Connections, gauge theory and characteristic classes

Arie Peterson and Loek Spitz
Universiteit van Amsterdam

\footnote{Supervised by prof. dr. Erik Verlinde and prof. dr. Eric Opdam}
Abstract

This paper tries to explain the mathematics behind gauge theory, especially fibre bundles and connections. We use those concepts to describe electromagnetism as a gauge theory, and show how the Aharonov-Bohm effect can be understood as a gauge-theoretic phenomenon. Finally, we introduce characteristic classes of vector bundles. It is aimed at advanced undergraduate students with an interest in both mathematics and physics.
## Contents

0.1 Introduction .................................................. 3

1 Fibre bundles and vector bundles 4
  1.1 Fibre bundles ............................................. 5
  1.2 \(G\)-bundles ............................................. 8
  1.3 \(G\)-torsors ............................................. 8
  1.4 Principal \(G\)-bundles .................................. 10
  1.5 Vector bundles .......................................... 12

2 Connections 14
  2.1 Connections ............................................. 14
     2.1.1 Linear connections ................................ 18
     2.1.2 Invariant connections ............................. 19
  2.2 Parallel transport ..................................... 21
  2.3 Curvature ............................................... 24
  2.4 Holonomy and monodromy ............................... 26
  2.5 Covariant derivative ................................... 29

3 Applications 31
  3.1 Electromagnetism as a gauge theory .................. 31
     3.1.1 The abelian Aharonov-Bohm effect ............... 33
  3.2 Chern-Simons theory .................................... 37
  3.3 \(S^2\) is no Lie group ................................. 38

4 Characteristic classes 40
  4.1 Introduction ........................................... 40
  4.2 Characteristic classes .................................. 41
  4.3 Covariant derivative revisited ....................... 42
     4.3.1 Curvature ......................................... 44
  4.4 Invariant polynomials .................................. 45
  4.5 Chern classes .......................................... 51
0.1 Introduction

This paper is written as a bachelor thesis for the physics and mathematics programme at the Universiteit van Amsterdam. We were aided in this work by prof. dr. Erik Verlinde and prof. dr. Eric Opdam. While working on this project we have had, at different times, several different goals in mind: among those were understanding and explaining the Verlinde algebra, Chern-Simons theory, an article by Witten about topological quantum field theory, an article by Verlinde about the non-abelian Aharonov-Bohm effect, gauge theory, and characteristic classes. In the end we settled for explaining the mathematics behind gauge theory, describing electromagnetism as a gauge theory, showing how the Aharonov-Bohm effect can be understood as a gauge-theoretic phenomenon, and giving an introduction to characteristic classes.

In the first chapter we explain the concept of a fibre bundle quite extensively, as this is the basis on which everything else depends. In the second chapter we widen the mathematical basis by talking about connections, parallel transport, curvature, holonomy and monodromy, and covariant derivatives. Throughout these two chapters we try to give examples of the concepts we introduce, but most of these examples will be mathematical in nature (recurring most often is the Möbius strip). The physical examples will have to wait until chapter three, where we discuss electromagnetism and the Aharonov-Bohm effect. In the last chapter we treat characteristic classes, through use of invariant polynomials and the curvature of a connection.
Chapter 1

Fibre bundles and vector bundles

If you have two topological spaces, you can use them to define a new one: the product space. In a way, this is not a very interesting construction, as you have only one way of doing it. It is straightforward and easily pictured by putting one space on a horizontal axis, the other on a vertical one; the plane then represents the product space.

Now sometimes we want to have a little more options: we would like to be able to make a product of two spaces, but in a twisted way.

Example 1.0.1. In physics, this might occur in the following simple situation. Imagine you are walking along a path winding through a mountainous landscape, and you want to know what velocities you can have at a certain point on the path. Obviously, your velocity can only be in the direction of the path, either forwards or backwards. It can have any size (putting relativistic objections aside for a moment, for the sake of this simple example); therefore, we could say it can take values in the vector space $\mathbb{R}$. This goes for any point on the path. To describe the whole situation at once then, it would be useful to be able to look at a space which consists of the path with the vector space $\mathbb{R}$ attached to it at every point, in such a way as to make it tangent to the path. Since the path winds through the mountains, each copy of $\mathbb{R}$ has to be attached in its own way; we need (mathematical) tools to do this. These tools are precisely given by the concept of a fibre bundle (actually, in this example we are dealing with a very special sort of fibre bundle, namely a vector bundle, as we’ll see in section 1.5).

Example 1.0.2. A very simple mathematical use of fibre bundles is given by the Möbius strip (see figure 1.1). The Möbius strip is almost a product of the circle with the unit interval, but with one twist given to it. Another way of saying this is that the Möbius strip locally (to be shortly defined precisely) looks like a product, but globally it is given a twist. This kind of construction is exactly what a fibre bundle does for you.
1.1 Fibre bundles

Having given some intuitive ideas about fibre bundles, it is now time to give the mathematical definition and look at some examples more closely.

Definition 1.1.1. A fibre bundle consists of the following data:

- Three topological spaces: The base space $X$, the fibre $F$ and the total space $E$.
- A map $p : E \to X$ called the projection.
- An open covering $\bigcup_{\alpha} U_{\alpha} = X$ of the base space $X$.
- For each $\alpha$ a homeomorphism $h_{\alpha} : p^{-1}(U_{\alpha}) \to U_{\alpha} \times F$ called a local trivialisation, with the property that if $y \in p^{-1}(x)$, then $h_{\alpha}(y) = (x, f)$ for some $f \in F$.

Furthermore, if $U_{\alpha} \cap U_{\beta} \neq \emptyset$ then the map

$$h_{\alpha} \circ h_{\beta}^{-1} : (U_{\alpha} \cap U_{\beta}) \times F \to (U_{\alpha} \cap U_{\beta}) \times F$$

can be written as

$$h_{\alpha} \circ h_{\beta}^{-1}(x, f) = (x, h_{\alpha\beta}(x)(f)),$$  \hspace{1cm} (1.1)

where

$$h_{\alpha\beta} : U_{\alpha} \cap U_{\beta} \to \text{Aut}(F)$$

are called the transition maps.

Now before going any further, let us examine this definition more closely to try and find out why we want it to be like this. To begin with it might be a good idea to look at the simplest example of a fibre bundle; the product of two spaces $X$ and $F$. Just going through the definition, we have:

A base space $X$, a fibre $F$ and a total space $E$ which we define as $E = X \times F$. The projection $p$ we will take to be the usual projection onto the base space. The open covering of $X$ will be given by just the set $X$ and the local trivialisation can then be the identity on $p^{-1}(X) = E = X \times F$. Equation (1.1) is automatically satisfied by taking the sole transition map (for $U_{\alpha} = U_{\beta} = X$) to take every element of $X$ to the identity in $\text{Aut}(F)$. The fibre bundle we have thus created is called the trivial bundle over $X$ with fibre $F$. 
So it is possible to see a product of two spaces as a fibre bundle over one of them. Furthermore, this bundle is called the trivial bundle. In the definition above we have already seen the word ‘trivial’; we said that to have a fibre bundle you have to have some maps called local trivialisations. Now this is an interesting name for these maps; there must be a good reason for using it.

We call these maps local trivialisations because they trivialise the bundle locally; they tell you how the bundle can locally (in the inverse image of an $U_\alpha$ under $p$) be seen as a product of $U_\alpha$ with $F$, which as we saw is a trivial example of a fibre bundle. In the example above local is the same as global (the only open set in the covering being $X$ itself), therefore the bundle is not only trivialised locally, but even globally. Hence the name: trivial bundle.

There is another name in the definition of a fibre bundle that might give us a hint as to how we could visualize the concept: the transition maps. In order to appreciate their name, it is a good idea to examine them a bit closer. First, it is easy to see that these maps satisfy the following relations for $U_\alpha \cap U_\beta \neq \emptyset$ and $U_\alpha \cap U_\beta \cap U_\gamma \neq \emptyset$:

\begin{align*}
h_{\alpha\alpha} &= Id_F \\
h_{\beta\alpha} &= (h_{\alpha\beta})^{-1} \\
h_{\alpha\beta} \circ h_{\beta\gamma} \circ h_{\gamma\alpha} &= Id_F \tag{1.2}
\end{align*}

The thing to notice about these three relations is that they are very similar to the three conditions for an equivalence relation. If we want to understand the transition maps it might therefore be a good idea to see if we could use them to define an equivalence relation. This is indeed the way to go and it is done as follows.

**Remark 1.1.2.** Instead of being given a bundle and looking at the transition maps themselves to find out what they do, we will examine the situation where we have a base space $X$ with an open covering $\bigcup U_\alpha = X$, a fibre $F$ and a set of maps satisfying (1.2), and use those data to define a bundle. First we define a space $\tilde{E}$ as the disjoint union of the products of the open sets with the fibre:

\[ \tilde{E} = \bigcup_{\alpha} \{(x, f, \alpha) \mid x \in U_\alpha, f \in F \} . \]  \tag{1.3}

Next, we define an equivalence relation on $\tilde{E}$ by

\[ (x, f, \alpha) \sim (x, h_{\alpha\beta}(x)(f), \beta) \]  \tag{1.4}

for all $x \in U_\alpha \cap U_\beta$ and for all $U_\alpha$ and $U_\beta$ with non-empty intersection. The fact that the maps $h_{\alpha\beta}$ satisfy (1.2) ensures that this is indeed an equivalence relation (check this yourself!). Now that we have an equivalence relation on $\tilde{E}$ we can define a new space $E$ as the quotient space of $\tilde{E}$ under $\sim$:

\[ E = \tilde{E} / \sim \]  \tag{1.5}

It can be proven without much difficulty that, if we take $E$ as the total space and define a projection and local trivialisations in a natural way, this will give us a fibre bundle. We will not do so here, but the reader can verify it for himself or herself.
Before going through all this we promised that it would give some insight into why the maps \( h_{\alpha\beta} \) are called the transition maps. Now that we have done some work, we can see that the transition maps are indeed accurately named; they describe the transition from looking at the fibre bundle locally at \( U_\alpha \) to looking at it locally at \( U_\beta \), whenever these two have nonempty intersection. Equation (1.4) describes exactly how we go from the one viewpoint to the other.

