sprung on the mathematical community and the pandemonium that ensued has been told many times, but is best heard from someone who was there so for this as well as a lovely introduction to what is to come here and a pleasant afternoon's entertainment we heartily recommend [**Tau2**]. We will now get on with the business of describing the mathematical side of the new classical gauge theory. When it is possible, with the background we have at our disposal, to provide details we will do so; when it is not, we will try to provide a sense of what is involved, what needs to be learned and where it can be learned.

A.2 Clifford Algebra and Spin^c-Structures

Seiberg-Witten invariants are defined, just as the Donaldson invariants are, from a moduli space of solutions to certain partial differential equations. They are much simpler to deal with than those of Donaldson, but the price one must pay is that just writing down the equations involves a rather substantial investment of time in various algebraic preliminaries.

Much of the algebraic background we require is most conveniently phrased in the language of Clifford algebras. We recall that any finite dimensional, real vector space V with an inner product $\langle\ ,\ \rangle$ has a Clifford algebra $Cl(V)$ which can be described abstractly as the quotient of the tensor algebra $\mathcal{J}(V)$ by the 2-sided ideal $\mathcal{I}(V)$ generated by elements of the form $v \otimes v + \langle\ ,\ \rangle 1$ with $v \in V$. More concretely, if $\{e_1, \ldots, e_n\}$ is an orthonormal basis for V, then $Cl(V)$ is the real associative algebra with unit 1 generated by $\{e_1, \ldots, e_n\}$ and subject to the relations

$$e_i e_j + e_j e_i = -2\langle e_i, e_j \rangle 1, \qquad i, j = 1, \ldots, n. \tag{A.2.1}$$

We intend to be even more concrete and construct an explicit matrix model for the Clifford algebra $Cl(4) = Cl(\ \ ^4)$ of $\ \ ^4$ with its usual positive definite inner product. The procedure will be to identify $\ \ ^4$ with a real linear subspace of a matrix algebra, find an orthonormal basis for this copy of $\ \ ^4$ satisfying the defining conditions (A.2.1), where the product is matrix multiplication and 1 is the identity matrix, and form the subalgebra it generates.

One can, of course, identify $\ \ ^4$ with the algebra $\ \ $ of quaternions $q = q^1 + q^2\mathrm{i} + q^3\mathrm{j} + q^4\mathrm{k}$, but we wish to embed this into the real, associative algebra $\ \ ^{2\times 2}$ of 2×2 quaternionic matrices:

$$\ \ ^{2\times 2} = \left\{ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} : q_{ij} \in \ \ , \qquad i, j = 1, 2 \right\}$$

Specifically, we identify $\ \ ^4$ with the real linear subspace of $\ \ ^{2\times 2}$ consisting of all elements of the form

$$x = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix}, \quad q \in \ \ \tag{A.2.2}$$

(this is, of course, *not* a subalgebra of $\;^{2\times 2}$). Notice that $\det x = \|q\|^2$ so, defining a norm on the set of x given by (A.2.2) by

$$\|x\|^2 = \det x \tag{A.2.3}$$

an inner product by polarization ($\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$ we find that the subspace of $\;^{2\times 2}$ consisting of all x of the form (A.2.2) is isomorphic to $\;^4$ as an inner product space. One easily checks that $\{e_1, e_2, e_3, e_4\}$ given by

$$e_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ e_2 = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix},\ e_3 = \begin{pmatrix} 0 & \mathrm{j} \\ \mathrm{j} & 0 \end{pmatrix},\ e_4 = \begin{pmatrix} 0 & \mathrm{k} \\ \mathrm{k} & 0 \end{pmatrix} \tag{A.2.4}$$

is an orthonormal basis and, moreover, satisfies

$$e_i e_j + e_j e_i = -2\langle e_i, e_j\rangle \mathbb{1}, \qquad i,j = 1,2,3,4, \tag{A.2.5}$$

where we use $\mathbb{1}$ generically for the identity matrix of any size (2×2 in this case). Note that it follows from (A.2.5) that

$$xy + yx = -2\langle x,y\rangle, \qquad x,y \in \;^4. \tag{A.2.6}$$

The real subalgebra of $\;^{2\times 2}$ generated by $\{e_1, e_2, e_3, e_4\}$ is the **real Clifford algebra** of $\;^4$ and is denoted $Cl(4)$. Writing out products of basis vectors and using (A.2.5) to eliminate linear dependencies gives the following basis for $Cl(4)$:

$$\begin{gathered}
e_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{1} \\
e_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\ e_2 = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}\ e_3 = \begin{pmatrix} 0 & \mathrm{j} \\ \mathrm{j} & 0 \end{pmatrix}\ e_4 = \begin{pmatrix} 0 & \mathrm{k} \\ \mathrm{k} & 0 \end{pmatrix} \\
e_1e_2 = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}\ e_1e_3 = \begin{pmatrix} \mathrm{j} & 0 \\ 0 & -\mathrm{j} \end{pmatrix}\ e_1e_4 = \begin{pmatrix} \mathrm{k} & 0 \\ 0 & -\mathrm{k} \end{pmatrix} \\
e_2e_3 = \begin{pmatrix} \mathrm{k} & 0 \\ 0 & \mathrm{k} \end{pmatrix}\ e_2e_4 = \begin{pmatrix} -\mathrm{j} & 0 \\ 0 & -\mathrm{j} \end{pmatrix}\ e_3e_4 = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & \mathrm{i} \end{pmatrix} \\
e_1e_2e_3 = \begin{pmatrix} 0 & \mathrm{k} \\ -\mathrm{k} & 0 \end{pmatrix}\ e_1e_2e_4 = \begin{pmatrix} 0 & -\mathrm{j} \\ \mathrm{j} & 0 \end{pmatrix} \\
e_1e_3e_4 = \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}\ e_2e_3e_4 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \\
e_1e_2e_3e_4 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}
\end{gathered} \tag{A.2.7}$$

Thus, $\dim Cl(4) = 16$. Since $\;^{2\times 2}$ itself has real dimension 16 we conclude that, in fact,

$$Cl(4) = \;^{2\times 2}. \tag{A.2.8}$$

Notice that the basis (A.2.7) gives $Cl(4)$ a natural $\mathbb{Z}_2$-grading

$$Cl(4) = Cl_0(4) \oplus Cl_1(4), \tag{A.2.9}$$

where $Cl_0(4)$ is spanned by $e_0,\ e_1e_2,\ e_1e_3,\ e_1e_4,\ e_2e_3,\ e_2e_4,\ e_3e_4$ and $e_1e_2e_3e_4$ and $Cl_1(4)$ is spanned by $e_1,\ e_2,\ e_3,\ e_4,\ e_1e_2e_3,\ e_1e_2e_4,\ e_1e_3e_4$ and $e_2e_3e_4$. The elements of $Cl_0(4)$ are said to be **even**, while those of $Cl_1(4)$ are **odd**. Regarding $\mathbb{Z}_2$ as $\{0, 1\}$ with addition modulo 2,

$$(Cl_i(4))\,(Cl_j(4)) \subseteq Cl_{i+j}(4) \tag{A.2.10}$$

for $i, j = 0, 1$, so $Cl(4)$ is a $\mathbb{Z}_2$-graded algebra, i.e., a superalgebra. From (A.2.7) it is clear that the decomposition (A.2.9) corresponds simply to

$$\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} q_{11} & 0 \\ 0 & q_{22} \end{pmatrix} + \begin{pmatrix} 0 & q_{12} \\ q_{21} & 0 \end{pmatrix}.$$

Lemma A.2.1 *The center $Z(Cl(4))$ of $Cl(4)$ is Span$\{e_0\} \cong\;$.*

Proof: Since $e_0 = \mathbb{1}$ it commutes with everything in $Cl(4)$ so $\mathrm{Span}\{e_0\} \subseteq Z(Cl(4))$ is clear. To complete the proof it will suffice to show that every

$$e_I = e_{i_1} \cdots e_{i_k}, \quad 1 \le k \le 4, \quad 1 \le i_1 < \cdots < i_k \le 4$$

fails to commute with something in $Cl(4)$. For $k = 1$ this is clear since $e_ie_j = -e_je_i$ for $i \ne j$. For $k = 4$, $e_I = e_1e_2e_3e_4$ so $e_1e_I = (e_1e_1)e_2e_3e_4 = -e_2e_3e_4$, whereas $e_Ie_1 = (e_1e_2e_3e_4)e_1 = (-1)^3(e_1e_1)e_2e_3e_4 = e_2e_3e_4$. Now suppose $1 < k < 4$. Then $e_Ie_{i_1} = (-1)^{k-1}e_{i_1}e_I$ and, if e_l is not among $e_{i_1}, \ldots, e_{i_k}$, $e_Ie_l = (-1)^k e_le_I$. Thus, e_I cannot commute with both e_{i_1} and e_l. ■

Lemma A.2.2 *If $x \in \;^4 \subseteq Cl(4)$ and $\|x\| = 1$, then x is a unit in $Cl(4)$ (i.e., is invertible) and $x^{-1} = -x$.*

Proof: $\langle x, x\rangle = 1$ and $xx + xx = -2\langle x, x\rangle \mathbb{1}$ imply $xx = -\mathbb{1}$. ■

We denote by $Cl^\times(4)$ the multiplicative group of units in $Cl(4)$ and by Pin(4) the subgroup of $Cl^\times(4)$ generated by all of the $x \in \;^4$ with $\|x\| = 1$ (see Lemma A.2.2). Now, an x of the form (A.2.2) has $\|x\| = 1$ if and only if $q \in \mathrm{Sp}(1)$ (the Lie group of unit quaternions) and the set of all such is closed under inversion ($x^{-1} = -x$). Thus, Pin(4) is just the set of all products of such elements. The even elements of Pin(4) are just its diagonal elements and they form a subgroup denoted

$$\begin{aligned}\mathrm{Spin}(4) &= \mathrm{Pin}(4) \cap Cl_0(4) \\ &= \left\{ \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix} : u_1, u_2 \in \mathrm{Sp}(1) \right\} \\ &\cong \mathrm{Sp}(1) \times \mathrm{Sp}(1).\end{aligned} \tag{A.2.11}$$

The topology and differentiable structure Spin(4) inherits from $^{2\times 2} \cong \ {}^{4} \cong \ {}^{16}$ are the product structures from Sp(1) (which is diffeomorphic to S^3) so Spin(4) is a compact, simply connected Lie group. Since the Lie algebra of Sp(1) can be identified with the pure imaginary quaternions Im , the Lie algebra of Spin(4) can be identified with

$$spin(4) = \left\{ \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} : Q_1,\ Q_2 \in \mathrm{Im} \quad \right\}. \tag{A.2.12}$$

The significance of Spin(4) lies in the following theorem.

Theorem A.2.3 Spin(4) *is a simply connected double cover of* $SO(4)$*.*

Proof: Since Spin(4) is a simply connected Lie group we need only show that it is a double cover of $SO(4)$. For this we consider the adjoint action of $Cl^\times(4)$ on $Cl(4)$, i.e., for each $u \in Cl^\times(4)$ we define a map

$$\mathrm{ad}_u : Cl(4) \longrightarrow Cl(4)$$

by

$$\mathrm{ad}_u(p) = upu^{-1} \tag{A.2.13}$$

for each $p \in Cl(4)$. This is clearly an algebra isomorphism that preserves the grading (A.2.9). Note that if $x \in \ {}^4 \subseteq Cl(4)$ has $\|x\| = 1$, then, for every $v \in \ {}^4 \subseteq Cl(4)$,

$$\mathrm{ad}_x(v) = xvx^{-1} = xv(-x) = -xvx$$

so the identity $vx + xv = -2\langle v, x\rangle \mathbb{1}$ implies

$$\begin{aligned} xvx + xxv &= -2\langle v,\ x\rangle\ x \\ xvx - v &= -2\langle v,\ x\rangle\ x \\ xvx &= v - 2\langle v,\ x\rangle\ x \\ xvx &= (v - \langle v,\ x\rangle\ x) - \langle v,\ x\}\ x. \end{aligned}$$

Now, $v - \langle v, x\rangle x$ is the projection of v into the hyperplane $x^\perp$ orthogonal to x so xvx is the reflection of v through $x^\perp$, written $\mathrm{Refl}_{x^\perp}(v)$. Thus,

$$\mathrm{ad}_x(v) = -\mathrm{Refl}_{x^\perp}(v). \tag{A.2.14}$$

In particular,

$$\mathrm{ad}_x : \ {}^4 \longrightarrow \ {}^4 \quad (x \in \ {}^4, \ \|x\| = 1).$$

But any element of Pin(4) is a product of elements $x \in \ {}^4$ with $\|x\| = 1$ so

$$\mathrm{ad}_u : \ {}^4 \longrightarrow \ {}^4 \quad (u \in \mathrm{Pin}(4)). \tag{A.2.15}$$

Since any product of reflections is an orthogonal transformation, ad_u is an orthogonal transformation for each $u \in \mathrm{Pin}(4)$. Since an element of Spin(4) is a product of an even number of $x \in \ {}^4$ with $\|x\| = 1$ and any product of an even number of reflections is a rotation we have

$$\mathrm{ad}_u \in SO(\ {}^4) \quad (u \in \mathrm{Spin}(4)). \tag{A.2.16}$$

Thus, we have a map, called the **spinor map**,

$$\mathrm{Spin} : \mathrm{Spin}(4) \longrightarrow SO(\ {}^4) \cong SO(4) \tag{A.2.17}$$

$$\mathrm{Spin}(u) = \mathrm{ad}_u.$$

Since any reflection can clearly be written as $-\mathrm{ad}_x$ for some $x \in \ {}^4$ with $\|x\| = 1$ and since any rotation can be written as a product of an even number of reflections, the spinor map is a surjective group homomorphism. Finally, to see that $\ker(\mathrm{Spin}) = \mathbb{Z}_2 = \{\pm \mathbb{1}\}$, so that it is precisely two-to-one, note that ad_u is the identity in $SO(4)$ if and only if $uxu^{-1} = x$ for each $x \in \ {}^4$. But then u must commute with everything in $Cl(4)$, i.e., $u \in Z(Cl(4))$. By Lemma A.2.1, $u = ae_0 = a\mathbb{1}$ for some $a \in$. By (A.2.11), $a^2 = 1$ and $u = \pm\mathbb{1}$. ■

Remark: Globalizing these constructions leads to the notion of a "spin structure" on a manifold. In the context of spacetime (as opposed to Riemannian) manifolds this is just the "spinor structure" we introduced and studied in Sections 3.5 and 6.5 so we will just briefly describe the definition and then explain why we need the more general concept of a "spin^c structure". We let B denote a compact, oriented, smooth 4-manifold with a Riemannian metric $\boldsymbol{g}$. Let $SO(4) \hookrightarrow F_{SO}(B) \xrightarrow{\pi_{SO}} B$ denote the corresponding oriented, orthonormal frame bundle. A **spin structure** $\mathcal{S}$ consists of a principal Spin(4)-bundle

$$\mathrm{Spin}(4) \hookrightarrow S(B) \xrightarrow{\pi_S} B$$

over B and a smooth map

$$\lambda : S(B) \longrightarrow F_{SO}(B)$$

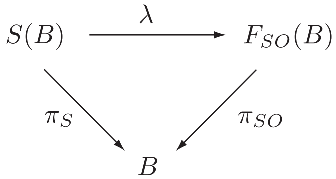
satisfying

$$\pi_{SO} \circ \lambda = \pi_S \tag{A.2.18}$$

and

$$\lambda(p \cdot u) = \lambda(p) \cdot \mathrm{Spin}(u) \tag{A.2.19}$$

for each $p \in S(B)$ and each $u \in \text{Spin}(4)$.



The fibers of $F_{SO}(B)$ are copies of $SO(4)$ so (A.2.18) says that we have a copy of Spin(4) "above" each of these and (A.2.19) says that the map λ of $S(B)$ onto $F_{SO}(B)$ is essentially the spinor map at each point of B. Now, unlike the frame bundle $F_{SO}(B)$, which exists for any manifold of the type we have described, there is an obstruction to the existence to a spin structure. The arguments of Section 6.5 carry over *verbatim* to show that B admits a spin structure if and only if the 2nd Stiefel-Whitney class $w_2(B) \in \check{H}^2(B; \mathbb{Z}_2)$ is trivial. Unfortunately, many interesting 4-manifolds (e.g., $\ ^2$) do not satisfy this condition and without a spin structure one cannot define "spinor fields" in the usual sense. Since spinor fields are crucial to Seiberg-Witten theory and since one would like this theory to apply to as many 4-manifolds as possible we seek a generalized notion of both "spin structure" and "spinor field". As it happens, there is a very natural generalization obtained by complexifying our previous algebraic considerations.

To define complex analogues of the algebraic objects we have introduced we will embed $Cl(4)$ into a complex algebra of matrices and form the complex subalgebra it generates. The basic tool we use is the usual matrix model of the quaternions. Specifically, we consider the map $\gamma : \ \longrightarrow \ ^{2\times 2}$ from the quaternions to the 2×2 complex matrices given by

$$\gamma(q) = \gamma(\alpha + \beta \mathrm{j}) = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \tag{A.2.20}$$

where we have written

$$q = q^1 + q^2\mathrm{i} + q^3\mathrm{j} + q^4\mathrm{k} = (q^1 + q^2\mathrm{i}) + (q^3 + q^4\mathrm{i})\mathrm{j} = \alpha + \beta\mathrm{j}.$$

One easily verifies that γ is real linear, injective, preserves products, carries $\bar{q}$ to $\overline{\gamma(q)}^{\top}$ and satisfies $\det(\gamma(q)) = \|q\|^2$ so that we can identify with the set of all 2×2 complex matrices of the form $\left(\begin{smallmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{smallmatrix}\right)$. More specifically, if we let

$$\gamma(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{1} \qquad \gamma(\mathrm{i}) = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix} = I$$
$$\gamma(\mathrm{j}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = J \qquad \gamma(\mathrm{k}) = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} = K \tag{A.2.21}$$

then we can identify $q = q^1 + q^2\mathrm{i} + q^3\mathrm{j} + q^4\mathrm{k}$ with

$$q = q^1\mathbb{1} + q^2 I + q^3 J + q^4 K. \tag{A.2.22}$$

Now we identify $Cl(4) = {}^{2\times 2}$ with a subset of ${}^{4\times 4}$. Define $\Gamma : {}^{2\times 2} \longrightarrow {}^{4\times 4}$ by

$$\Gamma\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} \gamma(q_{11}) & \gamma(q_{12}) \\ \gamma(q_{21}) & \gamma(q_{22}) \end{pmatrix}, \tag{A.2.23}$$

where each $\gamma(q_{ij})$ is a 2×2 block in the matrix on the right-hand side. This map Γ is also real linear, injective and preserves products so we can identify the real algebra $Cl(4)$ with its image.

$$Cl(4) = \Gamma({}^{2\times 2})$$

The restriction of Γ to ${}^4 \subseteq Cl(4)$ is

$$x = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix} \longrightarrow \Gamma(x) = \begin{pmatrix} 0 & \gamma(q) \\ -\overline{\gamma(q)}^{\top} & 0 \end{pmatrix}. \tag{A.2.24}$$

Since $\det \Gamma(x) = \det x = \|x\|^2 = \|q\|^2$ we can define an inner product via polarization on this copy of ^{4} from $\|\Gamma(x)\|^2 = \det \Gamma(x)$ and then $\Gamma|\, {}^4$ becomes an isometry. We now fully identify ^{4} with this copy and obtain the basis

$$E_1 = \gamma(e_1) = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \qquad E_2 = \gamma(e_2) = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$$
$$E_3 = \gamma(e_3) = \begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix} \qquad E_4 = \gamma(e_4) = \begin{pmatrix} 0 & K \\ K & 0 \end{pmatrix} \tag{A.2.25}$$

satisfying

$$E_i E_j + E_j E_i = -2\langle E_i, E_j \rangle \mathbb{1}, \quad i, j = 1, 2, 3, 4. \tag{A.2.26}$$

In this context, $Cl(4)$ is the *real* subalgebra of ${}^{4\times 4}$ generated by $\{E_1, E_2, E_3, E_4\}$ and a basis is as in (A.2.7), but with everything capitalized (and 1 changed to $\mathbb{1}$). Under γ, $\mathrm{Sp}(1)$ is mapped to $SU(2)$ so, in our new model of $Cl(4)$ we have the identifications

$$\mathrm{Spin}(4) = \left\{\begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} : U_1, U_2 \in SU(2)\right\} \cong SU(2) \times SU(2) \tag{A.2.27}$$

and

$$spin(4) = \left\{ \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} : A_1, A_2 \in su(2) \right\} \tag{A.2.28}$$

corresponding to (A.2.11) and (A.2.12).

Now we regard $^{4\times 4}$ as a complex algebra and define the **complexified Clifford algebra** $Cl(4) \otimes$ to be the complex subalgebra generated by $\{E_1, E_2, E_3, E_4\}$, i.e., by $Cl(4)$. A basis over is given by (A.2.7), with all of the e_i capitalized. Since $^{4\times 4}$ also has dimension 16 over we conclude that

$$Cl(4) \otimes \quad = \quad ^{4\times 4}. \tag{A.2.29}$$

Now let

$$S \quad = \quad ^4$$

be the complex vector space ^{4} with its usual Hermitian inner product $(\langle z, w \rangle = \bar{z}^1 w^1 + \bar{z}^2 w^2 + \bar{z}^3 w^3 + \bar{z}^4 w^4))$ and identify $Cl(4) \otimes$ with the vector space End $(S \ )$ of complex linear transformations of S to itself:

$$Cl(4) \otimes \quad = \text{End} \ (S \ ) \tag{A.2.30}$$

Thus, the elements of $Cl(4) \otimes$ (and therefore also $Cl(4)$, ^{4} and Spin(4)) act as endomorphisms of S . This action is called **Clifford multiplication** and will be written with a dot $\cdot$. In particular, we have a representation of the real Clifford algebra by endomorphisms of S :

$$Cl(4) \longrightarrow \text{End} \ (S \ )$$

(representations of algebras are by endomorphisms rather than isomorphisms since not all elements of an algebra are units). This representation of $Cl(4)$ is easily seen to be irreducible by writing out the real linear combinations of the basis $E_0, \ldots, E_1E_2E_3E_4$ for $Cl(4) \subseteq Cl(4) \otimes$. Restricting the Clifford action further to $\text{Spin}(4) \subseteq Cl(4) \otimes$ gives a *group* respresentation of Spin(4) on S :

$$\Delta \quad : \text{Spin}(4) \longrightarrow \text{Aut} \ (S \ )$$

(by automorphisms now since the elements of Spin(4) are all units). This is called the **complex spin representation** and, as we shall now see, is *not* irreducible. Indeed, if we write

$$S \quad \cong S^+ \oplus S^-$$

$$\begin{pmatrix} z^1 \\ z^2 \\ z^3 \\ z^4 \end{pmatrix} = \begin{pmatrix} z^1 \\ z^2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ z^3 \\ z^4 \end{pmatrix}, \tag{A.2.31}$$

then Clifford multiplication by elements of $Cl_0(4)$, because they are block diagonal, preserves S^+ and S^-, whereas Clifford multiplication by elements of $Cl_1(4)$, because they are block anti-diagonal, interchanges S^+ and S^-. In particular, Δ resolves into a direct sum

$$\Delta \quad = \Delta^+ \oplus \Delta^-$$

where

$$\Delta^\pm : \mathrm{Spin}(4) \longrightarrow SU(S^\pm)$$

(see (A.2.27) for the "SU"). Δ^+ and Δ^- are inequivalent, irreducible representations of Spin(4). Notice also that Clifford multiplication by the elements of $\ ^4$, which are odd, interchanges S^+ and S^- (this will be crucial when we define the "Dirac operator" shortly).

Recall that Spin(4) is the set of all even elements in the subgroup of multiplicative units in the Clifford algebra $Cl(4)$ generated by the unit sphere in $\ ^4 \subseteq Cl(4)$. For the complex analogue we add to the generators the unit circle in . More precisely, we identify $U(1)$ with the subset

$$U(1) = \{\mathrm{e}^{\theta\,\mathrm{i}}\,\mathbb{1} : \theta \in \quad \}$$

of $Cl(4) \otimes$ (often dropping the "$\mathbb{1}$" and thinking of $\mathrm{e}^{\theta\,\mathrm{i}}$ as an element of $Cl(4) \otimes$). Then

$$\mathrm{Spin}^c(4)$$

is defined to be the subgroup of the group of multiplicative units in $Cl(4) \otimes$ generated by Spin(4) and $U(1)$. Notice that the elements of $\mathrm{Spin}^c(4)$ are necessarily even, i.e., in $Cl_0(4) \otimes$. Since $U(1)$ is in the center of $Cl(4) \otimes$ we have

$$\begin{aligned} \mathrm{Spin}^c(4) &= \{\mathrm{e}^{\theta\,\mathrm{i}}\,u : \theta \in \quad , \quad u \in \mathrm{Spin}(4)\} \\ &= \left\{ \mathrm{e}^{\theta\,\mathrm{i}} \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} : \theta \in \quad , \quad U_1, U_2 \in SU(2) \right\}. \end{aligned} \tag{A.2.32}$$

Note that $U_1, U_2 \in SU(2)$ implies $\mathrm{e}^{\theta\,\mathrm{i}}\,U_1, \mathrm{e}^{\theta\,\mathrm{i}}\,U_2 \in U(2)$ and

$$\det(\mathrm{e}^{\theta\,\mathrm{i}}\,U_1) = \det(\mathrm{e}^{\theta\,\mathrm{i}}\,U_2) = \mathrm{e}^{2\theta\,\mathrm{i}}. \tag{A.2.33}$$

Since every element of $U(2)$ can be written as $\mathrm{e}^{\theta\,\mathrm{i}}\,U$, $U \in SU(2)$ (uniquely up to a simultaneous change of sign for both $\mathrm{e}^{\theta\,\mathrm{i}}$ and U) we have

$$\mathrm{Spin}^c(4) = \left\{ \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix} : U_\pm \in U(2), \quad \det U_+ = \det U_- \right\}. \tag{A.2.34}$$

There is yet another useful way of looking at $\mathrm{Spin}^c(4)$. The mapping

$$\begin{aligned} \mathrm{Spin}(4) \times U(1) &\longrightarrow \mathrm{Spin}^c(4) \\ (u, \mathrm{e}^{\theta\,\mathrm{i}}\,\mathbb{1}) &\longrightarrow \mathrm{e}^{\theta\,\mathrm{i}}\,u \end{aligned}$$

is a surjective homomorphism. Its kernel is the set of (α, α^{-1}), where $\alpha \in \mathrm{Spin}(4)$. But $\mathrm{Spin}(4)$ intersects the scalars only in $\pm\mathbb{1}$ so this kernel is $\mathbb{Z}_2 = \pm(\mathbb{1}, \mathbb{1})$. Thus,

$$\mathrm{Spin}^c(4) \cong \mathrm{Spin}(4) \times U(1)/\mathbb{Z}_2. \tag{A.2.35}$$

Finally, notice that, from Lemma A.2.1 and (A.2.32) it follows that the center of $\mathrm{Spin}^c(4)$ is

$$Z\left(\mathrm{Spin}^c(4)\right) = U(1). \tag{A.2.36}$$

Globalizing all of this to 4-manifolds will require a few mappings which we now introduce. First define

$$\delta : \mathrm{Spin}^c(4) \longrightarrow U(1)$$

as follows: For

$$\xi = \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{\theta\,\mathrm{i}}\, U_1 & 0 \\ 0 & \mathrm{e}^{\theta\,\mathrm{i}}\, U_2 \end{pmatrix} \in \mathrm{Spin}^c(4),$$

$$\delta(\xi) = \det U_+ = \det U_- = \mathrm{e}^{2\theta\,\mathrm{i}}. \tag{A.2.37}$$

Then δ is a surjective homomorphism with kernel $\mathrm{Spin}(4)$. Next define

$$\pi : \mathrm{Spin}^c(4) \longrightarrow SO(4)$$

as follows: The adjoint action of $\mathrm{Spin}(4)$ on ^{4} extends to an adjoint action of $\mathrm{Spin}^c(4)$ on ^{4}. Indeed, if $\xi = \mathrm{e}^{\theta\,\mathrm{i}}\, u \in \mathrm{Spin}^c(4)$, then, for each $x \in {}^4$, $\mathrm{ad}_\xi(x) = \xi x \xi^{-1} = u x u^{-1} = \mathrm{ad}_u(x)$ so, on ^{4},

$$\mathrm{ad}_\xi = \mathrm{ad}_u = \mathrm{Spin}(u) \in SO\left({}^4\right) \cong SO(4)$$

and we may take

$$\pi(\xi) = \mathrm{ad}_\xi = \mathrm{ad}_u = \mathrm{Spin}(u). \tag{A.2.38}$$

Finally, define

$$\mathrm{Spin}^c : \mathrm{Spin}^c(4) \longrightarrow SO(4) \times U(1)$$

by

$$\mathrm{Spin}^c(\xi) = \mathrm{Spin}^c\!\left(\mathrm{e}^{\theta\,\mathrm{i}}\, u\right) = \left(\pi\left(\xi\right), \delta\left(\xi\right)\right) = \left(\mathrm{Spin}(u), \mathrm{e}^{2\theta\,\mathrm{i}}\right). \tag{A.2.39}$$

Then Spin^c is a surjective homomorphism whose kernel is easily seen to be $\mathbb{Z}_2 = \pm\mathbb{1}$. It follows that $\mathrm{Spin}^c(4)$ is a double cover of $SO(4) \times U(1)$. Thus, the Lie algebra $\mathit{spin}^c\,(4)$ is $\mathit{so}(4) \times u(1) \cong \mathit{spin}(4) \times u(1)$ and can be identified with the subset

$$\mathit{spin}^c\,(4) = \left\{ \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} + ti \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix} : t \in \quad, A_1, A_2 \in su(2) \right\} \tag{A.2.40}$$

of $Cl(4) \otimes$ (in particular, $spin^c(4)$ also acts by Clifford multiplication on S, preserving both $S^{\pm}$).

Now, the identification $Cl(4) \otimes \ = \text{End}\ (S\ )$ and the fact that the elements of $\text{Spin}^c(4)$ are all units implies that the complex spin representation $\Delta\ : \text{Spin}(4) \longrightarrow \text{Aut}\ (S\ )$ extends to a representation

$$\hat{\Delta}\ : \text{Spin}^c(4) \longrightarrow \text{Aut}\ (S\ ).$$

Since the elements of $\text{Spin}^c(4)$ are block diagonal, $\hat{\Delta}$ also splits into

$$\hat{\Delta}\ = \hat{\Delta}^+ \oplus \hat{\Delta}^-$$

where

$$\hat{\Delta}^{\pm} : \text{Spin}^c(4) \longrightarrow U(S^{\pm}) \tag{A.2.41}$$

(see (A.2.34)).

One of the Seiberg-Witten equations relates the self-dual part of the curvature of a $U(1)$-connection to a certain trace free endomorphism of a positive spinor. The last of our algebraic preliminaries describes the relationship between 2-forms and endomorphisms. We note that there is a natural linear isomorphism from the space $\bigwedge^2(\ ^4,\ )$ of (complex-valued) 2-forms on ^{4} into $Cl(4) \otimes$. Indeed, if $\{e_1, e_2, e_3, e_4\}$ is the standard basis for ^{4} and $\{e^1, e^2, e^3, e^4\}$ is its dual, then we define

$$\rho : \bigwedge^2 (\ ^4,\ ) \longrightarrow Cl_0(4) \otimes$$

by

$$\rho(\eta) = \rho\left(\sum_{i<j} \eta_{ij} e^i \wedge e^j\right) = \sum_{i<j} \eta_{ij} E_i E_j =$$

$$\begin{pmatrix} (\eta_{12}+\eta_{34})I + (\eta_{13}-\eta_{24})J + (\eta_{14}+\eta_{23})K & 0 \\ 0 & (-\eta_{12}+\eta_{34})I + (-\eta_{13}-\eta_{24})J + (-\eta_{14}+\eta_{23})K \end{pmatrix}. \tag{A.2.42}$$

Notice that, although ρ is clearly a linear isomorphism, it is not multiplicative, e.g., $e^1 \wedge e^1 = 0$, but $E_1 E_1 = -$. There is, of course, an analogous map in any rank. Notice that if $\eta \in \bigwedge^2(\ ^4,\ )$ is real-valued (respectively, Im -valued), then $\rho(\eta)$ is skew-Hermitian (respectively, Hermitian). For example, if η is real-valued,

$$\begin{aligned}\overline{\rho(\eta)}^T &= \sum_{i<j} \bar{\eta}_{ij} \overline{E_i E_j}^T = \sum_{i<j} \eta_{ij} \bar{E}_j^T \bar{E}_i^T \\ &= \sum_{i<j} \eta_{ij}(-E_j)(-E_i) = \sum_{i<j} \eta_{ij} E_j E_i \\ &= \sum_{i<j} \eta_{ij}(-E_i E_j) \\ &= -\rho(\eta).\end{aligned}$$

Note also that, in (A.2.42), $\{e_a\}$ can be replaced by any oriented, orthonormal basis provided $\{E_a\}$ is replaced by its image under Γ (see (A.2.25)).

Now, being even (i.e., block diagonal) any $\rho(\eta)$ preserves the subspaces $S^{\pm}$ of $S\ $ and so we obtain endomorphisms of $S^{\pm}$ by setting

$$\rho^{\pm}(\eta) = \rho(\eta)\,|\,S^{\pm}. \tag{A.2.43}$$

For example, suppressing the two zero entries in S^+ (see (A.2.31)),

$$\rho^{+}(\eta) = (\eta_{12} + \eta_{34})I + (\eta_{13} + \eta_{42})J + (\eta_{14} + \eta_{23})K. \tag{A.2.44}$$

Thus, we have two maps

$$\rho^{\pm} : \bigwedge^2(\ ^4,\ ) \longrightarrow \text{End}\ \ (S^{\pm}). \tag{A.2.45}$$

Now let $\bigwedge^2(\ ^4,\ ) = \bigwedge^2_+(\ ^4,\ ) \oplus \bigwedge^2_-(\ ^4,\ )$ be the decomposition of $\bigwedge^2(\ ^4,\ )$ into self-dual and anti-self-dual 2-forms (relative to the Hodge star $*$ for the usual orientation and inner product on $\ ^4$). We show that $\rho^{\pm}$ carries $\bigwedge^2_{\pm}(\ ^4,\ )$ isomorphically onto the space $\text{End}_0(S^{\pm})$ of trace free (complex) endomorphisms of $S^{\pm}$.

Lemma A.2.4 $\rho^{\pm}|\bigwedge^2_{\pm}(\ ^4,\ )$ *is a complex linear isomorphism onto* $\text{End}_0(S^{\pm})$.