Another way to look at equation (1.4) (and hence at the transition maps) is to view it as gluing instruction for the total space \( \tilde{E} \). To understand this, let us look back at what we have done above. We constructed a space \( \tilde{E} \) consisting of products of open sets with the fibre, which are all totally disjoint from one another. Next, we defined an equivalence relation which related points in different products to each other, using the transition maps. Lastly, we defined our total space as the quotient of \( \tilde{E} \) under \( \sim \), thus taking equivalent points and making them identical. This last action could be visualised as gluing together the space \( E \) from the loose fragments of \( \tilde{E} \).

Fibre bundles are often constructed in exactly this way, by telling you how to glue together different products of open sets with the fibre; i.e. by giving the transition maps. A simple example of this is the Möbius strip.

**Example 1.1.3.** The Möbius strip can be constructed in the following way. Take the circle \( S^1 \) (seen as the circle in \( \mathbb{R}^2 \) in this example) as the base space and the interval \( F = [-1, 1] \) as the fibre. Take as an open cover of \( S^1 \) the sets \( U \) consisting of all points with second coordinate greater than \( -\frac{1}{2} \) and \( V \) consisting of all points with second coordinate smaller than \( \frac{1}{2} \). The intersection of these then exists of two parts; one to the left of the vertical axis (call it \( W_1 \)), and one to the right (\( W_2 \)). Now define the transition function \( h_{UV} : (U \cap V) \to \text{Aut}(F) \) so that \( h_{UV}(x)(i) = i \) for \( x \in W_1 \) and \( h_{UV}(x)(i) = -i \) for \( x \in W_2 \). The fibre bundle constructed from this data by the method described above is precisely the Möbius strip, as can be easily seen.

There is some additional terminology we will often use:

**Remark 1.1.4.** If \( F \longrightarrow E \longrightarrow X \) is a fibre bundle, we call \( F \) the abstract fibre of the bundle. This contrasts with the concrete fibre \( p^{-1}(x) \subseteq E \) ‘above’ a point in the base space \( x \in X \). In these terms, the trivialisation maps provide identifications of the concrete fibres with the abstract fibre.

**Definition 1.1.5.** A section of a fibre bundle \( F \longrightarrow E \longrightarrow X \) on a subset \( U \) of the base space \( X \) is a continuous map \( s : U \to E \), such that \( p \circ s = \text{id}_U \) (possible additional smoothness constraints on \( s \) depend on the context). In words, a section on \( U \) chooses for each \( x \in U \) an element of the fibre over \( x \). A global section is a section on the whole base space \( X \).

We could now go on by giving more examples of fibre bundles, defining what a morphism of fibre bundles is and inspecting what it means to be an isomorphism. However, since we will not need these in the remainder of the text we will not do so, but concentrate instead on those notions that will be important to us later on. The first of these is the concept of a \( G \)-bundle, which we explain in the next section.
1.2 $G$-bundles

Often it is useful to have some additional algebraic structure on a fibre bundle. An example of this is the notion of a $G$-bundle. If $G$ is a group, then a $G$-bundle is a fibre bundle together with a left action $\lambda$ of $G$ on the fibre such that the transition functions can be represented by elements of $G$. To be more precise: we must have that $h_{\alpha\beta}(x) \in \lambda(G)$ for all $x \in U_\alpha \cap U_\beta$ and for all $\alpha$ and $\beta$. Using multiplicative notation for the action: for every $x \in U_\alpha \cap U_\beta$ there must be an $g \in G$ such that $h_{\alpha\beta}(x)(f) = g \cdot f$.

Since we have, in a $G$-bundle, that $h_{\alpha\beta}(x) \in \lambda(G)$, we can make maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ such that $h_{\alpha\beta}(x) = \lambda(g_{\alpha\beta}(x))$. If we take the action to be effective (which can always be done by looking at $G/N$ instead of $G$, where $N$ is the normal subgroup of elements that act trivially), these maps are unique; we will then also speak of them as transition maps. It is easy to see from 1.2 that these maps satisfy

\[
g_{\alpha\alpha}(x) = e \quad (1.6)
\]
\[
g_{\beta\alpha}(x) = g_{\alpha\beta}(x)^{-1} \quad (1.7)
\]
\[
g_{\alpha\beta}(x)g_{\beta\gamma}(x)g_{\gamma\alpha}(x) = e \quad (1.8)
\]

**Remark 1.2.1.** There is a slightly different point of view: if we are given a $G$-bundle, we say that $G$ is the structure group of this bundle. Note, however, that this is somewhat sloppy language, since there are in general many possible structure groups for a given bundle. (For example, every bundle with fibre $F$ is an $\text{Aut}(F)$-bundle.) It would be more precise in this case to say that the bundle can be made into a $G$-bundle. Alternatively, we could change the definition of ‘$G$-bundle’, to make the group $G$ part of its data.

On a related note, suppose that we are given a $G$-bundle. One may wonder whether we really need all of the elements of $G$ for the transition maps; perhaps it is also a $H$-bundle, for some subgroup $H$ of $G$. This can be non-obvious: it may be necessary to pick a very specific covering of the base space, or transition maps that differ from the original ones. If the bundle is in fact also a $H$-bundle, this insight is called reduction of the structure group (from $G$ to $H$).

We shall see more examples in the section on principal $G$-bundles.

1.3 $G$-torsors

$G$-torsors are quite common in physics. Usually, their presence is not made explicit and their properties are described in an intuitive way only.

Let us first give a few examples.

**Example 1.3.1.** The energy of a system is not a well-defined quantity: only energy differences can be measured (and thereby have physical significance)$^2$. One may agree to elect a certain state of the system as ‘ground state’ and declare it to have zero energy. The energy of other states is then defined to be the energy difference with the ground state (just a real number).

However, which state qualifies as ground state depends on the situation and may be open to discussion.

\footnote{For those who are not familiar with the concept of an action it is explained in section 1.3.}

\footnote{In general relativity, this is no longer true.}
Example 1.3.2. Consider positions in our three-dimensional world. It does not make sense to add two such positions. What would be the sum of the summit of Kilimanjaro and the centre of the moon? We can subtract positions, though, obtaining something we might call a displacement: a vector having both a length and a direction. Furthermore, it does make sense to add two displacements (as vectors in 3-space).

After choosing an origin, positions are conveniently specified by giving the displacement between the position and the origin. This does introduce a certain danger, though: you might be tempted to add two positions (after all, now that we chose an origin, they are just vectors), even though we agreed that that is a meaningless thing to do.

Example 1.3.3. A more mathematical example concerns the many possible bases of a vector space. Let’s confine ourselves to a finite dimensional one $V$, say of dimension $n$. We consider the set $B(V)$ of all ordered bases of $V$. Given such a basis and an element of $GL(n)$ (the set of invertible $n \times n$-matrices), we can construct a new basis by taking respective linear combinations (specified by the columns of the matrix) of the old basis vectors. This construction has some nice properties:

- If we first apply a matrix $A$ to a basis, and then apply another matrix $B$ to the outcome, we get the same as when we apply the product $AB$ to the first basis.
- Given two ordered bases of $V$, there is a unique matrix in $GL(n)$ so that the matrix applied to the first basis gives the second one (we might call it the quotient of the second basis by the first).

It does not make sense to multiply two bases. However, once we pick some basis as ‘origin’, we can identify other bases with their quotients by the origin (elements of $GL(n)$).

The general pattern will be clear. The mathematical description is as follows:

Definition 1.3.4.

A left action of a group $G$ on a set $X$ is a map $\mu : G \times X \to X : (g, x) \mapsto g \cdot x$ such that

1. $(gh) \cdot x = g \cdot (h \cdot x)$ for all $g, h \in G$ and $x \in X$;
2. $e \cdot x = x$ for all $x \in X$, with $e$ the identity element of $G$.

A right action of a group $G$ on a set $X$ is a map $\nu : X \times G \to X : (x, g) \mapsto x \cdot g$ such that

1. $x \cdot (gh) = (x \cdot g) \cdot h$ for all $g, h \in G$ and $x \in X$;
2. $x \cdot e = x$ for all $x \in X$, with $e$ the identity element of $G$.

Alternatively, we can curry the map $\mu : G \times X \to X$ to get a map $\lambda : G \to \text{End}(X)$ (where $\text{End}(X)$ is the monoid of functions from $X$ to itself). The reader should check that $\mu$ is a left action if and only if the image of $\lambda$ is contained in $\text{Aut}(X)$ (the group of bijections from $X$ to itself) and $\lambda$ is a group homomorphism.
A set together with a left (right) action of a group $G$ on it is sometimes called a left (right) $G$-set. It is easy to check that left actions of $G$ on $X$ correspond to right actions of $G^{\text{op}} \cong G$ on $X$. If $\mu : G \times X \to X$ is a left action, then $\nu : X \times G \to X : (x, g) \mapsto \mu(g^{-1}, x)$ is a right action. The following definitions regarding left actions can be extended by declaring a right action to have a property if and only if its corresponding left action has it.

- If $G$ acts on $X$ and $x$ is a point of $X$, then the orbit of $x$ is the set $G \cdot x = \{ g \cdot x ; g \in G \}$ of points that can be reached by acting on $x$. One element being in another element’s orbit is an equivalence relation, so the set $X$ is partitioned in disjunct orbits.