Proof: We give the argument for $\rho^{+}|\bigwedge^2_{+}(\ ^4,\ )$. The $\rho^{-}|\bigwedge^2_{-}(\ ^4,\ )$ case is analogous. A simple computation from (A.2.42) gives

$$\begin{aligned}\rho(e^1 \wedge e^2 + e^3 \wedge e^4) &= 2\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \\ \rho(e^1 \wedge e^3 + e^4 \wedge e^2) &= 2\begin{pmatrix} J & 0 \\ 0 & 0 \end{pmatrix} \\ \rho(e^1 \wedge e^4 + e^2 \wedge e^3) &= 2\begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned} \tag{A.2.46}$$

Since $\{e^1 \wedge e^2 + e^3 \wedge e^4,\ e^1 \wedge e^3 + e^4 \wedge e^2,\ e^1 \wedge e^4 + e^2 \wedge e^3\}$ spans the set of self-dual 2-forms on $\ ^4$, it is clear that $\rho^+ \mid \bigwedge^2_+(\ ^4,\ \ )$ is a linear, injective map to $\mathrm{End}(S^+)$. Because $I,\ J$ and K are trace free, so is everything in the image of $\rho^+ \mid \bigwedge^2_+(\ ^4,\ \ )$. Furthermore, one can show that every 2×2 complex, trace free matrix is a complex linear combination of $I,\ J$ and K so $\rho^+ \mid \bigwedge^2_+(\ ^4,\ \ )$ maps onto $\mathrm{End}_0(S^+)$. ∎

It follows, in particular, from Lemma A.2.4 that the map $\rho^+ \mid \bigwedge^2_+(\ ^4,\ \ ) : \bigwedge^2_+(\ ^4,\ \ ) \longrightarrow \mathrm{End}_0(S^+)$ has an inverse that we will simply denote

$$\sigma^+ : \mathrm{End}_0(S^+) \longrightarrow \bigwedge\nolimits^2_+(\ ^4,\ \ ). \tag{A.2.47}$$

One can compute this inverse explicitly, but we will content ourselves with describing its action on the particular type of trace free endomorphism that arises in the Seiberg-Witten equations. For this we consider an element

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix}$$

of S^+ (temporarily suppress the two zero components in S^+). Define an endomorphism of S^+ by the matrix

$$\psi \otimes \psi^* = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} (\bar{\psi}^1\ \bar{\psi}^2) = \begin{pmatrix} |\psi^1|^2 & \psi^1 \bar{\psi}^2 \\ \bar{\psi}^1\, \psi^2 & |\psi^2|^2 \end{pmatrix}. \tag{A.2.48}$$

The trace free part of this endomorphism is

$$\begin{aligned} (\psi \otimes \psi^*)_0 &= \psi \otimes \psi^* - \frac{1}{2}\mathrm{tr}(\psi \otimes \psi^*) \\ &= \begin{pmatrix} \frac{1}{2}(|\psi^1|^2 - |\psi^2|^2) & \psi^1\, \bar{\psi}^2 \\ \bar{\psi}^1\, \psi^2 & \frac{1}{2}(|\psi^2|^2 - |\psi^1|^2) \end{pmatrix}. \end{aligned} \tag{A.2.49}$$

One can verify directly that the image of $(\psi \otimes \psi^*)_0 \in \mathrm{End}_0(S^+)$ under σ^+ is

$$\begin{aligned} \sigma^+((\psi \otimes \psi^*)_0) &= -\frac{1}{4}\mathrm{i}\, \big[(|\psi^1|^2 - |\psi^2|^2)(e^1 \wedge e^2 + e^3 \wedge e^4) \\ &\qquad -2\,\mathrm{Im}(\psi^1\, \bar{\psi}^2)(e^1 \wedge e^3 + e^4 \wedge e^2) \\ &\qquad -2\,\mathrm{Re}(\psi^1\, \bar{\psi}^2)(e^1 \wedge e^4 + e^2 \wedge e^3)\big] \\ &= -\frac{1}{4}\, \big[(\psi^*\ I\ \psi)(e^1 \wedge e^2 + e^3 \wedge e^4) \\ &\qquad + (\psi^* J \psi)(e^1 \wedge e^3 + e^4 \wedge e^2) \\ &\qquad + (\psi^* K \psi)(e^1 \wedge e^4 + e^2 \wedge e^3)\big] \end{aligned} \tag{A.2.50}$$

(apply ρ^+ to the right-hand side to get $(\psi \otimes \psi^*)_0$).

Finally, we have the algebraic spadework completed and we can proceed to the problem of globalizing all of these notions to manifolds and bundles. Henceforth, B will denote a compact, connected, simply connected, oriented smooth 4-manifold (simple connectivity is not essential here, but will streamline some of what we have to say). For any choice of Riemannian metric $\boldsymbol{g}$ on B,

$$SO(4) \hookrightarrow F_{SO}(B) \stackrel{\pi_{SO}}{\longrightarrow} B$$

will denote the corresponding oriented, orthonormal frame bundle of B. Should B happen to admit a spin structure (see the Remark following the proof of Theorem A.2.3), then the representations $\Delta^{\pm} : \mathrm{Spin}(4) \longrightarrow SU(S^{\pm})$ give associated spinor bundles $S(B) \times_{\Delta^{\pm}} S^{\pm}$ whose sections are spinor fields. Because such a spin structure need not exist (and because spinor fields are essential ingredients in Seiberg-Witten theory), we formulate a complex analogue of a spin structure, which always exists.

A **spin**^{c} **structure** $\mathcal{L}$ on B consists of a principal $\mathrm{Spin}^c(4)$-bundle

$$\mathrm{Spin}^c(4) \hookrightarrow S^c(B) \stackrel{\pi_{S^c}}{\longrightarrow} B \tag{A.2.51}$$

over B and a smooth map

$$\Lambda : S^c(B) \longrightarrow F_{SO}(B) \tag{A.2.52}$$

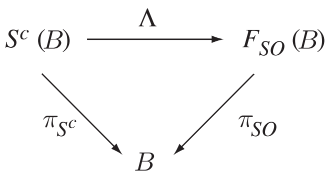
satisfying

$$\pi_{SO} \circ \Lambda = \pi_{S^c} \tag{A.2.53}$$

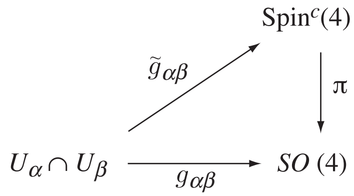
and

$$\Lambda(p \cdot \xi) = \Lambda(p) \cdot \pi(\xi) \tag{A.2.54}$$

for each $p \in S^c(B)$ and each $\xi \in \mathrm{Spin}^c(4)$. Here $\pi : \mathrm{Spin}^c(4) \longrightarrow SO(4)$ is defined by (A.2.38).



It is known that, for any B of the type we have described and any choice of the Riemannian metric $\boldsymbol{g}$, spin^c structures exist (see [**LM**]). In terms of transition functions this means that for any trivializing cover $\{U_\alpha\}$ for the frame bundle with transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \longrightarrow SO(4)$, there exist lifts $\tilde{g}_{\alpha\beta} : U_\alpha \cap U_\beta \longrightarrow \mathrm{Spin}^c(4)$



satisfying the cocycle condition.

Given a spin^c structure $\mathcal{L}$ on B each of the representations $\hat{\Delta}$: Spin$^c(4)$ $\longrightarrow$ Aut (S), $\hat{\Delta}^{\pm}$: Spin$^c(4) \longrightarrow U(S^{\pm})$ and δ : Spin$^c(4) \longrightarrow U(1)$ give rise to vector bundles associated to (A.2.51) which we will write as follows.

$$\mathcal{S}(\mathcal{L}) = S^c(B) \times_{\hat{\Delta}} S$$

$$\mathcal{S}^{\pm}(\mathcal{L}) = S^c(B) \times_{\hat{\Delta}^{\pm}} S^{\pm}$$

$$L(\mathcal{L}) = S^c(B) \times_{\delta}$$

$S(\mathcal{L})$ is called the **spinor bundle** of $\mathcal{L}, \mathcal{S}^{\pm}(\mathcal{L})$ are the **positive** and **negative spinor bundles** of $\mathcal{L}$ and $L(\mathcal{L})$ is the **determinant line bundle** of $\mathcal{L}$. The algebraic decomposition (A.2.31) persists in the bundle setting to give a Whitney sum decomposition

$$\mathcal{S}(\mathcal{L}) = \mathcal{S}^{+}(\mathcal{L}) \oplus \mathcal{S}^{-}(\mathcal{L}).$$

Remark: The Whitney sum of two vector bundles $\pi_i : E_i \longrightarrow X, i = 1, 2$, is just the natural vector bundle analogue of the direct sum of two vector spaces. Its fibers are, indeed, the direct sums $\pi_1^{-1}(x) \oplus \pi_2^{-1}(x)$ of the fibers of E_1 and E_2 and, if $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ is a trivializing cover for both E_1 and E_2, its transition functions are the direct sums of the transition functions for E_1 and E_2.

We will also need the principal $U(1)$-bundle

$$U(1) \hookrightarrow L^0(\mathcal{L}) \xrightarrow{\pi_{L^0}} B$$

associated to $L(\mathcal{L})$.

Remark: This bundle can be described as follows. Choose a Hermitian fiber metric (smoothly varying Hermitian inner products on the fibers) on the complex line bundle $L(\mathcal{L})$. Then $L^0(\mathcal{L})$ is the unit circle bundle in $L(\mathcal{L})$ i.e., it is the corresponding oriented, orthonormal frame bundle. One can retrieve $L(\mathcal{L})$ from $L^0(\mathcal{L})$ as the vector bundle associated to $L^0(\mathcal{L})$ by complex multiplication. One can show that $w_2(B) = c_1(L^0(\mathcal{L}))$ mod 2, where $w_2(B)$ is the second Stiefel-Whitney class of B, and that, conversely, given a $U(1)$-bundle L^0 over B with $w_2(B) = c_1(L^0)$ mod 2 there is a spin^c structure $\mathcal{L}$ on B with

$L^0(\mathcal{L}) = L^0$. More details on this and much of what follows are available in **[M1]**

We will also require a bundle associated to the frame bundle that does not require a spin or spin^c structure. Notice that Spin(4), being contained in $Cl^\times(4)$, acts on $Cl(4)$ by conjugation and, since $(-u)p(-u)^{-1} = upu^{-1}$, this gives an action of $SO(4) = \text{Spin}(4)/\mathbb{Z}_2$ on $Cl(4)$ which clearly preserves products $(u(pq)u^{-1} = (upu^{-1})(uqu^{-1}))$. The **Clifford bundle** $Cl(B)$ is the bundle with typical fiber $Cl(4)$ over B associated to the frame bundle by this action.

$$Cl(B) = F_{SO}(B) \times_{SO(4)} Cl(4)$$

Similarly, one has a **complexified Clifford bundle**

$$Cl(B) \otimes \quad = F_{SO}(B) \times_{SO(4)} (Cl(4) \otimes \quad).$$

These decompose into even and odd summands, e.g., $Cl(B) \cong Cl_0(B) \oplus Cl_1(B)$. Moreover, pointwise multiplication provides the spaces of sections of these bundles with algebra structures and such sections act on sections of the spinor bundle by pointwise Clifford multiplication.

Now, Spin$^c(4)$ double covers $SO(4) \times U(1)$ by the map Spin^c so $S^c(B)$ double covers the fiber product $F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ (this is just that part of the product bundle $SO(4) \times U(1) \hookrightarrow F_{SO}(B) \times L^0(\mathcal{L}) \hookrightarrow B \times B$ above the diagonal in $B \times B$ with this diagonal identified with B in the obvious way). We will use the symbol Spin^c also for this double cover.

$$\begin{array}{ccccc} \text{Spin}^c(4) & \hookrightarrow & S^c(B) & \longrightarrow & B \\ & & \big\downarrow \text{Spin}^c & & \\ SO(4) \times U(1) & \hookrightarrow & F_{SO}(B) \dot{\times} L^0(\mathcal{L}) & \longrightarrow & B \end{array} \tag{A.2.55}$$

Locally, this map is just

$$\begin{aligned} (b, \xi) \longrightarrow \big(b, \text{Spin}^c(\xi)\big) &= \big(b, (\pi(\xi), \delta(\xi))\big) \\ &= \big(b, \big(\text{Spin}(u), \mathrm{e}^{2\theta \mathrm{i}}\big)\big) \end{aligned} \tag{A.2.56}$$

where $\xi = \mathrm{e}^{\theta \mathrm{i}} u \in \text{Spin}^c(4)$.

Next observe that the algebraic isomorphism σ^+ in (A.2.47) globalizes as follows. The map ρ of (A.2.42) is independent of the choice of oriented, orthonormal basis for $\;^4$ so, using local oriented, orthonormal frame fields on B (i.e., sections of $F_{SO}(B)$) it gives a map from 2-forms on B to sections of $Cl(B) \otimes \;$. These sections in turn act on the spinor bundle $\mathcal{S}(\mathcal{L}) \cong \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L})$ of any spin^c structure. Since the action is fiberwise, the image of a self-dual 2-form preserves $\mathcal{S}^+(\mathcal{L})$ and is, at each point, a trace free endomorphism of $\mathcal{S}^+$. Thus, a self-dual 2-form on B gives rise to a section of the trace free endomorphism bundle $\text{End}_0(\mathcal{S}^+(\mathcal{L}))$ of $\mathcal{S}^+(\mathcal{L})$ and we have an isomorphism (also denoted σ^+) from the sections $\mathbf{\Gamma}(\mathrm{End}_0(\mathcal{S}^+(\mathcal{L})))$ of $\mathrm{End}_0(\mathcal{S}^+(\mathcal{L}))$ to the self-dual 2-forms on B:

$$\sigma^+ : \mathbf{\Gamma}\left(\mathrm{End}_0\left(\mathcal{S}^+(\mathcal{L})\right)\right) \longrightarrow \bigwedge\nolimits_+^2 (B, \quad) \tag{A.2.57}$$

In particular, if ψ is a positive spinor field on B (i.e., a section of $\mathcal{S}^+(\mathcal{L})$), then $(\psi \otimes \psi^*)_0 \in \mathbf{\Gamma}(\mathrm{End}_0(\mathcal{S}^+(\mathcal{L})))$ is given pointwise by (A.2.49) and $\sigma^+((\psi \otimes \psi *)_0) \in \bigwedge_+^2(B, \quad)$ is given pointwise relative to a local oriented, orthonormal frame field on B by (A.2.50).

A.3 Seiberg-Witten Equations

At this point we can begin to give some idea of where all of this is going. To write the Seiberg-Witten equations for B one chooses a Riemannian metric $\boldsymbol{g}$ and a spin^c structure $\mathcal{L}$. The field content of the theory consists of a $U(1)$-connection $\boldsymbol{A}$ on $L^0(\mathcal{L})$ (the gauge field) and a positive spinor field ψ. They are related by two equations, one of which requires that the self-dual 2-form $\sigma^+((\psi \otimes \psi^*)_0)$ should coincide with the self-dual part of the curvature of $\boldsymbol{A}$ i.e., $\boldsymbol{F}_{\boldsymbol{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0)$. The other equation still requires a bit of preparation, however.

Recall (Section 3.3) that the frame bundle $SO(4) \hookrightarrow F_{SO}(B) \longrightarrow B$ has a distinguished (Levi-Civita) connection which we will denote $\boldsymbol{\omega}_{LC}$. This can be characterized locally as follows. If $\{e_1, e_2, e_3, e_4\}$ is a local oriented, orthonormal frame field on B (i.e., a section of $F_{SO}(B)$) with dual 1-form field $\{e^1, e^2, e^3, e^4\}$, then $\boldsymbol{\omega}_{LC}$ is represented by a skew-symmetric matrix $(\boldsymbol{\omega}_j^i)$ of -valued 1-forms satisfying $\mathrm{d}e^i = -\boldsymbol{\omega}_j^i \wedge e^j$, $i = 1, 2, 3, 4$. Now notice that if B had a spin structure $\mathrm{Spin}(4) \hookrightarrow S(B) \longrightarrow B$, then the map $\lambda : S(B) \longrightarrow F_{SO}(B)$ is a double cover that respects the group actions so that any connection on $F_{SO}(B)$, e.g., $\boldsymbol{\omega}_{LC}$, automatically lifts to a connection on $S(B)$ (think of the connection as a distribution of horizontal spaces). However, if B has only a spin^c structure $\mathrm{Spin}^c(4) \hookrightarrow S^c(B) \longrightarrow B$, then the map Λ of (A.2.52) is not a finite covering so $\boldsymbol{\omega}_{LC}$ alone will not determine a connection on $S^c(B)$. However, $\mathrm{Spin}^c : S^c(B) \longrightarrow F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ is a double cover (A.2.55) and if $\boldsymbol{A}$ is any connection on $L^0(\mathcal{L})$, then $\boldsymbol{A}$ and $\boldsymbol{\omega}_{LC}$ together determine a connection on $F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ which will then lift to a connection on $S^c(B)$. Specifically, if pr_F and pr_{L^0} denote the restrictions to $F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ of the projections of $F_{SO}(B) \times L^0(\mathcal{L})$ onto $F_{SO}(B)$ and $L^0(\mathcal{L})$, respectively, then

$$\mathrm{pr}_F{}^* \boldsymbol{\omega}_{LC} \oplus \mathrm{pr}_{L^0}{}^* \boldsymbol{A}$$

is a connection on the fiber product and, identifying *spin*$^c(4)$ with the subset of $Cl(4) \otimes \quad$ given in (A.2.40),

$$\boldsymbol{\omega}_{\boldsymbol{A}} = (\mathrm{Spin}^c)^* \left(\mathrm{pr}_F{}^* \boldsymbol{\omega}_{LC} \oplus \mathrm{pr}_{L^0}{}^* \boldsymbol{A}\right) \tag{A.3.1}$$

is a connection on $S^c(B)$. Any such connection $\boldsymbol{\omega_A}$ is called a **spin**^{c} **connection** for $\mathcal{L}$ and with one of these we can introduce the basic differential operator of Seiberg-Witten theory.