- An action $\mu : G \times X \to X$ is transitive, if for all $x, y \in X$ there exists some $g \in G$ that sends $x$ to $y$: $g \cdot x = y$. Equivalently, $\mu$ is transitive iff $X$ has exactly one orbit (namely $X$ itself).

- It is called free, if different elements of $G$ act differently on every point of $X$: $g \cdot x = h \cdot x \Rightarrow g = h$.

- Now, a $G$-torsor is a set $X$ together with a free transitive action of $G$ on $X$. $X$ is then also called a homogeneous space for $G$. The unique element $g$ such that $g \cdot x = y$ is called the quotient of $y$ by $x$ and denoted by $y / x$ (in case of a right action, we denote the element $g$ such that $x \cdot g = y$ by $x \backslash y$).

You can think of a $G$-torsor as a copy of $G$ that has forgotten which of its elements is the identity. As soon as we choose a $x_0 \in X$, we can identify $X$ with $G$ by sending $x$ to $x / x_0$.

Remark 1.3.5. Typically, if $G$ is a topological group and $X$ a topological space, we demand that the map $\mu : G \times X \to X$ is continuous, and then say that the action is continuous. Analogously, if $G$ is a Lie group, $X$ a smooth manifold and $\mu$ a smooth map, then the action is said to be smooth as well.

### 1.4 Principal $G$-bundles

**Definition 1.4.1.** A principal $G$-bundle is a $G$-bundle $F \xrightarrow{p} E \xrightarrow{\pi} X$ with a right action of $G$ on $F$ and a right action of $G$ on $E$, such that

1. the orbits of the action on $E$ are precisely the fibres of the bundle, and every fibre is a $G$-torsor (that is, restricting the action of $G$ to a fibre makes that fibre a $G$-torsor);

2. for every trivialisation $h_U : p^{-1}(U) \to U \times F$, and every $a \in p^{-1}(U)$ and $g \in G$, if $h_U(a) = (x, f)$

   \[ h_U(a \cdot g) = (x, f \cdot g) \quad (1.10) \]

   holds. Note that in the last equation the left-hand side features the action of $G$ on $E$, while the right-hand side mentions the action of $G$ on $F$.\footnote{$G^{\text{op}}$, the opposite of $G$, is a group with the same underlying set as $G$, but with multiplication reversed: $a \ast^{\text{op}} b = b \ast a$. $G^{\text{op}}$ and $G$ are isomorphic via $a \mapsto a^{-1}$.}
Remarks.

• Explicitly, condition 1 means that for \( x, y \in E \) the equation \( x \cdot g = y \) can be solved for \( g \in G \) only if \( p(x) = p(y) \) and that it then has a unique solution.

• Note that in addition to the local triviality of a general fibre bundle, condition 2 demands that the right action of \( G \) on the principal bundle is locally ‘trivial’ as well: on a chart, it must be given by the action of \( G \) on the abstract fibre \( F \).

• Since the action on a fibre is free and transitive, the action of \( G \) on \( F \) is free and transitive as well: \( F \) is a \( G \)-torsor. Because of this, many authors define a principal \( G \)-bundle to have \( G \) itself as abstract fibre instead of some \( F \), and replace the action of \( G \) on \( F \) by right multiplication of \( G \) on itself. This is no restriction (for all \( G \)-torsors are isomorphic to \( G \) itself), but it sometimes imposes an inconvenient (because unnatural) choice.

• Global sections on a principal \( G \)-bundle \( E \xrightarrow{p} X \) can be hard to come byootnote{In this instance, it is convenient to have \( G \) as the abstract fibre, so we will assume this.}: they exist if and only if the bundle is trivial! The ‘if’ part is obvious: just pick an element \( g \in G \), take the global section \( x \mapsto (x, g) \) of the trivial bundle and use a global trivialisation to make this into a global section of \( p \). The ‘only if’ part goes as follows: let \( s : X \to E \) be a global section. We can make a global trivialisation \( h : E \to X \times G \) by defining \( h^{-1}(x, g) = s(x) \cdot g \) (this is invertible; \( h \) can be given explicitly as \( h(e) = (p(e), s(p(e)) \backslash e) \)), so the bundle is trivial.

Example 1.4.2. A major example of a principal \( G \)-bundle is the so called frame bundle of a vector bundle. We will treat it together with vector bundles in 1.5.

Example 1.4.3. The map \( p : S^1 \to S^1 : z \mapsto z^2 \) (where we see \( S^1 \) as the unit complex numbers) is a fibre bundle with a two-point fibre. We can make it a principal \( \mathbb{Z}/2\mathbb{Z} \)-bundle by letting the non-identity element \( 1 \in \mathbb{Z}/2\mathbb{Z} \) act by switching on the abstract fibre and by inversion \((z \mapsto -z)\) on \( E = S^1 \). The latter action preserves the fibres (since \((-z)^2 = z^2\)) and acts freely and transitively on them (since all fibres are of the form \( \{z, -z\} \)). Condition 2 is also satisfied: the case \( g = 0 \) is trivial, while in the case \( g = 1 \) both \( h_U(a \cdot g) = h_U(-a) \) and \( (x, f \cdot g) \) are the other element (other than \( h_U(a) = (x, f) \)) of the (two-element) fibre above \( x \). If you take a covering of the base space by two connected open proper subsets \( U \) and \( V \) and examine the transition maps, you will see that this fibre bundle is much like the Möbius strip of example 1.0.2: at one component of the intersection \( U \cap V \), the fibres are connected in the trivial manner, while in the other component, they are switched (twisted). In a sense, this bundle and the Möbius strip have different fibres, but are twisted in same way. The next definition makes this intuition precise.

Definition 1.4.4. Given a principal \( G \)-bundle \( E \xrightarrow{p} X \) and a right action \( \mu \) of \( G \) on some set (or topological space, or smooth manifold) \( A \), we construct a new \( G \)-bundle \( \tilde{A} \xrightarrow{\bar{p}} E \times_\mu A \xrightarrow{p} X \), the associated bundle.
Intuitively, we would like to replace the fibre $G$ by $A$, and obtain the transition maps by composing the old ones with the action of $G$ on $A$. In order to do so, we first define a right action of $G$ on $E \times A$ by
\[(e, a) \cdot g := (e \cdot g, a \cdot g). \tag{1.11}\]
We let the total space $E \times_{\mu} A$ of the associated bundle be the quotient of $E \times A$ by the action of $G$ (so points of $E \times_{\mu} A$ are orbits of that action). Since the action of $G$ on $E$ preserves the fibres, we can define the projection $\tilde{p} : E \times_{\mu} A \to X$ by
\[\tilde{p}([[e, a]]) := p(e). \tag{1.12}\]
The abstract fibre $\tilde{A}$ of the associated bundle is defined by $\tilde{A} := (F \times A)/G$, so points of $\tilde{A}$ are orbits of the logical right action of $G$ on $F \times A$ (the action on $F$ is the one of the original (principal) bundle, and the action on $A$ is of course $\mu$). This fibre is of course homeomorphic to $A$, and even isomorphic to it as a right $G$-set, but – once more – not in a canonical way, because $F$ cannot be canonically identified with $G$. For every trivialisation $h : p^{-1}(U) \to U \times F$ of $E$, we define a trivialisation $\tilde{h} : \tilde{p}^{-1}(U) \to U \times \tilde{A}$ on $E \times_{\mu} A$ by
\[\tilde{h}([[e, a]]) := (p(e), [h_2(e), a]) \tag{1.13}\]
, where $h_2 : p^{-1}(U) \to F$ is the second component of $h$.

Example 1.4.5. The motivating example is formulated in this language as follows: take the principal $\mathbb{Z}/2\mathbb{Z}$-bundle of example 1.4.3 as the starting point. Let $A$ be the unit interval $I = [0, 1]$, and define the action $\mu$ of $\mathbb{Z}/2\mathbb{Z}$ on $I$ by letting $1$ act by reflection $x \mapsto 1 - x$. Then the associated bundle $S^1 \times_{\mu} I$ is isomorphic to the Möbius strip.

Example 1.4.6. If the new fibre $A$ is a vector space and the action is linear (so $\mu$ is a representation of $G$ on $A$), then the associated bundle is what we call a vector bundle (see the next section). This has a great physical significance: in the language of gauge theory, the ‘charge’ of a particle is given by a representation of the ‘gauge group’ $G$, and the associated bundle construction describes how the presence of a ‘gauge field’, which lives on a principal $G$-bundle, affects the particle in question (which then lives on a bundle with fibre $A$).

1.5 Vector bundles

Another way to add additional structure to the concept of a fibre bundle is to look at bundles where fibres have some structure in addition to being a topological space. One instance of this is the concept of a vector bundle. This is a fibre bundle in which the abstract fibre is also a vector space. Of course, for this to be in any way useful, we need the transition functions to respect this structure; thus for a vector bundle we furthermore require that each $h_{\alpha\beta}(x) : F \to F$ is a linear isomorphism.

Consequently, in a vector bundle the concrete fibre $F_x = p^{-1}(x)$ above a point $x$ can be made into a vector space in a meaningful way. That is, for $(x, f_1, \alpha) \sim (x, h_{\alpha\beta}(x)(f_1), \beta)$, $(x, f_2, \alpha) \sim (x, h_{\alpha\beta}(x)(f_2), \beta)$ and scalars $a_1$ and $a_2$ we have $(x, a_1 f_1 + a_2 f_2, \alpha) \sim (x, a_1 h_{\alpha\beta}(f_1) + a_2 h_{\alpha\beta}(f_2), \beta)$, following directly
from the fact that \( h_{\alpha\beta} \) is a homomorphism and implying that we can define addition and scalar multiplication on \( F_x \) by the same operations in a trivialisation.