Remark: The Levi-Civita connection $\boldsymbol{\omega}_{LC}$ on $F_{SO}(B)$ is to be regarded as fixed. The $U(1)$-connection $\boldsymbol{A}$ on $L^0(\mathcal{L})$, on the other hand, is the gauge field of Seiberg-Witten theory and will be constrained only by the field equations we eventually write down. The task of $\boldsymbol{A}$ is to produce, with $\boldsymbol{\omega}_{LC}$, a spin^c connection $\boldsymbol{\omega_A}$ on $S^c(B)$.

Now any connection $\boldsymbol{\omega}$ on any principal bundle $G \hookrightarrow P \longrightarrow X$ determines a differential operator $\nabla_{\boldsymbol{\omega}}$ (called a **covariant derivative**) on the space $\mathbf{\Gamma}(P \times_\rho \mathcal{V})$ of sections of any associated vector bundle. One simply recalls (Section 6.8 and Appendix B of [**N4**]) that such sections s correspond bijectively to $\mathcal{V}$-valued 0-forms φ on P that are tensorial of type ρ, that the covariant exterior derivative $d^{\boldsymbol{\omega}}\varphi$ is also tensorial of type ρ (Theorem 4.5.4) and so it, in turn, corresponds to a 1-form $\nabla_{\boldsymbol{\omega}} s$ on X with values in $P \times_\rho \mathcal{V}$. Regarding the 1-form $\nabla_{\boldsymbol{\omega}} s$ as acting on vector fields we can think of the covariant derivative $\nabla_{\boldsymbol{\omega}} s$ of the section $s \in \mathbf{\Gamma}(P \times_\rho \mathcal{V})$ as a 1-form with values in the sections $\mathbf{\Gamma}(P \times_\rho \mathcal{V})$ so

$$\nabla_{\boldsymbol{\omega}} : \mathbf{\Gamma}(P \times_\rho \mathcal{V}) \longrightarrow \bigwedge^1(X) \otimes \mathbf{\Gamma}(P \times_\rho \mathcal{V}).$$

A spin^c connection $\boldsymbol{\omega_A}$ on $\mathrm{Spin}^c(4) \hookrightarrow S^c(B) \xrightarrow{\pi_{S^c}} B$ induces covariant derivatives on the associated spinor bundles $\mathcal{S}(\mathcal{L})$, $\mathcal{S}^+(\mathcal{L})$ and $\mathcal{S}^-(\mathcal{L})$, all of which will be denoted $\nabla_{\boldsymbol{A}}$ when thought of as operating on sections of vector bundles and $\mathrm{d}_{\boldsymbol{A}}$ when operating on S -valued equivariant maps on $S^c(B)$. We will use

$$\nabla_{\boldsymbol{A}} : \mathbf{\Gamma}(\mathcal{S}(\mathcal{L})) \longrightarrow \bigwedge^1(B) \otimes \mathbf{\Gamma}(\mathcal{S}(\mathcal{L}))$$

to define the **Dirac operator**

$$\tilde{\not{D}}_{\boldsymbol{A}} : \mathbf{\Gamma}(\mathcal{S}(\mathcal{L})) \longrightarrow \mathbf{\Gamma}(\mathcal{S}(\mathcal{L}))$$

as follows. Let $\{e_1, e_2, e_3, e_4\}$ be a local oriented, orthonormal frame field on $U \subseteq B$ (i.e., a local section of $F_{SO}(B)$). Each e_i can be regarded as a vector field on U and also as a section of the Clifford bundle $Cl(B)$ which therefore acts by Clifford multiplication on sections of $\mathcal{S}(\mathcal{L})$ defined on U. Thus, for each $\Psi \in \mathbf{\Gamma}(\mathcal{S}(\mathcal{L}))$ we can define $\tilde{\not{D}}_{\boldsymbol{A}}\Psi$ on U by

$$\tilde{\not{D}}_{\boldsymbol{A}}\Psi = \sum_{i=1}^{4} e_i \cdot \nabla_{\boldsymbol{A}}\Psi(e_i). \tag{A.3.2}$$

One shows that this is independent of the choice of $\{e_1, e_2, e_3, e_4\}$ and so defines $\tilde{\not{D}}_{\boldsymbol{A}}\Psi$ globally. We will write out a concrete example shortly.

Since $\mathcal{S}(\mathcal{L}) = \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L})$ we may restrict $\tilde{\not{D}}_{\boldsymbol{A}}$ to sections of either $\mathcal{S}^+(\mathcal{L})$ or $\mathcal{S}^-(\mathcal{L})$. Since Clifford multiplication by e_i switches $\mathcal{S}^\pm(\mathcal{L})$, so will these restrictions. We will write these as

$$\not{D}_{\boldsymbol{A}} : \boldsymbol{\Gamma}\big(\mathcal{S}^{+}(\mathcal{L})\big) \longrightarrow \boldsymbol{\Gamma}\big(\mathcal{S}^{-}(\mathcal{L})\big) \tag{A.3.3}$$

and

$$\not{D}_{\boldsymbol{A}}^{\,*} : \boldsymbol{\Gamma}\big(\mathcal{S}^{-}(\mathcal{L})\big) \longrightarrow \boldsymbol{\Gamma}\big(\mathcal{S}^{+}(\mathcal{L})\big) \tag{A.3.4}$$

(these are, in fact, adjoints relative to the L^2 inner product on sections induced by the pointwise Hermitian inner product on fibers). We will also follow the custom in mathematics of referring to $\not{D}_{\boldsymbol{A}}$ also as a **Dirac operator**.

With this we can (at last) formulate the Seiberg-Witten equations. Thus, we let B denote a compact, connected, simply connected, oriented, smooth 4-manifold. Select a Riemannian metric $\boldsymbol{g}$ for B and then a spin^c structure $\mathcal{L}$ for the corresponding oriented, orthonormal frame bundle $F_{SO}(B)$. A pair $(\boldsymbol{A}, \psi)$ consisting of a $U(1)$-connection $\boldsymbol{A}$ on $U(1) \hookrightarrow L^0(B) \longrightarrow B$ and a positive spinor field $\psi \in \boldsymbol{\Gamma}(\mathcal{S}^{+}(\mathcal{L}))$ satisfies the **Seiberg-Witten (SW) equations** if

$$\not{D}_{\boldsymbol{A}}\psi = 0 \qquad (\textbf{Dirac Equation}) \tag{A.3.5}$$

and

$$\boldsymbol{F}_{\boldsymbol{A}}^{+} = \sigma^{+}((\psi \otimes \psi^{*})_0) \qquad (\textbf{Curvature Equation}), \tag{A.3.6}$$

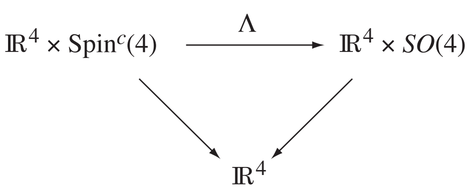
where $\boldsymbol{F}_{\boldsymbol{A}}^{+}$ is the $\boldsymbol{g}$-self-dual part of the curvature of $\boldsymbol{A}$.

Remark: The curvature of $\boldsymbol{A}$ is actually a $u(1)$-valued 2-form $\boldsymbol{\Omega}_{\boldsymbol{A}} = \mathrm{d}\boldsymbol{A}$ on $L^0(\mathcal{L})$, but, since $U(1)$ is Abelian, this projects to a $u(1)$-valued 2-form on B and this is what we mean by $\boldsymbol{F}_{\boldsymbol{A}}$.

To gain some sense of what these equations actually look like we will write them out explicitly on $\;^4$. More precisely, we consider $\;^4$ with its usual Riemannian metric and orientation. Since $\;^4$ is contractible, all of the relevant bundles over it are trivial and we will work with explicit trivializations. Thus, the oriented, orthonormal frame bundle is

$$SO(4) \hookrightarrow \;^4 \times SO(4) \longrightarrow \;^4$$

and there is an essentially unique spin^c structure



where $\Lambda(b,\xi) = (b, \pi(\xi))$ with π given by (A.2.38). The spinor bundles are therefore also trivial so their sections can be identified with globally defined functions on $\ ^4$ which we will write

$$\mathbf{\Psi} = \begin{pmatrix} \psi^1 \\ \psi^2 \\ \psi^3 \\ \psi^4 \end{pmatrix} : \ ^4 \longrightarrow S \ \cong \ ^4$$

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \\ 0 \\ 0 \end{pmatrix} : \ ^4 \longrightarrow S^+ \cong \ ^2$$

$$\phi = \begin{pmatrix} 0 \\ 0 \\ \psi^3 \\ \psi^4 \end{pmatrix} : \ ^4 \longrightarrow S^- \cong \ ^2.$$

For convenience we will often abuse the notation and write $\Psi = \binom{\psi}{\phi}$ by suppressing the zero components. We use x^1, x^2, x^3, x^4 for the standard coordinates on $\ ^4$ and write ∂_i for $\frac{\partial}{\partial x^i}, i = 1,2,3,4$ (these being applied componentwise to spinor fields).

The determinant line bundle is likewise trivial, as is the corresponding principal $U(1)$-bundle

$$U(1) \hookrightarrow \ ^4 \times U(1) \longrightarrow \ ^4.$$

A connection on this $U(1)$-bundle is then uniquely determined by a globally defined $u(1) = \mathrm{Im}\ $-valued 1-form on $\ ^4$:

$$\boldsymbol{A} = A_i \mathrm{d}x^i$$

$$A_i : \ ^4 \longrightarrow \mathrm{Im}\ , \qquad i = 1,2,3,4.$$

In orthonormal coordinates the covariant exterior derivative induced by the Levi-Civita connection is just ordinary (componentwise) exterior differentiation so the covariant derivative $\nabla_{\boldsymbol{A}}$ induced by it and the $U(1)$-connection $\boldsymbol{A}$ takes the form $\nabla_{\boldsymbol{A}} = \mathrm{d} + \boldsymbol{A}$, i.e.,

$$\nabla_{\boldsymbol{A}} = \nabla_i \mathrm{d}x^i = (\partial_i + A_i)\mathrm{d}x^i,$$

so that

$$\nabla_{\boldsymbol{A}}\Psi = (\partial_i \Psi + A_i \Psi)\mathrm{d}x^i$$

$$= \begin{pmatrix} (\partial_i\psi^1 + A_i\psi^1)\mathrm{d}x^i \\ (\partial_i\psi^2 + A_i\psi^2)\mathrm{d}x^i \\ (\partial_i\psi^3 + A_i\psi^3)\mathrm{d}x^i \\ (\partial_i\psi^4 + A_i\psi^4)\mathrm{d}x^i \end{pmatrix}.$$

Thus, with $\{e_i\} = \{\partial_i\}$ the standard oriented, orthonormal frame field on ^{4}, we have $\nabla_{\mathbf{A}}\Psi(e_i) = \partial_i\Psi + A_i\Psi$ and, for convenience, we will write this as

$$\nabla_i\Psi = (\partial_i + A_i)\Psi = \begin{pmatrix} \partial_i\psi^1 + A_i\psi^1 \\ \partial_i\psi^2 + A_i\psi^2 \\ \partial_i\psi^3 + A_i\psi^3 \\ \partial_i\psi^4 + A_i\psi^4 \end{pmatrix}.$$

The Dirac operator $\tilde{\not{D}}_{\boldsymbol{A}}\Psi = \sum_{i=1}^4 e_i \cdot \nabla_i\Psi$ requires that we Clifford multiply by the basis elements e_i, i.e., matrix multiply by $E_i = \Gamma(e_i) \in Cl(4) \otimes$ as in (A.2.25). For this we write $\Psi = \binom{\psi}{\phi}$ so that

$$\begin{aligned} \tilde{\not{D}}_{\boldsymbol{A}}\Psi &= \sum_{i=1}^4 e_i \cdot \nabla_i\Psi = \sum_{i=1}^4 E_i\nabla_i\Psi \\ &= \begin{pmatrix} \nabla_1\phi + I\nabla_2\phi + J\nabla_3\phi + K\nabla_4\phi \\ -\nabla_1\psi + I\nabla_2\psi + J\nabla_3\psi + K\nabla_4\psi \end{pmatrix}. \end{aligned}$$

Note that, as expected, $\tilde{\not{D}}_{\boldsymbol{A}}$ sends positive spinor fields $\Psi = \binom{\psi}{0}$ to negative spinor fields and vice versa. The restriction of $\tilde{\not{D}}_{\boldsymbol{A}}$ to positive spinor fields will be written

$$\not{D}_{\boldsymbol{A}}\psi = -\nabla_1\psi + I\nabla_2\psi + J\nabla_3\psi + K\nabla_4\psi \tag{A.3.7}$$

by suppressing the zero components (but now one must remember that $\not{D}_{\boldsymbol{A}}\psi$ is a negative spinor field). The first Seiberg-Witten equation (A.3.5) then becomes

$$\nabla_1\psi = I\nabla_2\psi + J\nabla_3\psi + K\nabla_4\psi \tag{A.3.8}$$

or, in complete detail,

$$\begin{pmatrix} -(\partial_1 + A_1) + \mathrm{i}(\partial_2 + A_2) & (\partial_3 + A_3) + \mathrm{i}(\partial_4 + A_4) \\ -(\partial_3 + A_3) + \mathrm{i}(\partial_4 + A_4) & -(\partial_1 + A_1) - \mathrm{i}(\partial_2 + A_2) \end{pmatrix} \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{A.3.9}$$

This part is linear, of course.

For the second Seiberg-Witten equation (A.3.6) we use the (pointwise) expressions (A.2.50) for $\sigma^+((\psi \otimes \psi^*)_0)$ and the following local description of $\boldsymbol{F}_{\boldsymbol{A}}^+$. With $\boldsymbol{A} = A_i\mathrm{d}x^i$ we have $\boldsymbol{F}_{\boldsymbol{A}} = \mathrm{d}\boldsymbol{A} = \sum_{i<j} F_{ij}\mathrm{d}x^i \wedge \mathrm{d}x^j$, where $F_{ij} = \partial_i A_j - \partial_j A_i$. A basis for the self-dual 2-forms is given by $\{\mathrm{d}x^1 \wedge \mathrm{d}x^2 + \mathrm{d}x^3 \wedge \mathrm{d}x^4, \mathrm{d}x^1 \wedge \mathrm{d}x^3 + \mathrm{d}x^4 \wedge \mathrm{d}x^2, \mathrm{d}x^1 \wedge \mathrm{d}x^4 + \mathrm{d}x^2 \wedge \mathrm{d}x^3\}$ so

$$\boldsymbol{F}_{\boldsymbol{A}}^{+} = \frac{1}{2}(\boldsymbol{F}_{\boldsymbol{A}} + {}^{*}\boldsymbol{F}_{\boldsymbol{A}}) = \frac{1}{2}(F_{12}+F_{34})(\mathrm{d}x^1 \wedge \mathrm{d}x^2 + \mathrm{d}x^3 \wedge \mathrm{d}x^4) + \frac{1}{2}(F_{13}+F_{42})(\mathrm{d}x^1 \wedge \mathrm{d}x^3 + \mathrm{d}x^4 \wedge \mathrm{d}x^2) + \frac{1}{2}(F_{14}+F_{23})(\mathrm{d}x^1 \wedge \mathrm{d}x^4 + \mathrm{d}x^2 \wedge \mathrm{d}x^3).$$

Thus, (A.3.6) becomes

$$\begin{aligned} F_{12}+F_{34} &= -\frac{1}{2}\psi^* I\psi \\ F_{13}+F_{42} &= -\frac{1}{2}\psi^* J\psi \\ F_{14}+F_{23} &= -\frac{1}{2}\psi^* K\psi \end{aligned} \tag{A.3.10}$$

or, in more detail,

$$\begin{aligned} (\partial_1 A_2 - \partial_2 A_1) + (\partial_3 A_4 - \partial_4 A_3) &= -\frac{1}{2}\mathrm{i}\left(|\psi^1|^2 - |\psi^2|^2\right) \\ (\partial_1 A_3 - \partial_3 A_1) + (\partial_4 A_2 - \partial_2 A_4) &= -\mathrm{i}\ \mathrm{Re}\left(\bar{\psi}^1\psi^2\right) \\ (\partial_1 A_4 - \partial_4 A_1) + (\partial_2 A_3 - \partial_3 A_2) &= -\mathrm{i}\ \mathrm{Im}\left(\bar{\psi}^1\psi^2\right). \end{aligned} \tag{A.3.11}$$

Note that these are only rather mildly nonlinear.

Remark: It is perhaps worth pointing out that these equations do have nontrivial solutions. We will produce some solutions with $\psi = 0$ (these will turn out to be reducible solutions with respect to the gauge action to be introduced shortly). When $\psi = 0$ the Dirac equation is satisfied identically and the curvature equation reduces to $d\boldsymbol{F}_{\boldsymbol{A}}^{+} = 0$ so what we need is an anti-self-dual $U(1)$-connection, i.e., an $\boldsymbol{A} = A^i dx^i$ for which $d\boldsymbol{A}$ is of the form

$$\boldsymbol{F} = F_1\left(dx^1 \wedge dx^2 - dx^3 \wedge dx^4\right) + F_2\left(dx^1 \wedge dx^3 - dx^4 \wedge dx^2\right) + F_3\left(dx^1 \wedge dx^4 - dx^2 \wedge dx^3\right).$$

What we will do is seek functions F_1, F_2, F_3 that are independent of x^4 and for which this 2-form is closed. The Poincaré Lemma then implies that it is exact so it must be $d\boldsymbol{A}$ for some $\boldsymbol{A}$ and this $\boldsymbol{A}$ (together with $\psi = 0$) gives our solution. Now,

$$\begin{aligned} & d(F_1(dx^1 \wedge dx^2 - dx^3 \wedge dx^4)) = dF_1 \wedge (dx^1 \wedge dx^2 - dx^3 \wedge dx^4) \\ &= (\partial_1 F_1 dx^1 + \partial_2 F_1 dx^2 + \partial_3 F_1 dx^3) \wedge (dx^1 \wedge dx^2 - dx^3 \wedge dx^4) \\ &= -(\partial_1 F_1) dx^1 \wedge dx^3 \wedge dx^4 - (\partial_2 F_1) dx^2 \wedge dx^3 \wedge dx^4 + (\partial_3 F_1) dx^1 \wedge dx^2 \wedge dx^3. \end{aligned}$$

Computing the remaining terms similarly one finds that the coefficient of $dx^1 \wedge dx^3 \wedge dx^4$ is $-(\partial_1 F_1 + \partial_2 F_2 + \partial_3 F_3)$ and this is just (minus) the divergence of the map $F = (F_1, F_2, F_3)$. The remaining coefficients are just the components of the curl of this same map F. Consequently, $\boldsymbol{F}$ will be closed if we choose F to be a map with $\operatorname{div} F = 0$ and curl $F = 0$ and these are a dime-a-dozen (think about the static, source-free Maxwell equations, or just make some up).

We return now to the general development so that B is a compact, connected, simply connected, oriented, smooth 4-manifold. Choosing a Riemannian metric $\boldsymbol{g}$ on B gives an oriented, orthonormal frame bundle and one can then select a spin^c structure $\mathcal{L}$. The corresponding **Seiberg-Witten configuration space** $\mathcal{A}(\mathcal{L})$ consists of all pairs $(\boldsymbol{A}, \psi)$, where $\boldsymbol{A}$ is a connection on the principal $U(1)$-bundle $L^0(\mathcal{L})$ and $\psi \in \boldsymbol{\Gamma}(\mathcal{S}^+(\mathcal{L}))$ is a positive spinor field. An $(\boldsymbol{A}, \psi) \in \mathcal{A}(\mathcal{L})$ is a **Seiberg-Witten monopole** (**SW monopole**) if it satisfies (A.3.5) and (A.3.6) (these two equations together will henceforth be denoted simply (SW)).

A.4 The Moduli Space and Invariant

As in the case of Donaldson theory, our real interest is in a moduli space of SW monopoles so we begin by isolating the appropriate gauge group. This will be a subgroup of the group of automorphisms of the spin^c bundle (diffeomorphisms σ of $S^c(B)$ onto itself satisfying $\sigma(p \cdot \xi) = \sigma(p) \cdot \xi$ for each $p \in S^c(B)$ and each $\xi \in \operatorname{Spin}^c(4)$ and $\pi_{S^c} \circ \sigma = \pi_{S^c}$).

Recall that $S^c(B)$ double covers the fiber product $F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ via the map Spin^c. Letting pr_F and pr_{L^0} be the projections of $F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ onto $F_{SO}(B)$ and $L^0(\mathcal{L})$ we obtain maps

$$\begin{array}{ccccc} \operatorname{Spin}^c(4) & \hookrightarrow & S^c(B) & \longrightarrow & B \\ & & \Big\downarrow \mathrm{pr}_F \circ \operatorname{Spin}^c & & \\ SO(4) & \hookrightarrow & F_{SO}(B) & \longrightarrow & B \end{array} \tag{A.4.1}$$

and

$$\begin{array}{ccccc} \operatorname{Spin}^c(4) & \hookrightarrow & S^c(B) & \longrightarrow & B \\ & & \Big\downarrow \mathrm{pr}_{L^0} \circ \operatorname{Spin}^c & & \\ U(1) & \hookrightarrow & L^0(\mathcal{L}) & \longrightarrow & B \end{array} \tag{A.4.2}$$

We will say that an automorphism $\sigma : S^c(B) \longrightarrow S^c(B)$ **covers the identity on** $F_{SO}(B)$ if

$$\mathrm{pr}_F \circ \operatorname{Spin}^c \circ \sigma = \mathrm{pr}_F \circ \operatorname{Spin}^c. \tag{A.4.3}$$

The collection of all such is a group $\mathcal{G}(\mathcal{L})$ under composition which we call the (**Seiberg-Witten**) **gauge group** and which we will show acts naturally on the solutions to (SW).

Lemma A.4.1 *If $\gamma \in C^\infty(B, U(1))$ is any smooth map of B into $U(1) \subseteq Spin^c(4)$, then the map*

$$\sigma_\gamma : S^c(B) \longrightarrow S^c(B)$$
$$\sigma_\gamma(p) = p \cdot \gamma(\pi_{S^c}(p))$$

is an automorphism of $S^c(B)$ that covers the identity on $F_{SO}(B)$. Conversely, every element of $\mathcal{G}(\mathcal{L})$ is σ_γ for some $\gamma \in C^\infty(B, U(1))$ and

$$\mathcal{G}(\mathcal{L}) \cong C^\infty(B, U(1))$$

where the group operation in $C^\infty(B, U(1))$ is pointwise multiplication in $U(1)$.

Proof: First note that, because $U(1) = Z(\mathrm{Spin}^c(4)), \sigma_\gamma(p \cdot \xi) = (p \cdot \xi) \cdot \gamma(\pi_{S^c}(p \cdot \xi)) = (p \cdot \xi) \cdot \gamma(\pi_{S^c}(p)) = (p \cdot \gamma(\pi_{S^c}(p))) \cdot \xi = \sigma_\gamma(p) \cdot \xi$ so σ_γ is a bundle map. Since p and $p \cdot \gamma(\pi_{S^c}(p))$ are in the same fiber of π_{S^c}, σ_γ covers the identity on B so σ_γ is an automorphism. It covers the identity on $F_{SO}(B)$ as well because $\mathrm{pr}_F \circ \mathrm{Spin}^c$ is locally given by $(b, \xi) = (b, \mathrm{e}^{\theta \mathrm{i}} u) \longrightarrow (b, \text{ Spin } (u))$ so (b, ξ) and $(b, \xi) \cdot \mathrm{e}^{\phi \mathrm{i}} = (b, \mathrm{e}^{(\theta + \phi)\mathrm{i}} u)$ have the same image, i.e., $\mathrm{pr}_F \circ \mathrm{Spin}^c \circ \sigma_\gamma = \mathrm{pr}_F \circ \mathrm{Spin}^c$.

For the converse, it is easy to verify that $p_1, p_2 \in S^c(B)$ have the same image under $\mathrm{pr}_F \circ \mathrm{Spin}^c$ if and only if they differ by the action of something in $U(1)$, i.e., $p_2 = p_1 \cdot \mathrm{e}^{\phi \mathrm{i}}$ for some $\phi \in$. Thus, an automorphism $\sigma : S^c(B) \longrightarrow S^c(B)$ that covers the identity on $F_{SO}(B)$ must satisfy $\sigma(p) = p \cdot$ (something in $U(1)$) for each $p \in S^c(B)$. We claim that this "something" must be the same for all points in the same fiber of π_{S^c}. Indeed, $\pi_{S^c}(p_1) = \pi_{S^c}(p_2)$ implies $p_2 = p_1 \cdot \xi$ for some $\xi \in \mathrm{Spin}^c(4)$ and if $\sigma(p_1) = p_1 \cdot \mathrm{e}^{\phi \mathrm{i}}$, then $\sigma(p_2) = \sigma(p_1 \cdot \xi) = \sigma(p_1) \cdot \xi = (p_1 \cdot \mathrm{e}^{\phi \mathrm{i}}) \cdot \xi = (p_1 \cdot \xi) \cdot \mathrm{e}^{\phi \mathrm{i}} = p_2 \cdot \mathrm{e}^{\phi \mathrm{i}}$ also. Thus, $\sigma(p) = p \cdot \gamma(\pi_{S^c}(p))$ for some $\gamma \in C^\infty(B, U(1))$ as required and the rest is clear. ∎

We will use whichever view of the gauge group $\mathcal{G}(\mathcal{L})$ is most convenient in any particular situation.

Our goal now is to show that $\mathcal{G}(\mathcal{L})$ acts naturally on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ and, indeed, preserves the set of solutions to (SW). We begin by defining the action of $\mathcal{G}(\mathcal{L})$ on positive spinor fields $\psi \in \mathbf{\Gamma}(S^+(\mathcal{L}))$. For this we identify ψ with an equivariant S^+ -valued map on $S^c(B)$ and define the action of $\sigma_\gamma \in \mathcal{G}(\mathcal{L})$ by pullback, i.e.,

$$\psi \cdot \sigma_\gamma = \psi \cdot \gamma = {\sigma_\gamma}^* \psi = \psi \circ \sigma_\gamma. \tag{A.4.4}$$

Thus, at each $p \in S^c(B)$,

$$\begin{aligned}\big(\psi\cdot\sigma_\gamma\big)(p) &= \psi\big(\sigma_\gamma(p)\big) = \psi\big(p\cdot\gamma(\pi_{S^c}(p))\big)\\ &= \big(\gamma(\pi_{S^c}(p))\big)^{-1}\psi(p).\end{aligned}\tag{A.4.5}$$

Thus, if we think instead of ψ as a section of $\mathcal{S}^+(\mathcal{L})$ we have

$$\psi\cdot\gamma = (\gamma\circ\pi_{S^c})^{-1}\psi. \tag{A.4.6}$$

The same formulas define the action of $\mathcal{G}(\mathcal{L})$ on negative spinor fields.

Turning next to the connection $\boldsymbol{A}$ on $U(1)\hookrightarrow L^0(\mathcal{L})\longrightarrow B$ we note that the automorphism σ_γ of $S^c(B)$ induces an automorphism ${\sigma'}_\gamma$ of $L^0(\mathcal{L})$ as follows:

$$\begin{array}{ccc} S^c(B) & \xrightarrow{\ \sigma_\gamma\ } & S^c(B)\\ {\scriptstyle \mathrm{pr}_{L^0}\circ\,\mathrm{Spin}^c}\downarrow & & \downarrow{\scriptstyle \mathrm{pr}_{L^0}\circ\,\mathrm{Spin}^c}\\ L^0(\mathcal{L}) & \xrightarrow[\ {\sigma'}_\gamma\ ]{} & L^0(\mathcal{L}) \end{array}$$

$${\sigma'}_\gamma\circ\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c = \mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\circ\sigma_\gamma \tag{A.4.7}$$

(we will write out an explicit local expression for ${\sigma'}_\gamma$ shortly). Now we define the action of $\sigma_\gamma\in\mathcal{G}(\mathcal{L})$ on $\boldsymbol{A}$ by

$$\boldsymbol{A}\cdot\sigma_\gamma = \boldsymbol{A}\cdot\gamma = ({\sigma'}_\gamma)^*\boldsymbol{A}. \tag{A.4.8}$$

It will be convenient to have (A.4.8) expressed locally in terms of gauge potentials. Thus, we let s be a local section of $L^0(\mathcal{L})$ and write $\boldsymbol{\mathcal{A}} = s^*\boldsymbol{A}$. Also define

$$\boldsymbol{\mathcal{A}}\cdot\gamma = s^*(\boldsymbol{A}\cdot\gamma) = s^*\Big(({\sigma'}_\gamma)^*\boldsymbol{A}\Big) = ({\sigma'}_\gamma\circ s)^*\boldsymbol{A}. \tag{A.4.9}$$

Now, since $\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c$ is locally given by

$$(b,\xi) = \big(b,\mathrm{e}^{\theta\mathrm{i}}u\big)\longrightarrow\big(b,\delta\left(\mathrm{e}^{\theta\mathrm{i}}u\right)\big) = \big(b,\mathrm{e}^{2\theta\mathrm{i}}\big)$$

it satisfies

$$\big(\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\big)(p\cdot\xi_0) = \big(\big(\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\big)(p)\big)\cdot\delta(\xi_0) \tag{A.4.10}$$

so

$$\begin{aligned}{\sigma'}_\gamma\big(\big(\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\big)(p)\big) &= \big(\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\big)\big(\sigma_\gamma(p)\big)\\ &= \big(\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\big)\big(p\cdot\gamma(\pi_{S^c}(p))\big)\\ &= \big(\big(\mathrm{pr}_{L^0}\circ\mathrm{Spin}^c\big)(p)\big)\cdot\delta\big(\gamma(\pi_{S^c}(p))\big).\end{aligned}$$

Thus, we may write

$$\sigma'_\gamma(x) = x \cdot \delta\big(\gamma\big(\pi_{L^0}(x)\big)\big) = x \cdot \big(\gamma\big(\pi_{L^0}(x)\big)\big)^2 \tag{A.4.11}$$

for each $x \in L^0(\mathcal{L})$. In particular,

$$\big(\sigma'_\gamma \circ s\big)(b) = s(b) \cdot (\gamma(b))^2. \tag{A.4.12}$$

It then follows from (A.4.8) that

$$\boldsymbol{\mathcal{A}} \cdot \gamma = (\gamma^2)^{-1}\boldsymbol{\mathcal{A}}(\gamma^2) + (\gamma^2)^{-1}\mathrm{d}(\gamma^2) = \boldsymbol{\mathcal{A}} + (\gamma^2)^{-1}(2\gamma \mathrm{d}\gamma)$$

and therefore

$$\boldsymbol{\mathcal{A}} \cdot \gamma = \boldsymbol{\mathcal{A}} + 2\gamma^{-1}\mathrm{d}\gamma. \tag{A.4.13}$$

Applying $\pi_{L^0}^*$ to both sides of (A.4.13) gives

$$\boldsymbol{A} \cdot \gamma = \boldsymbol{A} + \pi_{L^0}^*(2\gamma^{-1}\mathrm{d}\gamma) = \boldsymbol{A} + 2(\gamma \circ \pi_{L^0})^{-1}\mathrm{d}(\gamma \circ \pi_{L^0}). \tag{A.4.14}$$

We now have an action of the group $\mathcal{G}(\mathcal{L})$ on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ given by

$$\begin{aligned}(\boldsymbol{A}, \psi) \cdot \sigma_\gamma = (\boldsymbol{A}, \psi) \cdot \gamma &= \Big((\sigma'_\gamma)^* \boldsymbol{A}, {\sigma_\gamma}^* \psi\Big) \\ &= \Big(\boldsymbol{A} + 2(\gamma \circ \pi_{L^0})^{-1}\, \mathrm{d}(\gamma \circ \pi_{L^0}), (\gamma \circ \pi_{S^c})^{-1}\psi\Big)\end{aligned}$$

or, locally on B,

$$(\boldsymbol{\mathcal{A}}, \psi) \cdot \gamma = (\boldsymbol{\mathcal{A}} + 2\gamma^{-1}\mathrm{d}\gamma, \gamma^{-1}\psi). \tag{A.4.15}$$

In order to show that this action preserves solutions to (SW) we first observe that the spin^c connection corresponding to $\boldsymbol{A} \cdot \gamma$ is the pullback by σ_γ of that corresponding to $\boldsymbol{A}$, i.e.,

$$\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma} = {\sigma_\gamma}^* \boldsymbol{\omega}_{\boldsymbol{A}}. \tag{A.4.16}$$

To see this we note, from (A.3.1), that

$${\sigma_\gamma}^* \boldsymbol{\omega}_{\boldsymbol{A}} = (\mathrm{Spin}^c \circ \sigma_\gamma)^*({\mathrm{pr}_F}^* \boldsymbol{\omega}_{LC} + \mathrm{pr}_{L^0}^* \boldsymbol{A})$$

and

$$\begin{aligned}\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma} &= (\mathrm{Spin}^c)^*\Big({\mathrm{pr}_F}^* \boldsymbol{\omega}_{LC} + \mathrm{pr}_{L^0}^*\Big(({\sigma_\gamma}')^* \boldsymbol{A}\Big)\Big) \\ &= (\mathrm{pr}_F \circ \mathrm{Spin}^c)^* \boldsymbol{\omega}_{LC} + ({\sigma_\gamma}' \circ \mathrm{pr}_{L^0} \circ \mathrm{Spin}^c)^* \boldsymbol{A} \\ &= (\mathrm{pr}_F \circ \mathrm{Spin}^c \circ \sigma_\gamma)^* \boldsymbol{\omega}_{LC} + (\mathrm{pr}_{L^0} \circ \mathrm{Spin}^c \circ \sigma_\gamma)^* \boldsymbol{A}\end{aligned}$$

$$\begin{aligned}&\text{(by (A.4.3) and (A.4.7))}\\&= (\mathrm{Spin}^c \circ \sigma_\gamma)^*(\mathrm{pr}_F{}^*\boldsymbol{\omega}_{LC} + \mathrm{pr}_{L^0}^*\boldsymbol{A})\\&= \sigma_\gamma{}^*\boldsymbol{\omega}_{\boldsymbol{A}}.\end{aligned}$$

Another computation of $\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma}$ using (A.4.14) gives

$$\begin{aligned}\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma} &= (\mathrm{Spin}^c)^*\left(\mathrm{pr}_F^*\,\boldsymbol{\omega}_{LC} + \mathrm{pr}_{L^0}^*\left(\boldsymbol{A} + \pi_{L^0}^*\left(2\gamma^{-1}\mathrm{d}\gamma\right)\right)\right)\\&= \boldsymbol{\omega}_{\boldsymbol{A}} + (\mathrm{Spin}^c)^*\left(\mathrm{pr}_{L^0}^*\left(\pi_{L^0}^*\left(2\gamma^{-1}\mathrm{d}\gamma\right)\right)\right)\\&= \boldsymbol{\omega}_{\boldsymbol{A}} + \left(\pi_{L^0} \circ \mathrm{pr}_{L^0} \circ \mathrm{Spin}^c\right)^*\left(2\gamma^{-1}\mathrm{d}\gamma\right)\\&= \boldsymbol{\omega}_{\boldsymbol{A}} + \pi_{S^c}^*\left(2\gamma^{-1}\mathrm{d}\gamma\right)\\\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma} &= \boldsymbol{\omega}_{\boldsymbol{A}} + 2\left(\gamma \circ \pi_{S^c}\right)^{-1}\mathrm{d}(\gamma \circ \pi_{S^c}\Big).\end{aligned} \tag{A.4.17}$$