A well known example of a vector bundle is the **tangent bundle** of a smooth manifold \( M \). Take as base space the manifold, and take as the concrete fibre over \( x \in M \) the tangent space \( T_x M \). The projection is of course the map taking the whole of \( T_x M \) to \( x \) and the trivialisations can be constructed from the charts on your manifold. The abstract fibre can be taken to be \( \mathbb{R}^n \), where \( n \) is the dimension of \( M \).

Given a vector bundle \( V \rightarrow W \rightarrow X \) of dimension \( n \), the **frame bundle** of \( W \) is a principal \( \text{GL}(n) \)-bundle \( B(V) \rightarrow E \rightarrow X \) defined as follows:

- the base space of the frame bundle is \( X \), the base space of the vector bundle;
- the fibre \( q^{-1}(x) \) above a point \( x \in X \) is the set of all ordered bases of the vector space \( p^{-1}(x) \) (the corresponding fibre of the vector bundle), so the total space \( E \) is the set of all bases of all fibres of \( W \);
- for every trivialisation \( h : p^{-1}(U) \rightarrow U \times V \) above an open set \( U \subset X \), we construct a trivialisation \( \tilde{h} : q^{-1}(U) \rightarrow U \times B(V) \). Given \( C \in q^{-1}(U) \), say \( q(C) = x \), \( C \) is a basis of the vector space \( p^{-1}(x) \). By applying \( h \) to these basis vectors, we obtain a basis \( D \) of \( V \) (since \( h \) restricted to a fibre is an isomorphism of vector spaces). So, define \( \tilde{h}(C) = (x, D) \in U \times B(V) \);
- the right action of \( \text{GL}(n) \) on \( E \) (restricted to a fibre) the same as the one in example 1.3.3 of 1.3: if \( C \in E \) is a basis and \( A \in \text{GL}(n) \), we define \( C \cdot A \) to be the basis obtained by taking linear combinations of the vectors of \( C \) with coefficients as specified by the columns of \( A \).

This construction is most commonly applied to the tangent bundle \( TM \) of a smooth manifold \( M \). It is then called simply the frame bundle of \( M \). A section of this frame bundle is called a **frame**; it is a choice of basis of the various tangent spaces to \( M \).

By the way, a one-dimensional vector bundle is called a **line bundle**. But beware, this word is also used for one-dimensional *complex* vector spaces.
Chapter 2

Connections

Let us go back to our favourite example of a fibre bundle: the Möbius strip (call the total space $E$ for convenience). The tangent space $T_e E$ at some point $e$ is obviously isomorphic to $\mathbb{R}^2$ for all points $e$. It is intuitively clear what we would mean by the vertical subspace $V_e E \subset T_e E$; it is the space of all vectors tangent to the fibre of $e$ (we will be more precise in a moment, for now just try to see things in a picture). What is not clear, is what we would mean by a horizontal subspace $H_e E \subset T_e E$. There is just no obvious way of defining such a subspace. This is where the notion of a connection comes in; it is a choice of the horizontal subspace of $T_e E$. Note that since there is no canonical horizontal subspace of the tangent space, we have in general many different possible connections on a fibre bundle; we can really choose what we want ‘horizontal’ to mean on a fibre bundle.

Connections have many applications. For physicists their main use might be in the ability to define a covariant derivative (see section 2.5). It turns out that many field theories can be described naturally by viewing the field as a connection on a certain fibre bundle. We will show this explicitly for the case of electromagnetism (section 3.1).

2.1 Connections

In the introduction to this chapter we said that it was intuitively clear what we would mean by the vertical subspace $V_e E$ of the tangent space $T_e E$ at some point on the Möbius strip; it was the subspace tangent to the fibre of the point $e$. Now let us define in a more precise manner what the vertical subspace $V_e E$ of $T_e E$ is for an arbitrary fibre bundle $F \rightarrow E \rightarrow M$ in which all spaces are manifolds: it is the subspace of all vectors $\eta \in T_e E$ such that $dp(\eta) = 0$. If you think about the example of the Möbius strip this does exactly what we want; $dp$ sends a vector to zero if and only if it is tangent to the fibre.

As said, however, we have no canonical way of choosing a horizontal subspace. Instead we use a connection to define what vectors are horizontal. In other words; a connection is a smooth (we will specify what we mean by this in a moment) choice of a horizontal subspace $H_e E$ of $T_e E$ at each $e \in E$, such that $H_e E$ is complementary to $V_e E$ (as should clearly be the case for the horizontal subspace). Such a connection then gives us projections $\pi^e_v$ and $\pi^e_h$ from the
tangent space onto the vertical and horizontal subspaces respectively (since $V_eE$ and $H_eE$ are complementary to each other, there is a unique decomposition of any tangent vector $\eta = \eta^v + \eta^h$ as the sum of a vertical and a horizontal vector, and these $\eta^v$ and $\eta^h$ are the projections of $\eta$ onto $V_eE$ and $H_eE$). Now we can specify what we mean by a smooth choice; for any smooth vector field $X$ on $E$ the horizontally projected field $\pi^hX$ should also be smooth.

**Remark 2.1.1.** In order to define even the vertical projection, we needed the choice of horizontal subspace. You might be tempted to think this unnecessary:

![Figure 2.1: Vertical projection needs choice of horizontal subspace.](image)

we could have taken the *orthogonal* projection onto the vertical subspace. The horizontal subspace is then automatically the orthogonal complement of the vertical one. However, you need an inner product to even talk about orthogonality.

A manifold equipped with the extra structure of an inner product on the tangent space at each point is called a pseudo-Riemannian manifold. (There are two more requirements: the choice of inner product at all the different points must vary smoothly with the point, and all inner products must be non-degenerate.) So, if the total space of a fibre bundle is a pseudo-Riemannian manifold, this gives a natural choice of connection on the bundle.

Well, so much for the definition; let us look at some examples.

**Example 2.1.2.** Take again the Möbius strip. How could we define a connection on it? One way to do this is to embed the Möbius strip in $\mathbb{R}^3$ and then take the horizontal subspace to be the subspace orthogonal to the vertical one, which is well defined because in $\mathbb{R}^3$ we have an inner product. It is complementary to the vertical subspace by definition and that it is smooth is clear intuitively, but can be checked by introducing a specific parametrisation of the Möbius strip and looking at it in local coordinates.

**Example 2.1.3** (Connection on a covering). Consider a fibre bundle $F \longrightarrow E \longrightarrow X$ with $F$ a discrete space (a 0-dimensional manifold). (Note that fibre bundles with discrete fibres are just covering maps, so $p$ is a smooth covering map.) Then the vertical tangent space is 0-dimensional (just like the fibre), so any horizontal tangent space must be the entire tangent space: there is only one possible connection.
Now that we have seen some examples, let’s look at the concept more closely: what happens in a trivialisation?

First the tangent space. In the following let \( e \) be a point in \( p^{-1}(U) \) and let \( (x,f) = h_U(e) \) be its representation in \( U \times F \). We have, for the tangent space of a trivialisation, that \( T_{(x,f)}(U \times F) \cong T_x U \times T_f F \). Now, the horizontal and vertical subspaces of \( T_x E \) are carried over to the trivialisation by the isomorphism \( dh_U|_e : T_e(p^{-1}(U)) \cong T_e E \rightarrow T_{(x,f)}(U \times F) \). The vertical subspace in the trivialisation is easy to describe: since the following two diagrams commute

\[
\begin{array}{ccc}
p^{-1}(U) & \overset{h_U}{\rightarrow} & U \times F \\
p & \downarrow & \downarrow \text{id} \\
T_e T^{-1}(U) & \overset{dh_U}{\rightarrow} & T_{(x,f)}(U \times F) \\
pr_1 & \downarrow & \downarrow \text{id} \\
T_x U & \overset{dpr_1}{\rightarrow} & T_{(x,f)} \\
\end{array}
\]

(the first states that the trivialisation respects the projection (it is one of the conditions in the definition 1.1.1 of a fibre bundle); the second follows from the first by taking derivatives), we have \( dp(v) = (dpr_1 \circ dh_U)(v) \). A vector \( (\xi,y) \in T_x U \times T_f F \) which can of course be written as \( dh_U(v) \) for some \( v \in T_x E \) is vertical if and only if \( 0 = dp(v) = (dpr_1 \circ dh_U)(v) = dpr_1((\xi,y)) = \xi \).

We are now in a position to conclude what the vertical projection must look like in a trivialisation. If \( (\xi,y) \) is again a vector in \( T_{(x,f)}(U \times F) \), then (realising that the projections are linear maps):

\[
\pi^v_{(x,f)}(\xi,y) = (0, \tilde{\Gamma}(x,f)y + \Gamma(x,f)\xi)
\]

for some maps \( \tilde{\Gamma} : U \times F \rightarrow \text{Lin}(T F) \) and \( \Gamma : U \times F \rightarrow \text{Lin}(TU,TF) \). But we can do better than that, because we know that if a vector is vertical, then the vertical projection should leave it intact. So \( (0,y) \) should be sent to itself for all \( y \), giving that \( \tilde{\Gamma}(x,f) \) must be the identity for all \((x,f) \in (U \times F)\):

\[
\pi^v_{(x,f)}(\xi,y) = (0, y + \Gamma(x,f)\xi)
\]

This is all we can tell without actually specifying the horizontal subspace (choosing the connection). When a horizontal subspace is designated, that choice determines the vertical projection and thereby the map \( \Gamma \). As a further useful fact, we know that a vector has vertical projection zero if and only if it is itself horizontal, so we can conclude that the vector \((\xi, -\Gamma(x,f)\xi)\) is horizontal for all \(\xi \in T_x U\) (and any horizontal vector can be written in this way). The horizontal and vertical projection are complementary, so \( \pi^v_{(x,f)}(\xi,y) + \pi^h_{(x,f)}(\xi,y) = (\xi,y) \). This immediately gives us an equation for the horizontal projection in terms of \( \Gamma \):

\[
\pi^h_{(x,f)}(\xi,y) = (\xi, -\Gamma(x,f)\xi)
\]

That is about all we can say about the projections in this trivialisation; to make things interesting it’s a good plan to see what happens if we look at them from another trivialisation. To do this, define a map

\[
\psi(x,f) = (x, h_{VU}(x)(f))
\]
from \((U \cap V) \times F\) to \((U \cap V) \times F\) for two trivialisations \(h_U\) and \(h_V\). (Thus \(\psi\) is nothing other than \(h_V \circ h_U^{-1}\), as you can directly see in the definition of a fibre bundle (see 1.1.1).) We can now show that \(d\psi\) intertwines the local representatives of the vertical projection, meaning that the following relation holds:

\[
d\psi \circ \pi^v_{(x,f)} = \pi^v_{\psi(x,f)} \circ d\psi.
\]

Intuitively speaking, to project at \(U\) and then go over to \(V\) should be the same as first going over to \(V\) and then projecting there (in other words: the vertical projections in different trivialisation should be ‘compatible’ with one another). This is reasonable, because the vertical projections in different trivialisations are by definition the local representatives of the true vertical projection (in the tangent space of the fibre bundle), which is defined without reference to any trivialisation.

To prove (2.5), write the vertical projection \(\pi^v\) in terms of the respective local vertical projections:

\[
\pi^v = dh_U^{-1} \circ \pi^v_{(x,f)} \circ dh_U = dh_V^{-1} \circ \pi^v_{\psi(x,f)} \circ dh_V.
\]

Now pre-compose with \(dh_U^{-1}\) and post-compose with \(dh_V\):

\[
dh_V \circ dh_U^{-1} \circ \pi^v_{(x,f)} = \pi^v_{\psi(x,f)} \circ dh_V \circ dh_U^{-1}.
\]

This is the desired result, once we note that \(h_V \circ h_U^{-1} = \psi\).

We now write out (2.5) using the explicit formulas for the projections we derived above and the expression for \(d\psi\) in terms of its components:

\[
d\psi((\xi,y)) = (\xi, d_1 h_{UV}(\xi) + d_2 h_{UV}(y))
\]

where \(d_1 h_{UV}\) denotes the derivative of \(h_{UV}\) with respect to the first argument (in the base space) and \(d_2 h_{UV}\) the derivative of \(h_{UV}\) with respect to the second argument (in the fibre).

Performing those substitutions gives us:

\[
(0, d_2 h_{UV}(y + \Gamma_{U}(x,f)(\xi)))
= (0, d_1 h_{UV}(\xi) + d_2 h_{UV}(y) + \Gamma_V(x, h_{UV}(x)(f))(\xi)).
\]

Ignoring the left component and cancelling a common term \(d_2 h_{UV}(y)\), we get:

\[
d_2 h_{UV} \circ \Gamma_{U}(x,f) = d_1 h_{UV} + \Gamma_V(x, h_{UV}(x)(f)).
\]

We can solve this for \(\Gamma_V\) by replacing \(f\) with \(h_{UV}(x)^{-1}(f)\), resulting in:

\[
\Gamma_V(x,f) = d_2 h_{UV}(\Gamma_U(x, h_{UV}(x)^{-1}(f))) - d_1 h_{UV}.
\]

Now this may seem like a lot of work just to get some complicated formula relating the connection maps \(\Gamma(x,f)\) in different trivialisations, but it is an important observation that given maps \(\Gamma_U : U \times F \to \text{Lin}(TU,TF)\) satisfying (2.9) there is a unique connection having those maps as its connection maps (this
will later allow us to construct connections representing the electromagnetic field. This construction is very similar to that of remark 1.1.2: we can form a connection (c.q. fibre bundle) with given connection maps (c.q. transition maps) if they obey a certain compatibility relation. We do not prove this result; the reader should be able to do so quite easily (transfer the alleged horizontal subspace \( \{ (\xi, -\Gamma_U(x,f)\xi) \mid \xi \in T_xU \} \) from the trivialisation to the fibre bundle by \( h_U \), and show that the result of this procedure is the same for overlapping trivialisations).

Well now, there is a lot of freedom in the choice of connection: one may choose horizontal tangent spaces mostly independently (only subject to smoothness demands) in different points of the total space, even within a fibre. For many applications featuring additional structure on the fibre, this is too much freedom – we need the connection to behave itself within a fibre, with respect to this structure. We will encounter two instances of this principle. On a principal \( G \)-bundle, there is the notion of an invariant connection; as an example of its use, the gauge field (the electromagnetic field, in the case we describe in chapter 3) is described by an invariant connection on the gauge group. We review invariant connections in two subsections time. First, we examine linear connections on a vector bundle.

### 2.1.1 Linear connections

Let us now study the notion of a connection in a context where we have some additional structure; in this case a vector bundle. First we remark that there is a canonical identification \( V_e E \sim F_{p(e)}(e) \); for \( q \in F_{p(e)}(e) \) define a curve \( \gamma_q(t) = e + tq \in F_{p(e)} \). Then identification of \( q \) with \( [\gamma_q] \) gives an isomorphism \( F_{p(e)}(e) \sim V_e E \).

**Definition 2.1.4.** A connection is called **linear** if the map \( e \mapsto \pi^h_e \) is affine and the canonical zero section of \( E \) is horizontal (by which we actually mean that the space tangent to the zero section is everywhere horizontal).

Using the canonical identification mentioned above, we can represent \( \pi^h_e : T_{(\cdot,\cdot)}(U \times F) \to H_{(\cdot,\cdot)}(U \times F) \) in a trivialisation by a map \( \tilde{\pi}^h_e : TU \times F \to TU \times F \). The condition that the map \( e \mapsto \pi^h_e \) be affine for the connection to be linear now gives us that we can write (omitting the \( x \)-dependence, as we are working within a single fibre):

\[
\pi^h_e(\xi, y) = L(f)(\xi, y) + c(\xi, y)
\]  

(2.10)

for some linear map \( L \) (from the fibre to the space of linear operators on the tangent space of the trivialisation in \( (x, f) \)), and some constant linear operator \( c \). From the horizontality of the zero section (call it \( \sigma \), and \( \tilde{\sigma} \) in a trivialisation) we can derive

\[
\tilde{\sigma}(x) = (x, 0) \quad \text{ (because } \sigma \text{ is the zero section} )
\]

\[
\Rightarrow d\tilde{\sigma}(\xi) = (\xi, 0) \quad \text{ (because } T_{(x,0)}(\sigma(U)) = d\tilde{\sigma}(T_xU) \text{) }
\]

\[
\Rightarrow H_{(x,0)}E = d\tilde{\sigma}(T_xU) = T_xU \times 0 \quad \text{ (also using equation (2.3))}
\]

\[
\Rightarrow \pi^h_e(\xi, y) = (\xi, 0)
\]
If we substitute this result into equation (2.10), we get (using the linearity of $L$)

$$(\xi, 0) = \pi_0^1(\xi, y) = L(0)(\xi, y) + c(\xi, y) = c(\xi, y)$$

which gives us a simple expression for $c$. From equations (2.3) and (2.10), using this expression for $c$, we can then find an equation for $\Gamma$:

$$(0, -\Gamma(x, f)\xi) = L(f)(\xi, y)$$

and since $L$ is linear in $f$, $\Gamma$ must be linear in $f$ too. This means, that if we are working in the context of a vector bundle with a linear connection defined on it we can write:

$$\Gamma(x, f)\xi = \Gamma(x)(\xi)f$$

where $\Gamma(x)(\xi)$ is a linear map from the fibre to its tangent space (varying linearly in $\xi$, of course). This also gives us a new way to write equation (2.9) for the specific case of a linear connection (we will leave out the arguments of the maps here, for brevity):

$$\Gamma_V = h_V U h_V^{-1} - dh_V h_V^{-1}$$

(2.12)

This is all we need to know about linear connections, for now.

### 2.1.2 Invariant connections

As promised, we introduce another type of well-behaved connection, namely an **invariant** connection on a principal $G$-bundle $PG$, where $G$ is a Lie group. Before we define anything, let’s study this situation a bit. We have a canonical identification $V_e PG \cong g$ (where $g$ is the associated Lie algebra of $G$) by sending a point $L \in g$ to the equivalence class $L_e$ of the curve $\gamma(t) = e \cdot \exp(tL)$ in $V_e PG$ (this is clearly an identification if we keep the definition of $g$ (as tangent space to $G$ at $e$) in mind). We then have that the vertical projection $\pi_v^e$ can for all $e \in PG$ be represented by a map $\omega_e : T_e PG \to g$. We shall call the map $\omega$, which actually is a $g$-valued 1-form, the **connection 1-form** (note that, although the vertical tangent space does not depend upon the connection, the vertical projection, and thus $\omega_e$, do (see Remark 2.1.1)). Note that $\omega$ should obviously satisfy

$$\omega_e(L_e) = L$$

(2.13)

a result we shall use later on in this section.

Now for the promised definition, we have:

**Definition 2.1.5.** A connection is called **invariant** if all the horizontal subspaces $H_e E$ satisfy $H_e g E = dR_g(H_e E) = H_e E \cdot g$, where $R_g$ is the map $e \mapsto e \cdot g$ (the second equality is always true1, the constraint is in the first equality).