Theorem A.4.2 *The action of $\mathcal{G}(\mathcal{L})$ on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ carries solutions to (SW) onto other solutions to (SW). More precisely, if $(\boldsymbol{A},\ \psi) \in \mathcal{A}(\mathcal{L})$ satisfies*

$$\begin{cases} \not{D}_{\boldsymbol{A}}\psi = 0 \\ \boldsymbol{F}_{\boldsymbol{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0) \end{cases}$$

then for any $\sigma_\gamma \in \mathcal{G}(\mathcal{L})$, $(\boldsymbol{A},\ \psi)\cdot\gamma = (\boldsymbol{A}\cdot\gamma,\ \psi\cdot\gamma)$ satisfies

$$\begin{cases} \not{D}_{\boldsymbol{A}\cdot\gamma}(\psi\cdot\gamma) = 0 \\ \boldsymbol{F}_{\boldsymbol{A}\cdot\gamma}^+ = \sigma^+(((\psi\cdot\gamma) \otimes (\psi\cdot\gamma)^*)_0) \end{cases}.$$

Proof: For the curvature equation we observe that (A.4.16), the usual transformation equation for the curvature and the fact that $U(1)$ is Abelian imply that

$$\boldsymbol{F}_{\boldsymbol{A}\cdot\gamma}^+ = \gamma^2\boldsymbol{F}_{\boldsymbol{A}}^+(\gamma^2)^{-1} = \boldsymbol{F}_{\boldsymbol{A}}^+.$$

Similarly, the commutativity of $U(1)$ gives

$$\begin{aligned}(\psi\cdot\gamma)\otimes(\psi\cdot\gamma)^* &= (\gamma^{-1}\psi)\otimes(\gamma^{-1}\psi)^*\\&= (\gamma^{-1}\psi)\otimes(\gamma\,\psi^*)\\&= (\gamma^{-1}\gamma)(\psi\otimes\psi^*)\\&= \psi \otimes \psi^*.\end{aligned}$$

Thus, $\boldsymbol{F}_{\boldsymbol{A}}^+ = \sigma^+((\psi\otimes\psi^*)_0)$ implies $\boldsymbol{F}_{\boldsymbol{A}\cdot\gamma}^+ = \sigma^+(((\psi\cdot\gamma)\otimes(\psi\cdot\gamma)^*)_0)$. To verify the analogous statement for the Dirac equation it will surely be enough to show that

$$\not{D}_{\boldsymbol{A}\cdot\gamma}(\psi\cdot\gamma) = (\not{D}_{\boldsymbol{A}}\psi)\cdot\gamma\,. \tag{A.4.18}$$

For this it will be convenient to identify ψ with an equivariant S^+-valued map on $S^c(B)$ and compare the covariant exterior derivatives $\mathrm{d}_{\boldsymbol{A}}\psi$ and $\mathrm{d}_{\boldsymbol{A}\cdot\gamma}(\psi\cdot\gamma)$. There are standard formulas for such derivatives (see, for example, (6.8.4) of [**N4**]) which, in our case, give

$$\mathrm{d}_{\boldsymbol{A}}\psi = \mathrm{d}\psi + \frac{1}{2}\,\boldsymbol{\omega}_{\boldsymbol{A}}\psi$$

and

$$\mathrm{d}_{\boldsymbol{A}\cdot\gamma}(\psi\cdot\gamma) = \mathrm{d}(\psi\cdot\gamma) + \frac{1}{2}\,\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma}(\psi\cdot\gamma)\,,$$

where, e.g., $\boldsymbol{\omega}_{\boldsymbol{A}}$ takes values in $spin^c(4)$, identified with a subset of $Cl(4)\otimes$ as in (A.2.40), and $\boldsymbol{\omega}_{\boldsymbol{A}}\psi$ is a matrix product.

Remark: The factor $\frac{1}{2}$ arises from the actual identification of $spin^c(4)$ with $so(4)\oplus u(1)$ via the derivative at the identity of the double cover map $\mathrm{Spin}^c : \mathrm{Spin}^c(4) \longrightarrow SO(4)\times U(1)$.

Now we compute

$$\begin{aligned}
\mathrm{d}_{\boldsymbol{A}\cdot\gamma}(\psi\cdot\gamma) &= \mathrm{d}_{\boldsymbol{A}\cdot\gamma}\Big((\gamma\circ\pi_{S^c})^{-1}\,\psi\Big)\\
&= \mathrm{d}\Big((\gamma\circ\pi_{S^c})^{-1}\,\psi\Big) + \frac{1}{2}\,\boldsymbol{\omega}_{\boldsymbol{A}\cdot\gamma}\Big((\gamma\circ\pi_{S^c})^{-1}\,\psi\Big)\\
&= (\gamma\circ\pi_{S^c})^{-1}\,\mathrm{d}\,\psi - (\gamma\circ\pi_{S^c})^{-2}\,\mathrm{d}\,(\gamma\circ\pi_{S^c})\,\psi\\
&\qquad + \frac{1}{2}\Big(\boldsymbol{\omega}_{\boldsymbol{A}} + 2(\gamma\circ\pi_{S^c})^{-1}\,\mathrm{d}(\gamma\circ\pi_{S^c})\Big)\Big((\gamma\circ\pi_{S^c})^{-1}\,\psi\Big)\\
&= (\gamma\circ\pi_{S^c})^{-1}\Big(\mathrm{d}\,\psi + \frac{1}{2}\,\boldsymbol{\omega}_{\boldsymbol{A}}\,\psi\Big) - (\gamma\circ\pi_{S^c})^{-2}\,\mathrm{d}\,(\gamma\circ\pi_{S^c})\,\psi\\
&\qquad + (\gamma\circ\pi_{S^c})^{-2}\,\mathrm{d}\,(\gamma\circ\pi_{S^c})\,\psi\\
&= (\gamma\circ\pi_{S^c})^{-1}\,(\mathrm{d}_{\boldsymbol{A}}\,\psi)
\end{aligned}$$

and from this conclude that

$$\boldsymbol{\nabla}_{\boldsymbol{A}\cdot\gamma}\,(\psi\cdot\gamma) = (\gamma\circ\pi_{S^c})^{-1}\,\boldsymbol{\nabla}_{\boldsymbol{A}}\,\psi. \tag{A.4.19}$$

Finally,

$$\begin{aligned}
\not{D}_{\boldsymbol{A}\cdot\gamma}\,(\psi\cdot\gamma) &= \sum_{i=1}^{4} e_i\cdot\boldsymbol{\nabla}_{\boldsymbol{A}\cdot\gamma}\,(\psi\cdot\gamma)(e_i)\\
&= \sum_{i=1}^{4} e_i\cdot(\psi\circ\pi_{S^c})^{-1}\,\boldsymbol{\nabla}_{\boldsymbol{A}}\,\psi\,(e_i)\\
&= (\psi\circ\pi_{S^c})^{-1}\sum_{i=1}^{4} e_i\cdot\boldsymbol{\nabla}_{\boldsymbol{A}}\,\psi\,(e_i)\\
&= (\psi\circ\pi_{S^c})^{-1}\,\not{D}_{\boldsymbol{A}}\,\psi\\
&= (\not{D}_{\boldsymbol{A}}\,\psi)\cdot\gamma
\end{aligned}$$

which proves (A.4.18) and therefore Theorem 6. ■

Thus, the space of solutions to the Seiberg-Witten equations is invariant under the action of the gauge group $\mathcal{G}(\mathcal{L})$ and we may, as for the anti-self-dual equations, consider the **moduli space** $\mathcal{M}(\mathcal{L})$ of gauge equivalence classes of Seiberg-Witten monopoles:

$$\mathcal{M}(\mathcal{L}) = \{(\boldsymbol{A}, \psi) \in \mathcal{A}(\mathcal{L}) : \not{D}_{\boldsymbol{A}}\, \psi = 0,\ \boldsymbol{F}_{\boldsymbol{A}}^{+} = \sigma^{+}((\psi \otimes \psi^{*})_{0})\}/\mathcal{G}(\mathcal{L}). \quad \text{(A.4.20)}$$

The ensuing analysis required to manufacture a differential topological invariant of B from $\mathcal{M}(\mathcal{L})$ is in many ways analogous to that which we outlined for Donaldson theory. For this reason we will simply sketch with much broader strokes those aspects of the construction that are much the same and linger a bit longer over those that present something new. As for Donaldson theory we will ignore the (important) issue of replacing spaces of smooth objects with "appropriate Sobelev completions."

Seiberg-Witten theory is in many ways technically much simpler than Donaldson theory and very often the simplifications can be attributed to the fact that $U(1)$, unlike $SU(2)$, is Abelian. Our first manifestation of this is the ease with which we identify the reducible elements of the configuration space.

Lemma A.4.3 *An element $(\boldsymbol{A},\ \psi)$ of $\mathcal{A}(\mathcal{L})$ is left fixed by some non-identity element σ_{γ} of $\mathcal{G}(\mathcal{L})$ if and only if $\psi \equiv 0$ and, in this case, $\gamma : B \longrightarrow U(1)$ must be a constant map.*

Proof: It will suffice to argue locally so we let $\boldsymbol{\mathcal{A}}$ be a gauge potential for $\boldsymbol{A}$. Then $(\boldsymbol{\mathcal{A}},\ \psi) \cdot \gamma = (\boldsymbol{\mathcal{A}},\ \psi)$ if and only if

$$(\boldsymbol{\mathcal{A}} + 2\,\gamma^{-1} \mathrm{d}\,\gamma,\ \gamma^{-1}\,\psi) = (\boldsymbol{\mathcal{A}},\ \psi)$$

and this, in turn, is the case if and only if

$$\gamma^{-1}\,\psi = \psi \quad \text{and} \quad 2\,\gamma^{-1}\,\mathrm{d}\,\gamma = 0\,.$$

Since $\gamma \neq 1$, the first of these can be true if and only if $\psi \equiv 0$. The second implies $\mathrm{d}\,\gamma = 0$ and, since B is connected, γ must be constant. ∎

An $(\boldsymbol{A},\ \psi) \in \mathcal{A}(\mathcal{L})$ is said to be **reducible** if $\psi \equiv 0$ and **irreducible** otherwise.

As was the case in Donaldson theory, the configuration space $\mathcal{A}(\mathcal{L})$ is an affine space and therefore an infinite-dimensional manifold. The tangent space at any $(\boldsymbol{A},\ \psi) \in \mathcal{A}(\mathcal{L})$ can be identified with

$$T_{(\boldsymbol{A},\psi)}(\mathcal{A}(\mathcal{L})) = \bigwedge\nolimits^{1}(B,\ \mathrm{Im}\quad) \oplus \boldsymbol{\Gamma}(S^{+}(\mathcal{L})) \qquad \text{(A.4.21)}$$

(no adjoint bundle required in the first summand because $U(1)$ is Abelian). The gauge group $\mathcal{G}(\mathcal{L})$ has the structure of a Hilbert Lie group whose Lie algebra can be identified with $\bigwedge^{0}(B,\ \mathrm{Im}\quad)$. Fixing $(\boldsymbol{A},\ \psi) \in \mathcal{A}(\mathcal{L})$, the action of $\mathcal{G}(\mathcal{L})$ on $(\boldsymbol{A},\ \psi)$ gives a map $\mathcal{G}(\mathcal{L}) \longrightarrow \mathcal{A}(\mathcal{L})$ whose derivative at the identity is

$$\begin{aligned} \bigwedge^0(B,\ \mathrm{Im}\quad) &\longrightarrow \bigwedge^1(B,\ \mathrm{Im}\quad)\oplus \boldsymbol{\Gamma}(S^+(\mathcal{L})) \\ a &\longrightarrow (2\,\mathrm{d}\,a,\ -a\cdot\psi)\,, \end{aligned} \tag{A.4.22}$$

where $a\cdot\psi$ is the rotation of ψ obtained by regarding each value of a as an element of the complexified Clifford algebra and Clifford multiplying by a.

Remark: Here is a formal, i.e., C^∞, local argument to persuade you of (A.4.22). The map is $\gamma \in \mathcal{G}(\mathcal{L}) \longrightarrow (\boldsymbol{\mathcal{A}},\ \psi)\cdot\gamma = (\boldsymbol{\mathcal{A}} + 2\gamma^{-1}\mathrm{d}\,\gamma,\ \gamma^{-1}\psi)$. Let $a \in \bigwedge^0(B,\ \mathrm{Im}\quad)$ and write $a(x) = \mathrm{i}\,\theta(x)$, where θ is a real-valued function on B. Then $\alpha(t) = \mathrm{e}^{\mathrm{i}\,t\theta(x)}$ is a curve in $\mathcal{G}(\mathcal{L}) = C^\infty(B,\ U(1))$ with $\alpha(0) = 1$ and $\alpha'(0) = a$. Thus, the derivative at $t = 0$ is

$$\begin{aligned} \frac{d}{dt}\Big((\boldsymbol{\mathcal{A}},\ \psi)\cdot \mathrm{e}^{\mathrm{i}\,t\theta(x)}\Big)\Big|_{t=0} &= \frac{d}{dt}\Big(\boldsymbol{\mathcal{A}} + 2\mathrm{i}\,t\mathrm{d}\theta,\ \mathrm{e}^{-\mathrm{i}\,t\theta(x)}\psi\Big)\Big|_{t=0} \\ &= (2\mathrm{i}\,\mathrm{d}\theta,\ -\mathrm{i}\theta(x)\psi) \\ &= (2\mathrm{d}\,a,\ -a\psi) \\ &= (2\mathrm{d}\,a,\ -a\cdot\psi)\,. \end{aligned}$$

Again as in Donaldson theory, the "large" moduli space

$$\mathcal{B}(\mathcal{L}) = \mathcal{A}(\mathcal{L})/\mathcal{G}(\mathcal{L})$$

of configurations is a smooth Banach manifold away from the reducible configurations (i.e., away from those $[\boldsymbol{A},\ \psi]$ with $\psi \equiv 0$) and the monopole moduli space $\mathcal{M}(\mathcal{L})$ is a subset of it. Define the **Seiberg-Witten map**

$$F : \mathcal{A}(\mathcal{L}) \longrightarrow \bigwedge\nolimits_+^2(B,\ \mathrm{Im}\quad)\oplus \boldsymbol{\Gamma}(S^-(\mathcal{L}))$$

by

$$F(\boldsymbol{A},\ \psi) = (\boldsymbol{F}_{\boldsymbol{A}}^+ - \sigma^+((\psi\otimes\psi^*)_0),\ \not{D}_{\boldsymbol{A}}\,\psi)\,. \tag{A.4.23}$$

Then $(\boldsymbol{A},\ \psi)$ satisfies (SW) if and only if $F(\boldsymbol{A},\ \psi) = (0,0)$. The derivative of F at $(\boldsymbol{A},\ \psi) \in \mathcal{M}(\mathcal{L})$, i.e., the linearization of the Seiberg-Witten equations at $(\boldsymbol{A},\ \psi)$, is

$$\begin{aligned} F_{*(\boldsymbol{A},\psi)} : \bigwedge^1(B,\ \mathrm{Im}\quad)\oplus\boldsymbol{\Gamma}(S^+(\mathcal{L})) &\longrightarrow \bigwedge\nolimits_+^2(B,\ \mathrm{Im}\quad)\oplus\boldsymbol{\Gamma}(S^-(\mathcal{L})) \\ F_{*(\boldsymbol{A},\psi)} &= \begin{pmatrix} \mathrm{d}^+ & -D_\psi \\ \cdot\frac{1}{2}\psi & \not{D}_{\boldsymbol{A}} \end{pmatrix} \end{aligned} \tag{A.4.24}$$

where

$$\mathrm{d}^+ : \bigwedge^1(B,\ \mathrm{Im}\quad) \longrightarrow \bigwedge\nolimits_+^2(B,\ \mathrm{Im}\quad)$$

is d followed by the projection onto the self-dual part,

$$\cdot \frac{1}{2}\psi : \bigwedge^1(B,\ \mathrm{Im}\quad) \longrightarrow \boldsymbol{\Gamma}(\mathcal{S}^-(\mathcal{L}))$$

takes a 1-form $\boldsymbol{\alpha}$ to $\boldsymbol{\alpha} \cdot \frac{1}{2}\psi$ (which is understood to mean Clifford multiplication by the vector field dual to the 1-form $\boldsymbol{\alpha}$) and

$$D_\psi : \boldsymbol{\Gamma}(\mathcal{S}^+(\mathcal{L})) \longrightarrow \bigwedge^2_+(B,\ \mathrm{Im}\quad)$$

is given by

$$D_\psi(\eta) = \sigma^+\left(\psi \otimes \eta^* + \eta \otimes \psi^* - \frac{1}{2}\left(\langle \psi,\ \eta\rangle + \overline{\langle \psi,\ \eta\rangle}\right)\mathbb{1}\right)$$

(the object inside the parentheses being a section of $\mathrm{End}_0(\mathcal{S}^+(\mathcal{L}))$ which σ^+ identifies with a self-dual 2-form on B).

Remark: (A.4.24) can be verified with a local argument analogous to that for (A.4.22).

Associated to any solution $(\boldsymbol{A},\ \psi)$ to (SW) is a **fundamental elliptic complex** $\mathcal{E}(\boldsymbol{A},\ \psi)$:

$$0 \longrightarrow \bigwedge^0(B,\ \mathrm{Im}\quad) \longrightarrow \bigwedge^1(B,\ \mathrm{Im}\quad) \oplus \boldsymbol{\Gamma}(\mathcal{S}^+(\mathcal{L})) \\ \longrightarrow \bigwedge^2_+(B,\ \mathrm{Im}\quad) \oplus \boldsymbol{\Gamma}(\mathcal{S}^-(\mathcal{L})) \longrightarrow 0\,,$$

where the second and third maps are, respectively, the derivative (A.4.22) of the action of $\mathcal{G}(\mathcal{L})$ on $(\boldsymbol{A},\ \psi)$, and the derivative (A.4.24) of the Seiberg-Witten map at $(\boldsymbol{A},\ \psi)$. This complex has finite-dimensional cohomology groups $H^i(\boldsymbol{A},\ \psi),\ i = 0, 1, 2$, which admit interpretations analogous to those in Donaldson theory:

$H^0(\boldsymbol{A},\ \psi) =$ tangent space to the stablizer of $(\boldsymbol{A},\ \psi)$ in $\mathcal{G}(\mathcal{L})$ so

$$\begin{aligned} H^0(\boldsymbol{A},\ \psi) = 0 \quad &\Longleftrightarrow \quad (\boldsymbol{A},\ \psi) \text{ irreducible} \\ &\Longleftrightarrow \quad \psi \not\equiv 0 \end{aligned}$$

$H^1(\boldsymbol{A},\ \psi) =$ formal tangent space to $\mathcal{M}(\mathcal{L})$ at $[\boldsymbol{A},\ \psi]$

$H^2(\boldsymbol{A},\ \psi) =$ obstruction space, i.e.,

$$H^2(\boldsymbol{A},\ \psi) = 0 \Longleftrightarrow \begin{array}{l} \text{Implicit Function Theorem} \\ \text{gives a local manifold} \\ \text{structure for } F^{-1}(0,0) \text{ near} \\ (\boldsymbol{A},\ \psi) \text{ of dimension} \dim H^1(\boldsymbol{A},\ \psi). \end{array}$$

If $H^2(\boldsymbol{A},\ \psi) = 0$ and $H^0(\boldsymbol{A},\ \psi) = 0$, then the local manifold structure for $F^{-1}(0,0)$ near $(\boldsymbol{A},\ \psi)$ projects injectively into the moduli space $\mathcal{M}(\mathcal{L})$ and, near $[\boldsymbol{A},\ \psi]$, $\mathcal{M}(\mathcal{L})$ is a smooth manifold of dimension

$$\begin{aligned}\dim H^1(\boldsymbol{A},\ \psi) &= -\dim H^0(\boldsymbol{A},\ \psi) + \dim H^1(\boldsymbol{A},\ \psi) - \dim H^2(\boldsymbol{A},\ \psi)\\ &= -\text{Index}\ (\mathcal{E}(\boldsymbol{A},\ \psi))\\ &= \frac{1}{4}(c_1(L^0)^2 - 2\chi(B) - 3\sigma(B)),\end{aligned}$$

where the last expression comes from the Atiyah-Singer Index Theorem and $c_1(L^0)^2$ means $\int_B c_1(L^0) \wedge c_1(L^0)$.

As in Donaldson theory, $H^0(\boldsymbol{A},\ \psi) = 0$ and $H^2(\boldsymbol{A},\ \psi) = 0$ are the "generic" situation when $b_2^+(B) > 1$, but this means something slightly different here. First, the part that is the same.

Theorem A.4.4 (Generic Metrics Theorem) *Let B denote a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(B) > 0$. Then there is a dense subset $Gen(\mathcal{R})$ of the space $\mathcal{R}(B)$ of Riemannian metrics on B with the following property: For any $\boldsymbol{g} \in Gen(\mathcal{R})$ and any corresponding spin^c structure $\mathcal{L}$, any solution $(\boldsymbol{A},\ \psi)$ to the Seiberg-Witten equations is irreducible, i. e., satisfies $H^0(\boldsymbol{A},\ \psi) = 0$. If $b_2^+(B) > 1$, then for any generic path $\boldsymbol{g}(t),\ 0 \le t \le 1$, of Riemannian metrics in $\mathcal{R}(B)$ there are no reducible solutions to the Seiberg-Witten equations for any spin^c structure corresponding to any of the metrics $\boldsymbol{g}(t),\ 0 \le t \le 1$.*

For $H^2(\boldsymbol{A},\ \psi) = 0$ there is no known generic metrics theorem of this sort. In this case one must perturb, not the metric, but the equations themselves. More precisely, we fix the metric $\boldsymbol{g}$ and the spin^c structure $\mathcal{L}$. For any fixed $\boldsymbol{\eta} \in \bigwedge^2_+(B,\ \text{Im}\quad)$ we introduce the $(\boldsymbol{\eta})$ **perturbed Seiberg-Witten (PSW) equations**

$$\not{D}_{\boldsymbol{A}}\psi = 0 \tag{A.4.25}$$

$$\boldsymbol{F}^+_{\boldsymbol{A}} = \sigma^+((\psi \otimes \psi^*)_0) + \boldsymbol{\eta}. \tag{A.4.26}$$

Remark: The motivation here is easy to understand. Solutions to (SW) are solutions to the equation $F(\boldsymbol{A},\ \psi) = (0,0)$, where F is the Seiberg-Witten map (A.4.23). For this to be a manifold, $(0,0)$ must be a regular value of F. If it is not, the infinite-dimensional version of Sard's Theorem suggests that a small perturbation of $(0,0)$ in $\bigwedge^2_+(B,\ \text{Im}\quad) \oplus \boldsymbol{\Gamma}(S^-(\mathcal{L}))$ of the form $(\boldsymbol{\eta},\ 0)$ will be a regular value so that $F(\boldsymbol{A},\ \psi) = (\boldsymbol{\eta},\ 0)$ will define a manifold of $(\boldsymbol{A},\ \psi)$. But $F(\boldsymbol{A},\ \psi) = (\boldsymbol{\eta},\ 0)$ is just (PSW).

The linearized complex at any solution to (PSW) is given by the same maps as for (SW) so the cohomology is the same. $\mathcal{G}(\mathcal{L})$ acts on solutions to (PSW) in the same way so there is a moduli space $\mathcal{M}(\mathcal{L},\ \boldsymbol{\eta})$ of solutions and everything we have said above for (SW) is also true for (PSW).

Theorem A.4.5 (Generic Perturbations Theorem) *Let B denote a compact, connected, simply connected, oriented, smooth 4-manifold. Fix a Riemannian metric $\boldsymbol{g}$ and a spin^c structure $\mathcal{L}$ for B. Then there is a dense subset $Gen(\bigwedge_+^2)$ in the space $\bigwedge_+^2(B,\ Im\ \ )$ of Im -valued self-dual 2-forms on B with the following properties: For $\boldsymbol{\eta} \in Gen(\bigwedge_+^2)$, every solution $(\boldsymbol{A},\ \psi)$ to the perturbed Seiberg-Witten equations (A.4.25) and (A.4.26) has $H^2(\boldsymbol{A},\ \psi) = 0$. If $b_2^+(B) > 0$ and $\boldsymbol{g} \in Gen(\mathcal{R})$, then, for any $\boldsymbol{\eta} \in Gen(\bigwedge_+^2)$, the moduli space $\mathcal{M}(\mathcal{L},\ \boldsymbol{\eta})$ is a smooth submanifold of $\mathcal{B}(\mathcal{L})$ of dimension $\frac{1}{4}(c_1(L^0)^2 - 2\chi(B) - 3\sigma(B))$.*

Remark: In particular, if $c_1(L^0)^2 - 2\chi(B) - 3\sigma(B) < 0$, then the moduli space is generically empty.

Exactly as in the case of Donaldson theory one can show that, for a fixed generic metric $\boldsymbol{g}$ and perturbation $\boldsymbol{\eta}$ and any associated spin^c structure $\mathcal{L}$, a choice of orientation for the vector space $H_+^2(B;\ \ )$ canonically orients all of the moduli spaces $\mathcal{M}(\mathcal{L},\ \boldsymbol{\eta})$. Likewise as in Donaldson theory, when $b_2^+(B) > 1$ there is a cobordism result which roughly says that for a generic l-parameter family $\boldsymbol{g}(t),\ 0 \leq t \leq 1$, of metrics and a generic l-parameter family $\boldsymbol{\eta}(t),\ 0 \leq t \leq 1$, of perturbations, the moduli spaces parametrized by t fit together to form a smooth manifold with boundary containing no points corresponding to reducible solutions. The boundary is the disjoint union of moduli spaces for $(\boldsymbol{g}(0),\ \boldsymbol{\eta}(0))$ and $(\boldsymbol{g}(1),\ \boldsymbol{\eta}(1))$. Moreover, selecting an orientation for $H_+^2\ (B;\ \ )$ orients this parametrized moduli space and the two boundary moduli spaces inherit opposite orientations.

Remark: There is a technical point which we glossed over here and should mention because it has no analogue in Donaldson theory. Changing the metric $\boldsymbol{g}$ changes the orthonormal frame bundle and so, one would think, the spin^c structure. It would appear that the discussion above is incomplete without a specification of a spin^c structure for each t. In fact, however, one can show that frame bundles for different metrics are naturally isomorphic and so one can pull back spin^c structures by the isomorphisms, thus effectively "fixing" $\mathcal{L}$ (up to equivalence) regardless of the choice of $\boldsymbol{g}$.

Except for a few minor simplifications and adaptations the story of the Seiberg-Witten moduli space thus far has been virtually indistinguishable from what we had to say about the anti-self-dual moduli space. The one aspect of Seiberg-Witten theory that differs significantly from Donaldson theory (and that accounts for its relative simplicity) is that there is no need for an "Uhlenbeck- style compactification":

> *For any metric $\boldsymbol{g}$, and spin^c structure $\mathcal{L}$ and any perturbation $\boldsymbol{\eta}$, the moduli space $\mathcal{M}(\mathcal{L},\ \boldsymbol{\eta})$ is always compact.*

The proof of this involves what is called an elliptic "bootstrapping" argument (which we will not describe) based on the crucial fact that the spinor field ψ and curvature $\boldsymbol{F}_{\boldsymbol{A}}$ for any solution $(\boldsymbol{A},\ \psi)$ to (PSW) satisfy **uniform a priori bounds** (this is categorically false for the anti-self-dual equations because these are conformally invariant in dimension four). Because of its significance we will sketch a proof of this but, since the perturbation adds only arithmetic to the argument, we will do this for the unperturbed equations, written in the form

$$\not{D}_{\boldsymbol{A}}\psi = 0 \tag{A.4.27}$$

$$\rho^+(\boldsymbol{F}_{\boldsymbol{A}}^+) = (\psi \otimes \psi^*)_0 \tag{A.4.28}$$

(recall that ρ^+ is the inverse of σ^+). We will appeal to the famous *Weitzenböck formula* from differential geometry which, in our present circumstances, reads

$$\not{D}_{\boldsymbol{A}}^* \circ \not{D}_{\boldsymbol{A}}\psi = \nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi + \frac{1}{4}\kappa\psi + \rho^+(\boldsymbol{F}_{\boldsymbol{A}}^+)\psi, \tag{A.4.29}$$

where $D_{\boldsymbol{A}}^*$ is the formal adjoint of $\not{D}_{\boldsymbol{A}} : \boldsymbol{\Gamma}(S^+(\mathcal{L})) \longrightarrow \boldsymbol{\Gamma}(S^-(\mathcal{L}))$, $\nabla_{\boldsymbol{A}}^*$ is the formal adjoint of the covariant derivative $\nabla_{\boldsymbol{A}} : \boldsymbol{\Gamma}(S^+(\mathcal{L})) \longrightarrow \wedge^1(B) \otimes \boldsymbol{\Gamma}(S^+(\mathcal{L}))$ and κ is the scalar curvature of B (for the metric $\boldsymbol{g}$). Because $(\boldsymbol{A},\ \psi)$ is a solution to the Seiberg-Witten equations, (A.4.29) reduces to

$$\nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi + \frac{1}{4}\kappa\psi + (\psi \otimes \psi^*)_0\psi = 0. \tag{A.4.30}$$

Take the pointwise inner product with ψ to obtain

$$\langle\, \nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi, \psi\,\rangle + \frac{1}{4}\kappa\|\,\psi\,\|^2 + \frac{1}{2}\|\,\psi\,\|^4 = 0 \tag{A.4.31}$$

(for the last term compute $(\psi \otimes \psi^*)_0\psi$ from (A.2.49) and then take the inner product with ψ).

Now, $\|\,\psi(x)\,\|$ is a continuous function on the compact space B so there is an $x_0 \in B$ at which it achieves an absolute maximum. We claim that, at this point, the first term in (A.4.31) is non-negative, i.e.,

$$\langle\, \nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi(x_0), \psi(x_0)\,\rangle \geq 0 \tag{A.4.32}$$

(note that (A.4.31) implies that $\langle\, \nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi, \psi\,\rangle$ must be real). The proof depends on the identity

$$\Delta_{\boldsymbol{g}}\|\,\psi\,\|^2 = -2\|\,\nabla_{\boldsymbol{A}}\psi\,\|^2 + 2\langle\, \nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi, \psi\,\rangle \tag{A.4.33}$$

where $\Delta_{\boldsymbol{g}} = \mathrm{d}^* \circ\ d$ is the scalar Laplacian corresponding to $\boldsymbol{g}$. This can be verified by writing out $\Delta_{\boldsymbol{g}}\|\,\psi\,\|^2$ in a local orthonormal frame field. Now, at x_0, (A.4.33) gives

$$2\langle\, \nabla_{\boldsymbol{A}}^* \circ \nabla_{\boldsymbol{A}}\psi(x_0), \psi(x_0)\,\rangle = \triangle_{\boldsymbol{g}}\|\,\psi\,\|^2(x_0) + 2\|\,\nabla_{\boldsymbol{A}}\psi\,\|^2(x_0). \tag{A.4.34}$$

Now, obviously $2\| \nabla_{\boldsymbol{A}}\psi \|^2(x_0) \geq 0$. Moreover, since $\| \psi \|^2$ achieves a maximum at $x_0, \triangle_{\boldsymbol{g}}\| \psi \|^2(x_0) \geq 0$ as well so (A.4.32) is proved. It then follows from (A.4.31), evaluated at x_0, that

$$\frac{1}{4}\kappa(x_0)\| \psi(x_0) \|^2 + \frac{1}{2}\| \psi(x_0) \|^4 \leq 0$$

so

$$\| \psi(x_0) \|^4 \leq -\frac{1}{2}\kappa(x_0)\| \psi(x_0) \|^2.$$

There are two possibilities. Either $\| \psi(x_0) \| = 0$, in which case $\psi \equiv 0$ and $(\boldsymbol{A}, \psi)$ is a reducible solution (and so $\| \psi \|$ is certainly uniformly bounded). Otherwise we have

$$\| \psi(x_0) \|^2 \leq -\frac{1}{2}\kappa\ (x_0)$$

and therefore

$$\| \psi(x) \|^2 \leq -\frac{1}{2}\kappa(x_0) \qquad \text{(A.4.35)}$$

for every x in B. Now, despite appearences, the right-hand side of (A.4.35) depends on ψ (through x_0) so, to get a bound on the spinor field of every solution $(\boldsymbol{A}, \psi)$ we define $k(x_0) = \max\ \{-\frac{1}{2}\kappa(x_0), 0\}$ for each $x_0 \in B$ and

$$k(B) = \max\{k(x_0) : x_0 \in B\}$$

and conclude that for any fixed metric and any spin^c structure, any solution $(\boldsymbol{A}, \psi)$ to the Seiberg-Witten equations has spinor field ψ bounded by the geometrical constant $k(B)$:

$$\| \psi(x) \|^2 \leq k(B) \qquad \forall x \in B. \qquad \text{(A.4.36)}$$

The second Seiberg-Witten equation (A.3.6) then gives a uniform bound on the self-dual part of the curvature for any solution. A bit more work then gives a bound on the anti-self-dual part of the curvature. From these one can deduce that, for a given $\boldsymbol{g}$ (and $\boldsymbol{\eta}$ in the perturbed case) there are at most finitely many (equivalence classes of) spin^c structures for which the moduli space is nonempty (Theorem 5.2.4 of [**M1**]).

The bounds described thus far are not sufficient to prove the compactness of the moduli space. For this one must bound the connection parts $\boldsymbol{A}$ of solutions $(\boldsymbol{A}, \psi)$ "up to gauge". This is generally accomplished by a gauge fixing argument and the bootstrapping referred to above (see Section 5.3 of [**M1**]). In any case, it can be done and the end result is that Seiberg-Witten moduli spaces are *always* compact (and, generically, are smooth manifolds). This compactness simplifies enormously the task of defining and computing "Donaldson-like" invariants in Seiberg-Witten theory (because there is no need, as there is in the anti-self-dual case, to compactify the moduli space before integrating cohomology classes over it). Even so we intend to consider what appears to be only a very special case. Thus, we fix a generic metric $\boldsymbol{g}$ and perturbation $\boldsymbol{\eta}$ and suppose that there exists a spin^c structure $\mathcal{L}$ for which the formal dimension of the moduli space is zero, i.e.,

$$c_1(L^0)^2 = 2\chi(B) + 3\sigma(B) \tag{A.4.37}$$

(such an $\mathcal{L}$ need not exist). Assuming that an orientation for the vector space $H^2_+(B;\ \ )$ has been fixed, the moduli space is a finite set of isolated points each of which is equipped with a sign ± 1. The sum of these signs is an integer and, as in Donaldson theory, when $b_2^+(B) > 1$ a cobordism argument shows that the integer is independent of the choice of (generic) metric and perturbation. We call this integer the **0-dimensional Seiberg-Witten invariant of** B **associated with** $\mathcal{L}$ and denote it

$$SW_0(B, \mathcal{L}).$$

This is, indeed, an invariant in the sense that, if $f : B' \longrightarrow B$ is a diffeomorphism for which the induced map $f^* : H^2_+(B;\ \ ) \longrightarrow H^2_+(B';\ \ )$ preserves orientation, then the induced spin^c structure $f^*\mathcal{L}$ for B' also satisfies (A.4.37) and $SW_0(B', f^*\mathcal{L}) = SW_0(B, \mathcal{L})$.