If a connection is invariant this immediately has a nice consequence: if at $e$ we have a decomposition of a tangent vector $\eta = \eta^h + \eta^v$ into horizontal and vertical components, then the decomposition at the point $e \cdot g$ is given by $\eta \cdot g = \eta^h \cdot g + \eta^v \cdot g$. That both sides of the latter equality are in fact equal is of course independent of the connection being invariant or not; it follows immediately from the fact that $dR_g$ is linear (see footnote 1 on page 19). We

---

1 Actually, the second equality can be read as the definition of a convenient notation: writing “$e \cdot g$” for “$dR_g(e)$.”
need the invariance of the connection to make sure that \( \eta^h \cdot g \) is again horizontal and \( \eta^v \cdot g \) is again vertical. Actually, even for the verticality we do not need any assumption about the connection; since \( R_g \) preserves the fibre, a vertical vector in the tangent space will stay vertical under \( dR_g \) (remember that the vertical tangent space is the same as the space tangent to the fibre). So all we need to show is that \( \eta^h \cdot g \) is horizontal. But \( \eta^h \cdot g \) is an element of \( H_e E \cdot g \) (by horizontality of \( \eta^h \)) and thus, by invariance of the connection, it is an element of \( H_e g E \), which is what we wanted. Note that this decomposition of \( \eta \cdot g \) also gives that

\[
\pi^v_{e \cdot g}(\eta \cdot g) = \pi^v_{e}(\eta) \cdot g \tag{2.14}
\]

We now want to find an expression for \( \omega_{e \cdot g}(\eta \cdot g) \). By the equation above and the definition of \( \omega \) we have that

\[
\omega_{e \cdot g}(\eta \cdot g) = \omega_{e}(\eta) \cdot g \tag{2.15}
\]

The right-hand side of (2.14) is of a general form for which we can derive

\[
L_e \cdot g = \frac{d}{dt} e \cdot \exp(tL)\bigg|_{t=0} = \frac{d}{dt} e \cdot g(g^{-1}\exp(tL)g)\bigg|_{t=0} = (\text{Ad}_{g^{-1}}L)_{e \cdot g},
\]

where the last equality follows directly from the definition of \( \text{Ad} \) as a map on the Lie algebra \( g \). Using this result on equation (2.14) we get:

\[
\pi^v_{e \cdot g}(\eta \cdot g) = (\text{Ad}_{g^{-1}}\omega_{e}(\eta))_{e \cdot g}.
\]

Applying the above-mentioned identification on both sides of this equation:

\[
\omega_{e \cdot g}(\eta \cdot g) = \text{Ad}_{g^{-1}}\omega_{e}(\eta) \tag{2.16}
\]

which is the desired expression for \( \omega_{e \cdot g} \).

As usual it is interesting to consider these notions in a trivialisation. Let therefore \( \omega^U \) be defined on \( U \times G \) as \( \omega^U = (h_U^{-1})^* (\omega) \). We will furthermore use the identification (by \( dR_{h^{-1}} \))

\[
T_{(x,h)}(U \times G) \cong T_x U \times T_h G \cong T_x E \times g
\]

If now \( \eta = (\xi, L) \in T_x U \times g \), then we have \( \eta \cdot g = (\xi, \text{Ad}_{g^{-1}}L) \) and furthermore, using equation (2.13), that \( \omega^U_{(x,h)}(\xi, L) = \omega^U_{(x,h)}(\xi, 0) + \omega^U_{(x,h)}(0, L) = \omega^U_{(x,h)}(\xi, 0) + L \). Using (2.16) and the fact that \( (x, e_G) \cdot h = (x, h) \) (where \( e_G \in G \) is the unit element of the group \( G \), not to be confused with the point \( e \in E \)), we get

\[
\omega^U_{(x,h)}(\xi, h^{-1} \cdot L \cdot h) = h^{-1} \cdot \omega^U_{(x,e_G)}(\xi, L) \cdot h = h^{-1} \cdot (\omega^U_{(x,e_G)}(\xi, 0) + L) \cdot h
\]

Taking \( L = 0 \) then gives us that we can define a \( g \)-valued 1-form \( A_U \) on \( U \), called the local principal gauge potential, such that \( \omega^U_{(x,h)}(\xi, 0) = h^{-1} \cdot A_U(x)(\xi) \cdot h \) (just take \( A_U(x) = \omega^U_{(x,e_G)}(\xi, 0) \)). Re-entering this into the equation above finally gives us

\[
\omega^U_{(x,h)}(\xi, L) = h^{-1} \cdot (A_U(x)(\xi) + L) \cdot h
\]

Lastly, it is of course of interest to know what happens when we go over to a different trivialisation. Since we have done similar work a couple of times before
in this paper we will just state the result and leave the calculation to the reader. The result is:

\[ A_V = \text{Ad}_{g_{V^U}}A_U - d_{g_{V^U}} \cdot g_{V^U}^{-1} \]  

(2.18)

Having gone through enough mathematical preparations, we can finally have a look at something more physical in nature and study the concept of parallel transport, which will finally illustrate what we need this whole mathematical framework for.

2.2 Parallel transport

Once we have a connection at our disposal, we can relate elements of different fibres. In a sense, a connection is a choice of identification of different fibres – in differential form. Parallel transport is the global version of the same identification.

In this section, we assume that all spaces are smooth manifolds and all maps are smooth. Let \( F \longrightarrow E \longrightarrow X \) be a fibre bundle equipped with a connection. In general, the identification of the fibre above \( x_0 \in X \) with the fibre above \( x_1 \in X \) depends on a choice of path from \( x_0 \) to \( x_1 \).

**Definition 2.2.1.** A smooth path \( a : [0, 1] \rightarrow E \) is called horizontal if at every point \( a(t) (t \in (0, 1)) \) the tangent vector to \( a \) is contained in the horizontal subspace of the tangent space \( T_{a(t)} E \) (as specified by the connection). A lift of a path can move in an arbitrary manner in the vertical direction; demanding that it be horizontal restricts that freedom, by prescribing in which direction the lift must move as we go along the path in the base space \( X \). You might guess that the horizontal lift of a path is essentially unique (it can of course be shifted as a whole in the vertical direction). We will see in a moment that this is true.

Let’s see what a horizontal lift \( \tilde{b} \) of a path \( b : [0, 1] \rightarrow X \) looks like in local coordinates. Let \( h : p^{-1}(U) \rightarrow U \times F \) be a trivialisation over the open set \( U \subseteq X \). We assume for the moment that the image of \( b \) lies completely in \( U \).

Write \( (h \circ \tilde{b})(t) := (x(t), f(t)) \), for \( t \in [0, 1] \). \( \tilde{b} \) being a lift of \( b \) means that \( x(t) = b(t) \). Horizontality of \( \tilde{b} \) at \( t \) means that the vertical projection of \( d\tilde{b}|_t \) is zero, or equivalently (since \( h \) is a diffeomorphism by assumption) that the vertical projection of \( d(h \circ \tilde{b})|_t =: (dx|_t, df|_t) \) is zero. According to (2.2), that vertical projection is \( (0, df|_t + \Gamma(x(t), f(t))dx|_t) = (0, df|_t + \Gamma(b(t), f(t))db|_t) \).

So \( \tilde{b} \) is horizontal if and only if

\[ df|_t = -\Gamma(b(t), f(t)) \circ db|_t \]  

(2.19)

holds for all \( t \in [0, 1] \).

For a given path \( b \), this is an ordinary differential equation for \( f \). As a boundary condition, we choose an element \( e_0 \) of the fibre above \( b(0) \) and demand that \( b(0) = e_0 \). According to the theory of ordinary differential equations there may or may not be a solution \( f \) for all \( t \), so there may or may not be a horizontal
lift \( \tilde{b} \) of \( b \) with \( \tilde{b}(0) = e_0 \), but if there is one, it is unique. In this case, we call \( \tilde{b}(1) \) the parallel transport of \( e_0 \) along the path \( b \).

Remarks.

1. Reparametrisation has no effect on parallel transport: if \( \tilde{b} \) is a lift of \( b \) (so \( p \circ \tilde{b} = b \) and \( \phi : [0, 1] \rightarrow [0, 1] \) is a smooth map with \( \phi(0) = 0 \) and \( \phi(1) = 1 \), then \( \tilde{c} := \tilde{b} \circ \phi \) is a lift of \( c := b \circ \phi \) (since \( p \circ \tilde{c} = \tilde{b} \circ \phi = b \circ \phi = c \)) with \( \tilde{c}(0) = \tilde{b}(0) \) and \( \tilde{c}(1) = \tilde{b}(1) \). Furthermore, if \( \tilde{b} \) is horizontal, then so is \( \tilde{c} \), since \( d\tilde{c} \big|_t = d\tilde{b} \big|_{\phi(t)} \circ d\phi \big|_t \) is a scalar multiple of \( d\tilde{b} \big|_{\phi(t)} \), so the former is in the horizontal tangent space if the latter is.

2. The composition of parallel transport along a path with parallel transport along another path is the same as parallel transport along the concatenation of the paths (traversed at double speed, if desired, to be a path with domain \([0, 1]\)). This follows directly from the remark above and the observation that the concatenation of horizontal lifts is a horizontal lift of the concatenation.

3. The earlier assumption that the image of \( b \) lies within a single trivialising open subset of \( X \) is no real restriction. Given an arbitrary path \( \gamma : [0, 1] \rightarrow X \), we can slice it up in a finite (since \([0, 1]\) is compact) number \( n \) of pieces \( \gamma_j : [t_{j-1}, t_j] \rightarrow X \) \((0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = 1)\), such that the image of each piece \( \gamma_j \) is completely within a trivialising open subset \( U_j \). We then define the parallel transport of a \( e_0 \in p^{-1}(\gamma(0)) \) along \( \gamma \) by consecutively transporting \( e_0 \) along the \( \gamma_j \) (by the definition above). (This does not depend on the choice of slicing. We will not prove that here, it follows from the two remarks above.)

There are two common cases where parallel transport always exists and we can be more explicit about its form.

Example 2.2.2. If the bundle is a vector bundle and the connection is linear, then the differential equation for \( f \) is linear, so it has a (unique) solution (given \( b \) and \( e_0 \)). Let us write \( A(t) := \Gamma(b(t), f) \circ db \big|_t \), so \( A(t) \) is a linear map from \( F \) to \( T_{f(t)}F \); (2.19) then reads

\[
df \big|_t = -A(t)f(t),
\]

while \( f(0) \) is determined by \( (b(0), f(0)) = \tilde{b}(0) = e_0 \).

Now, if \( A(t) \) were independent of \( t \), the solution would be simply \( f(t) = \exp \left( \int_0^t -A(s)ds \right) f(0) \). In the general case the exponential and integral must be replaced by what is known as the ordered exponential (or path-ordered exponential, or time-ordered exponential), commonly written as \( P\exp \) or \( T\exp \). In these terms, the solution to (2.20) is (we do not prove this)

\[
f(t) = P\exp \left( \int_0^t -A(s)ds \right) f(0).
\]

This notation is easy on the eye and signifies the similarity to the ordinary exponential well, but it unfortunately does not show that the \( P\), \( \exp \) and \( \int_0^t \) terms.
together form one mathematical operator. The ordered exponential can be defined by the differential equation above, or as a limit:

\[ P \exp \left( \int_0^t B(s) \, ds \right) := \lim_{n \to \infty} \prod_{k=n}^1 \exp(\Delta_n B(s_k)) \]  

(2.22)

where \( s_k = \frac{kt}{n} \), \( \Delta_n = \frac{t}{n} \) and the exp in the right-hand side is the exponential of linear operators from \( F \) to itself (which can be defined by its Taylor series: \( \exp(Q) = \sum_{i=0}^{\infty} \frac{Q^i}{i!} \)). This looks somewhat like the definition of the Riemann integral of analysis: we partition the interval \([0, t]\) in small pieces, represent \( \exp(B) \) on a small piece with the exponential of the length of the piece times \( B \) in some point of the piece, multiply those factors over the interval and finally let the partition become infinitely fine. Apart from the fact that we multiply the factors \( \exp(B) \), the major difference with the Riemann integral is that the factors of the product in the right-hand side of (2.22) may not commute: the order is important! The path-ordering or time-ordering now lies in the fact that we take the product in order of decreasing \( t \) from left to right.

With some suggestive notation (identifying \( A \) with \( \Gamma \)) we can write the parallel transport of \( e_0 = h^{-1}(b(0), f(0)) \) along the curve \( b \) as

\[ P \exp \left( -\int_b \Gamma \right) f(0). \]  

(2.23)

**Example 2.2.3.** If the bundle is a principal \( G \)-bundle and the connection is invariant, we can proceed in a way analogous to the solution above for linear connections on vector bundles. According to (2.17), in terms of the local principal gauge potential \( A_U \), horizontality is equivalent to

\[ f(t)^{-1} \cdot (A_U(db|\Gamma(1)) + df|\Gamma(1) \cdot f(t)^{-1}) \cdot f(t) = 0 \]  

(2.24)

or equivalently

\[ A_U(db|\Gamma(1)) + df|\Gamma(1) \cdot f(t)^{-1} = 0 \]  

(2.25)

so

\[ df|\Gamma(1) = -A_U(db|\Gamma(1)) \cdot f(t) . \]  

(2.26)

This differential equation is similar to the linear case (2.20); here the solution always exists and is given by (no proof):

\[ f(t) = P \exp \left( -\int_0^t A_U(b(s), db|s(1)) \, ds \right) f(0). \]  

(2.27)

where we define the ordered exponential by the same equation (2.22) as above. The only difference is that while before, the integrand \( B(s) \) was a linear operator on \( F \) and \( \exp \) the operator exponential, now \( B(s) \) is an element of the Lie algebra \( g \) and \( \exp \) is the Lie exponential (producing an element of the Lie group \( G \)). As above, we write the parallel transport of \( e_0 = h^{-1}(b(0), f(0)) \) along the path \( b \) in compact form as

\[ P \exp \left( -\int_b A_U \right) f(0). \]  

(2.28)

\[ ^2 \text{This is logical on an intuitive level: since we are trying to define something like the exponential of an integral (a continuous sum), it should be something like the (continuous) product of exponentials.} \]
2.3 Curvature

Before we introduce the concept of curvature, we illustrate it with a well known example.

Example 2.3.1. For this example, we must introduce some notions from Riemannian geometry. With a metric on a smooth manifold $X$, we here mean a smoothly varying choice of a nondegenerate inner product on the tangent spaces of $X$. A manifold equipped with such a metric is called a pseudo-Riemannian manifold. Pseudo, because the metric might not be positive-definite; if it is, the manifold together with this metric is called Riemannian. The metric makes it possible to talk about length and angle of tangent vectors, and by integration also about length of a path in the manifold.

One can define a linear connection on the tangent bundle of a pseudo-Riemannian manifold, that interacts nicely with the metric. It is called the Levi-Civita connection. Along a geodesic path, parallel transport preserves the tangent vectors to the path. This is usually taken as the definition of a geodesic path, but in the Riemannian case, geodesics are precisely the paths with (locally) minimal length (straight lines in Euclidian space $\mathbb{R}^n$, great circles on the sphere $S^2$).

Now, for a concrete example of parallel transport, let us consider the tangent bundle of the sphere $S^2$. There is a logical metric on the sphere, induced by the embedding of the sphere in $\mathbb{R}^3$, so we have the Levi-Civita connection at our disposal. It turns out that parallel transport with respect to this connection is just what you would hope it to be: along a geodesic path, a tangent vector $v$ is transported by preserving the angle between $v$ and the tangent vector to the path and keeping $v$ at constant length.

We will take the following path in $S^2$: start at the north pole; go down to the equator along a meridian; move east along the equator for a quarter of the circumference; return to the north pole by a meridian. This is a (sphere-)triangular piecewise-smooth path that encloses one eighth of the surface of the sphere. As our victim of parallel transport, we pick a tangent vector at the north pole that points in the direction of the first piece of our path. Transporting it along that first piece, it remains tangent to the path, so when we arrive at the equator it points due south. While moving east and keeping it perpendicular to the equator, it keeps pointing south. As we go up again, our tangent vector is constantly directed back along the meridian we are now following. Back at the north pole, the transported version of our vector is tangent to the meridian on which the last piece of our path lies: it has been rotated counterclockwise over 90 degrees with respect to the original tangent vector!

This example shows that parallel transport on the sphere indeed does depend on the path taken: if we had taken the constant path from the north pole to itself, our tangent vector would not have changed at all. If you perform the same experiment on the plane ($\mathbb{R}^2$) or a the surface of a cylinder ($S^1 \times \mathbb{R}$), you will see that parallel transport does not depend on the choice of path. The crucial difference is that the sphere is curved, while the plane and the surface...
of a cylinder are not. (The latter may seem to be curved, but not in the sense we mean. This is commonly illustrated by the fact that one can fold a (flat) piece of paper perfectly tight around a cylinder, while one can not do so around a sphere.)

**Definition 2.3.2.** We would like to define a general notion of curvature for a smooth fibre bundle $F \rightarrow E \xrightarrow{p} X$. In general, the degree of curvature might vary from point to point, so let’s fix a point $e \in E$. We must also choose two vectors $\eta, \xi \in T_p(e)X$. The curvature $R(e)(\eta, \xi)$ is by definition the tangent vector given by the parallel transport of $e$ around the infinitesimal parallelogram in $X$ based at $x = p(e)$ with subsequent corners $x, x + \xi, x + \eta, x + \eta, x$.

The rectangle is infinitesimal, so the result of parallel transport is not an element of the fibre, but an element of the vertical tangent space (vertical, since parallel transport along a closed path does not change the basepoint). We will see in a moment that it depends linearly on both $\eta$ and $\xi$. It is obviously antisymmetric in $\eta$ and $\xi$ (since switching them gives the same rectangle traversed in the opposite direction), so $R(e)$ is something like a $V_e E$-valued 2-form. $R$ is therefore sometimes called the curvature 2-form.

We say the curvature vanishes at $e \in E$ if $R(e)$ is the zero 2-form, and that it vanishes at a point $x \in X$ if it vanishes at the whole fibre above $x$. Finally, we call the connection flat if the curvature vanishes everywhere.