Remark: When one goes to the trouble of defining the Seiberg-Witten invariant even when (A.4.37) is not satisfied (which we will not) one obtains a map $SW(B, \cdot)$ on the set of (equivalence classes of) spin^c structures on B which assigns to each such an integer $SW(B, \mathcal{L}) \in \mathbb{Z}$, taken to be zero if the corresponding moduli space is empty. We have already noted that there can be at most finitely many $\mathcal{L}$ for which $SW(B, \mathcal{L}) \neq 0$. The empirical evidence suggests that, when $b_2^+(B) > 1, SW(B, \mathcal{L}) \neq 0$ occurs only for those $\mathcal{L}$ satisfying (A.4.37). We will say that a B with $b_2^+(B) > 1$ is of ***SW*-simple type** if $SW(B, \mathcal{L}) \neq 0$ implies that $\mathcal{L}$ must satisfy (A.4.37), i.e., if nonzero Seiberg-Witten invariants occur only for 0-dimensional moduli spaces. It has been conjectured that every B with $b_2^+(B) > 1$ is of SW-simple type. Finally, we shall refer to the elements of $H^2(B)$ which arise as $c_1(L^0)$ for some spin^c structure $\mathcal{L}$ satisfying (A.4.37) as **Seiberg-Witten (*SW*-) basic classes.** Thus, SW-basic classes are just the first Chern classes of complex line bundles corresponding to spin^c structures for which the SW moduli space is 0-dimensional.

A.5 The Witten Conjecture

As we saw in Section A.l, Witten was led to Seiberg-Witten gauge theory through topological quantum field theory and saw it as a dual version of Donaldson theory. Indeed, Witten was led to much more. He formulated a very concrete conjecture on the relationship between Donaldson and Seiberg-Witten invariants. The conjecture was remarkable. It was a very deep and purely mathematical statement, but one suggested entirely by physics. We will conclude with a very brief description of what Witten believed must be true (and was eventually proved by Feehan and Leness [**FeLe**]).

The sequence $\gamma_d(B), d = 0, 1, 2, \ldots$, of Donaldson invariants for B can be assembled into a single formal power series

$$\mathcal{D}_B(x) = \sum_{d=0}^{\infty} \frac{\gamma_d(B)(x)}{d!}$$

on $H^2(B)$ called the **Donaldson series** of B. For example, all of the invariants for $K3$ have been computed and one finds that

$$\mathcal{D}_{K3}(x) = \sum_{n=0}^{\infty} \frac{1}{n!}(Q_{K3}(x,\, x)/2)^n = \exp(Q_{K3}(x,\, x)/2).$$

Kronheimer and Mrowka [**KM**] formulated a condition on B which we will call **KM-simple type** that is sufficient to guarantee that if one is somehow given the Donaldson series $\mathcal{D}_B(x)$ one can inductively extract from it all of the individual Donaldson invariants $\gamma_d(B)$.

Remark: In order to state this condition precisely one requires more detailed information about how the $\gamma_d(B)$ are "really" defined (see the Remark following our "naive" definition in Section A.1). However, there is no known counterexample to the conjecture that every manifold B of the type we are considering is of KM-simple type.

They then proved a remarkable structure theorem for such manifolds the essential content of which was that all of the information contained in the (infinitely many) Donaldson invariants is also contained in a finite set of data (certain cohomology classes $K_1, \ldots, K_s$ of B and corresponding rational numbers $a_1, \ldots, a_s$). Specifically, they proved the following theorem.

Theorem A.5.1 *If B is of KM-simple type, then there exist cohomology classes $K_1, \ldots, K_s \in H^2(B)$, called **KM-basic classes**, and rational numbers $a_1, \ldots, a_s \in$, called **KM-coefficients**, such that*

$$\mathcal{D}_B(x) = \exp(Q_B(x,\, x)/2) \sum_{r=1}^{s} a_r \exp(K_r(x))$$

for every $x \in H_2(B)$. Moreover, each KM-basic class reduces mod 2 to the second Stiefel-Whitney class $w_2(B)$ of B.

The appearance of Theorem A.5.1 in the spring of 1994 was an extraordinary breakthrough in Donaldson theory. Enormously complicated calculations of an infinite set of invariants were suddenly replaced by the (certainly not easy, but at least finite) problem of determining the KM-basic classes and coefficients. As fate would have it, however, the fall of 1994 witnessed another event which rendered this triumph of Kronheimer and Mrowka moot. Edward Witten, in his now famous lecture at M.I.T. (described in [**Tau2**]), made the conjecture which, within weeks, brought about the demise of Donaldson theory and initiated an entirely new approach to the study of smooth 4-manifolds.

Witten's Conjecture *Let B be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(B) > 1$ and odd. Then*

1. *B is of SW-simple type if and only if B is of KM-simple type and, in this case,*
2. *SW-basic classes coincide with KM-basic classes and*

$$\mathcal{D}_B(x) = \exp\big(Q_B(x,\, x)/2\big) \sum 2^{m(B)} SW_0(B, \mathcal{L}) \exp(c_1(L^0)(x)),$$

where $m(B) = 2 + \frac{1}{4}(7\chi(B) + 11\sigma(B))$, $\chi(B)$ is the Euler characteristic of $B, \sigma(B)$ is the signature of B and the sum is over all (equivalence classes of) spin^c structures $\mathcal{L}$ for which (A.4.37) is satisffied.

The content of the conjecture is that, for manifolds of simple type, the basic classes are just those elements of $H_2(B)$ corresponding to spin^c structures with 0-dimensional SW moduli spaces and the coefficients are, up to the topological factor $2^{m(B)}$, just the corresponding 0-dimensional SW invariants.

There are a number of attitudes one might adopt toward a conjecture of this sort. One might, of course, try to prove it and, although this has not been the principal focus of work in this area, much deep and interesting mathematics has been directed toward this end. A strategy suggested by Pidstrigatch and Tyurin [**PT**] for relating the two sets of invariants by viewing the Donaldson and Seiberg-Witten moduli spaces as singular submanifolds of a larger moduli space of "nonabelian monopoles" has been taken up in earnest by Feehan and Leness in a long series of difficult and technical papers (see [**FeLe**]). By far the more prevalent attitude has been that, even if a proof is hard to come by, the conjecture has been checked in every case in which all of the invariants are known and has survived so that it would seem to make good practical sense for the topologist to (at least provisionally) abandon the Donaldson invariants for the much more tractable Seiberg-Witten invariants. Perhaps the most enlightened attitude, however, and one which has been emphasized by Atiyah, is that if physics is truly capable of casting such a penetrating light upon mathematics at the very deepest levels, then mathematicians will want to take heed and turn their attention once again to their historical roots in physics.

References

[AB] Aharonov, Y. and D. Bohm, "Significance of electromagnetic potentials in the quantum theory", Phys. Rev., 115(1959), 485–491.

[AJ] Atiyah, M.F. and L. Jeffrey, "Topological Lagrangians and Cohomology", J. Geo. Phys., 7(1990), 119–136.

[BB] Booss, B. and D.D. Bleecker, *Topology and Analysis: The Atiyah-Singer Index Formula and Gauge-Theoretic Physics*, Springer-Verlag, New York, Berlin, 1985.

[BGV] Berline, N., E. Getzler and M. Vergne, *Heat Kernels and Dirac Operators*, Springer-Verlag, New York, Berlin, 1996.

[Bl] Bleecker, D., *Gauge Theory and Variational Principles*, Addison-Wesley, Reading, MA, 1981.

[BPST] Belavin, A., A. Polyakov, A. Schwartz and Y. Tyupkin, "Pseudoparticle solutions of the Yang-Mills equations", Phys. Lett., 59B(1975), 85–87.

[Brow] Browder, A., *Mathematical Analysis: An Introduction*, Springer-Verlag, New York, Berlin, 1996.

[BT] Bott, R. and L.W. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag, New York, Berlin, 1982.

[DT] Dereli, T. and M. Tekmen, "Wu-Yang monopoles and non-Abelian Seiberg-Witten equations", hep-th/9806031.

[EGH] Eguchi, T., P.B. Gilkey and A.J. Hanson, "Gravitation, Gauge Theories and Differential Geometry", Physics Reports, 66, No. 6(1980), 213–393.

[Fang] Fang, J., *Abstract Algebra*, Schaum Publishing Co., New York, 1963.

[FeLe] Feehan, P.M.N. and T.G. Leness, "Witten's Conjecture for Many Four-Manifolds of Simple Type", arXiv:math/0609530v2[math.DG].

[Fey] Feynman, R.P., R.B. Leighton and M. Sands, *The Feynman Lectures on Physics*, Volumes I–III, Addison-Wesley, Reading, MA, 1964.

[FG-M] Feynman, R.P. and M. Gell-Mann, "Theory of Fermi interactions", Phys. Rev., 109, 193.

[FU] Freed, D.S., and K.K. Uhlenbeck, *Instantons and Four-Manifolds*, MSRI Publications, Springer-Verlag, New York, Berlin, 1984.

[G1] Geroch, R.P., "Spinor structure of space-times in general relativity, I", J. Math. Phys., 9(1968), 1739–1744.

[G2] Geroch, R.P., "Spinor structure of space-times in general relativity, II", J. Math. Phys., 11(1970), 343–348.

[GHVI] Greub, W., S. Halperin and R. Vanstone, *Connections, Curvature and Cohomology*, Vol. I, Academic Press, New York, 1972.

[Gre] Greiner, W., *Relativistic Quantum Mechanics*, Springer-Verlag, New York, Berlin, 1993.

[GO] Goddard, P. and D.I. Olive, "Magnetic monopoles in gauge field theories", Rep. Prog. Phys., Vol. 41(1978), 1357–1437.

[GP] Guillemin, V. and A. Pollack, *Differential Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[Groi1] Groisser, D., "$SU(2)$ Yang-Mills-Higgs Theory on ^{3}", Harvard Ph.D. thesis, 1983.

[Groi2] Groisser, D., "Integrality of the monopole number in $SU(2)$ Yang-Mills-Higgs theory on ^{3}", Comm. Math. Phys., 93(1984), 367–378.

[GR] Gunning, R.C. and H. Rossi, *Analytic Functions of Several Complex Variables*, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[Gui] Guidry, M., *Gauge Field Theories*, John Wiley & Sons, Inc., New York, 1991.

[HE] Hawking, S.W. and G.F.R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge University Press, Cambridge, England, 1976.

[Helg] Helgason, S., *Differential Geometry and Symmetric Spaces*, Academic Press, New York, 1962.

[Hir] Hirsch, M.W., *Differential Topology*, Springer-Verlag, GTM #33, New York, Berlin, 1976.

[Hol] Holstein, B.R., *Topics in Advanced Quantum Mechanics*, Addison-Wesley, Redwood City, CA, 1992.

[Howe] Howe, R., "Very basic Lie theory", Amer. Math. Monthly, November, 1983, 600–623.

[Huse] Husemoller, D., *Fibre Bundles*, 3^{rd} Ed., Springer-Verlag, New York, Berlin, 1994.

[Jost] Jost, J., *Riemannian Geometry and Geometric Analysis, Third Edition*, Springer-Verlag, New York, Berlin, 2002.

[JT] Jaffe, A. and C. Taubes, *Vortices and Monopoles: Structure of Static Gauge Theories*, Birkhauser, Boston, 1980.

[KM] Kronheimer, P.B. and T.S. Mrowka, "Embedded Surfaces and the Structure of Donaldson's Polynomial Invariants", J. Diff. Geo., 3(1995), 573–734.

[KN1] Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry*, Vol. 1, Wiley-Interscience, New York, 1963.

[KN2] Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry*, Vol. 2, Wiley-Interscience, New York, 1969.

[Lang] Lang, S., *Linear Algebra*, 2^{nd} Edition, Addison-Wesley, Reading, MA, 1971.

[Lan] Landkof, N.S., *Foundations of Modern Potential Theory*, Springer-Verlag, New York, Berlin, 1972.

[LM] Lawson, H.B. and M.L. Michelsohn, *Spin Geometry*, Princeton University Press, Princeton, NJ, 1989.

[M1] Morgan, J.W., *The Seiberg-Witten Equations and Applications to the Topology of Smooth Four-Manifolds*, Princeton University Press, Princeton, NJ, 1996.

[M2] Morgan, J.W., "An Introduction to Gauge Theory", in *Gauge Theory and the Topology of Four-Manifolds*, American Mathematical Society, 1998.

[Miln] Milnor, J., *Topology from the Differentiable Viewpoint*, University of Virginia Press, Charlottesville, VA, 1966.

[MQ] Mathai, V. and D. Quillen, "Superconnections, Thom Classes and Equivariant Differential Forms", Topology, 25(1986), 85–110.

[MS] Milnor, J.W. and J.D. Stasheff, *Characteristic Classes*, Princeton University Press, Princeton, NJ, 1974.

[MT] Madsen, I. and J. Tornehave, *From Calculus to Cohomology: De Rham Cohomology and Characteristic Classes*, Cambridge University Press, Cambridge, England, 1997.

[N1] Naber, G.L., *Topological Methods in Euclidean Spaces*, Dover Publications, Mineola, NY, 2000.

[N2] Naber, G.L., *Spacetime and Singularities*, Cambridge University Press, Cambridge, England, 1988.

[N3] Naber, G.L., *The Geometry of Minkowski Spacetime*, Springer-Verlag, Applied Mathematical Sciences Series #92, 2^{nd} Edition, New York, Berlin (to appear 2011).

[N4] Naber, G.L., *Topology, Geometry and Gauge Fields: Foundations*, Springer-Verlag, Texts in Applied Mathematics 25, 2^{nd} Edition, New York, Berlin, 2010.

[O'N] O'Neill, B., *Semi-Riemannian Geometry: with Applications to Relativity*, Academic Press, 1983.

[Pen] Penrose, R., *Techniques of Differential Topology in Relativity*, S.I.A.M., Philadelphia, PA, 1972.

[PT] Pidstrigach, V and A. Tyurin, "Localization of Donaldson's Invariants along Seiberg-Witten Classes", arXiv: dg-ga/9507004.

[Ry] Ryder, L.H., *Quantum Field Theory*, Cambridge University Press, Cambridge, England, 1996.

[SaW] Sachs, R.H. and H. Wu, *General Relativity for Mathematicians*, Springer-Verlag, New York, Berlin, 1977.

[Sp1] Spivak, M., *Calculus on Manifolds*, W. A. Benjamin, Inc., New York, 1965.

[Sp2] Spivak, M., *A Comprehensive Introduction to Differential Geometry,* Volumes I–V, Publish or Perish, Inc., Boston, 1979.

[St] Steenrod, N., *The Topology of Fibre Bundles*, Princeton Univ. Press, Princeton, N.J., 1951.

[SW] Seiberg, N. and E. Witten, "Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD", Nucl. Phys. B 431(1994), 581–640.

[Tau1] Taubes, C.H., "Self-Dual Connections on Non-Self-Dual 4-Manifolds", J. Diff. Geo., 17(1982), 139–170.

[Tau2] Taubes, C.H., *The Seiberg-Witten Invariants* (Videotape), American Mathematical Society, Providence, RI, 1995.

[vdW] van der Waerden, B.L., *Group Theory and Quantum Mechanics*, Springer-Verlag, New York, Berlin, 1974.

[W1] Witten, E., "Supersymmetry and Morse Theory", J. Diff. Geo., 17(1982), 661–692.

[W2] Witten, E., "Topological quantum field theory", Comm. Math. Phys. 117(1988), 353–386.

[W3] Witten, E., "Monopoles and 4-manifolds", Math. Res. Letters 1(1994), 764–796.

[Warn] Warner, F.W., *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, GTM #94, New York, Berlin, 1983.

[Wein] Weinberg, S., *The Quantum Theory of Fields*, Volumes I and II, Cambridge University Press, Cambridge, England, 1995 and 1996.

[YM] Yang, C.N. and R.L. Mills, "Conservation of isotopic spin and isotopic gauge invariance", Phys. Rev., 96(1954), 191–195.

Symbols

What follows is a list of the those symbols that are used consistently throughout the text, a brief discription of their meaning and/or a reference to the page on which such a description can be found.

, real numbers

, complex numbers

, quaternions

= , , or

$\langle\ ,\ \rangle$, inner product on ^{n}

$|\ \ |$, absolute value in , modulus in or

dim X, dimension of the manifold X

$C^\infty(X)$, real-valued C^∞-functions on X

id, identity map

$T_p(X)$, tangent space at p

D_i, i^{th} partial derivative operator

ι, inclusion map

$\alpha'(t_0)$, velocity vector to α at t_0

f_{*p}, derivative of f at p

$\mathcal{V}^*$, dual of vector space $\mathcal{V}$

$(U_S, \varphi_S), (U_N, \varphi_N)$, stereographic projection charts

$\|\ \ \|$, norm in Euclidean space

$SU(2)$, 5

$^{n-1}$, projective spaces

$[\xi]$, equivalence class containing ξ

$\mathcal{P}$, projection map

$\mathcal{GL}(n,\ )$, 5

$GL(n,\ )$, 5

$U(n,\ )$, 5

$O(n)$, 5

$U(n)$, 5

$Sp(n)$, 5

$SO(n)$, 5

$SU(n)$, 5

$\bar{A}^T$, conjugate transpose of A

$\varphi \times \psi$, 5

Index