To actually calculate the curvature in terms of the connection, we choose local coordinates $f^a$ ($a \in \{1, \ldots, \dim F\}$) of $F$ and $x^i$ ($i \in \{1, \ldots, \dim X\}$) of $X$ around $(x, f) = h(e)$ ($h$ a local trivialisation). We can obtain the parallel transport along a side of the rectangle up to second order in the length by using Picard iteration on the differential equation of parallel transport ((2.19)). Next we must compose those four operators, throwing out all higher order terms that come up. We will not do this calculation here. The answer is relatively simple: in coordinates and index notation, parallel transport is given by

$$(x^i, f^a) \mapsto (x^i, f^a + R(x, f)^a_{\,ij} \eta^i \xi^j) \quad (2.29)$$

where

$$R(x, f)^a_{\,ij} = \frac{\partial \Gamma^a_{\,i}}{\partial x^j}(x, f) - \frac{\partial \Gamma^a_{\,j}}{\partial x^i}(x, f) + \frac{\partial \Gamma^a_{\,i}}{\partial f^b}(x, f) \Gamma^b_{\,j}(x, f) - \frac{\partial \Gamma^a_{\,j}}{\partial f^b}(x, f) \Gamma^b_{\,i}(x, f). \quad (2.30)$$

**Example 2.3.3.** We will simplify this in the all-important case of a linear connection on a vector bundle. We use the same notation as above. Because of the linearity of the connection, we can write

$$\Gamma^a_{\,i}(x, f) = \Lambda^a_{\,ib}(x) f^b. \quad (2.31)$$

Substitute this in (2.30), noting that $\frac{\partial \Lambda^a_{\,ib}}{\partial x^j}(x, f) = \Lambda^a_{\,ib}(x)$:

$$R(x, f)^a_{\,ij} = \frac{\partial \Lambda^a_{\,ib}}{\partial x^j}(x) f^b - \frac{\partial \Lambda^a_{\,ib}}{\partial x^j}(x) f^b + \Lambda^a_{\,ib}(x) \Lambda^b_{\,jc}(x) f^c - \Lambda^a_{\,ib}(x) \Lambda^b_{\,jc}(x) f^c. \quad (2.32)$$

More formally, one should construct an actual parallelogram with (probably slight curved) sides of variable length in the forementioned directions, and then take the derivative at length 0 of the parallel transport along the parallelogram.
We see that this is linear in \( f \), so after identifying the vertical tangent space with \( F \) in the natural way, we can interpret \( R \) as a 2-form that takes values in the space of operators from \( F \) to itself (an \( \text{End}(F) \)-valued 2-form). We write \( R(x, f)_{ij}^a = R^a_{bij} f^b \) (dropping the \( x \)-dependence for esthetic reasons and abusing the same letter \( R \)), with

\[
R^a_{bij} = \frac{\partial \Lambda^a_{bj}}{\partial x^i} - \frac{\partial \Lambda^a_{bi}}{\partial x^j} + \Lambda^a_{ci} \Lambda^c_{bj} - \Lambda^a_{cj} \Lambda^c_{bi} .
\]

It is now clear that the first two terms are together the \( ij \)-component of \( d\Lambda^a_{bj} \), while the latter two terms are the \( ij \)-component of the 2-form that is usually rather tersely written as \([\Lambda, \Lambda]^a_{bj} \), defined by \([\Lambda, \Lambda](\eta, \xi) = [\Lambda(\eta), \Lambda(\xi)] \) (the commutator is that of linear operators on \( F \)). In this terse notation, we have the following simple formula for the curvature (identifying \( \Lambda \) with \( \Gamma \)):

\[
R = d\Gamma + [\Gamma, \Gamma] .
\] (2.32)

**Example 2.3.4.** In the case of an invariant connection on a principal \( G \)-bundle, the curvature is given by

\[
R = dA + [A, A] \tag{2.33}
\]

where \( A \) is the local principal gauge potential and the brackets denotes the Lie bracket of \( g \)-valued differential forms (\( g \) the Lie algebra of \( G \)).\(^5\) This can be proved with an argument analogous to the one above.

These two special cases can also be obtained from the respective ordered exponential solutions of the parallel transport equation, in combination with the Baker-Campbell-Hausdorff formula (the non-commutative analogue to the formula \( \exp(a) \exp(b) = \exp(a + b) \)).

### 2.4 Holonomy and monodromy

The path dependence of parallel transport is a symptom of a phenomenon called anholonomy. We can measure anholonomy by describing what changes parallel

\(^5\)Just as above, the \( g \)-valued 2-form \([C, D] \) is defined by \([C, D](\eta, \xi) = [C(\eta), D(\xi)] \), where now the right-hand side features the Lie bracket of the Lie algebra.

In the special (but common) case that the Lie algebra \( g \) is a matrix Lie algebra (that is, the elements of \( g \) are matrices and the Lie bracket \([x, y] \) is the commutator \( xy - yx \) of matrices), we can see a \( g \)-valued differential form as a matrix of real-valued forms. We then define the **exterior product** \( C \wedge D \) of two \( g \)-valued forms \( C \) and \( D \) as the matrix product of \( C \) and \( D \), multiplying **elements** of the matrices according to the exterior product of ordinary (real-valued) forms. In index notation, we write this as

\[
(C \wedge D)^e_c = C^a_b \wedge D^b_c .
\] (2.34)

This exterior product is associative, just like the real-valued case, but it is not in general antisymmetric on 1-forms, due to possible non-commutativity of the Lie-algebra.

For the special case where \( C \) and \( D \) are 1-forms, we have the formula \((C \wedge D)(\eta, \xi) = C^a(\eta)D^b(\xi) - C^b(\xi)D^a(\eta)\), so when \( C = D = A \), we have

\[
(A \wedge A)^e_c(\eta, \xi) = A^a_\xi(\eta)A^b_\xi(\xi) - A^a_\xi(\xi)A^b_\xi(\eta) = (A(\eta)A(\xi))^e_c - (A(\xi)A(\eta))^e_c = [A(\eta), A(\xi)]^e_c \tag{2.35}
\]

so dropping the indices this says that \( A \wedge A = [A, A] \). Formula (2.33) is thus also written as \( R = dA + A \wedge A \).
transport along a closed path can cause, in the form of the holonomy group (which might be more properly called the anholonomy group).

**Definition 2.4.1.** Let \( F \xrightarrow{p} E \xrightarrow{p^{-1}} X \) be a fibre bundle with a connection. The **holonomy group** of the bundle (and connection) at basepoint \( x \in X \) is the subgroup of \( \text{Aut}(p^{-1}(x)) \), the automorphism group of the fibre above \( x \), obtained by parallel transporting (according to the connection) along closed paths based at \( x \). The **local holonomy group** is the subgroup of the holonomy group obtained by parallel transporting only along contractible paths based at \( x \).

So, given a closed path \( \gamma \) based at \( x \), the map that sends \( e \in p^{-1}(x) \) to the parallel transport of \( e \) along \( \gamma \) is an element of the holonomy group at basepoint \( x \).

**Remarks.**

1. The (local) holonomy group is indeed a group, because of remarks 1 and 2 of section 2.2, and since the closed paths based at \( x \) modulo reparametrisation form a group.

2. If the base space \( X \) is path-connected, switching to another base point \( y \in X \) changes the holonomy group only by isomorphism (namely conjugation by parallel transport along a path from \( x \) to \( y \)), so the basepoint is then usually not mentioned. This situation is analogous to that regarding the fundamental group of a topological space. Especially when the basepoint is omitted, the fibre above the basepoint is identified with the abstract fibre \( F \), so the holonomy group can be seen as a subgroup of \( \text{Aut}(F) \). (Choosing an identification induced by another trivialisation only changes the holonomy group by conjugation with the transition map evaluated at the basepoint.)

3. The (local) holonomy group of a pseudo-Riemannian manifold is defined to be the (local) holonomy group of the tangent bundle of the manifold with respect to the Levi-Civita connection. Since the Levi-Civita connection is linear, parallel transport is always a linear transformation, so the holonomy group of a \( n \)-dimensional pseudo-Riemannian manifold is a subgroup of \( GL(n) \) (identifying the tangent space with \( \mathbb{R}^n \)). In fact, because of the special features of the Levi-Civita connection (loosely speaking, it preserves the metric), parallel transport is actually an orthogonal transformation (i.e. has determinant \( \pm1 \)), so the holonomy group is a subgroup of \( O(n) \). If the manifold is orientable, a parallel transport must have determinant 1, so the holonomy group is even a subgroup of \( SO(n) \), the special orthogonal transformations.

**Example 2.4.2** (Holonomy of \( S^2 \)). Let us return to example 2.3.1. What is the holonomy group of \( S^2 \)? We saw in remark 3 above that it is a subgroup of \( SO(2) \) (the rotations of the plane). But with a simple modification of the path of 2.3.1, we can obtain any rotation of the tangent space: start at the north pole; go down to the equator along a meridian; move east along the equator over a length (or angle) \( \alpha \); then return to the north pole along a meridian; parallel transport along this path is obviously rotation over \( \alpha \) in the counterclockwise direction.
direction. This shows that the holonomy group of $S^2$ is $SO(2)$. Because $S^2$ is simply connected, this is also the local holonomy group.

**Remark 2.4.3.** In general, oriented $n$-dimensional pseudo-Riemannian manifolds have as holonomy group the full $SO(n)$, just like the 2-sphere. If the holonomy group is smaller, this signifies some special property of the manifold. For example:

- The $2n$-dimensional oriented Riemannian manifolds with holonomy $U(n) \subseteq SO(2n)$ are precisely the Kähler manifolds (manifolds that admit a certain complex structure).
- The $2n$-dimensional oriented Riemannian manifolds with holonomy $SU(n) \subseteq SO(2n)$ are precisely the Calabi-Yau manifolds (Kähler manifolds with a certain additional topological property – namely that the first Chern class (a so-called characteristic class) is zero).

A serious investigation of these correspondences is beyond the scope of this paper.

**Example 2.4.4 (Parallel transport on a covering).** On a covering (fibre bundle with a discrete fibre) with its unique connection, parallel transport around a closed loop coincides with what is known in topology as the monodromy action. Any (continuous) lift of a path in the base space is automatically horizontal. Because the total space locally looks like the base space (or because the vertical tangent space is zero), parallel transport around an infinitesimally small closed path is the identity. In particular, the curvature vanishes.

However, if the the fibre has more than one point and the total space is path-connected, there are loops in the base space around which parallel transport is not the identity. Consider for example the covering $\exp : \mathbb{R} \to S^1$ of the circle by the line. Parallel transport around a path that winds one time around the circle moves an element of a fibre by $2\pi$.

The above example shows that even if the curvature is zero, so locally parallel transport is path-independent, there still can be anholonomy caused by global path dependence. This phenomenon is called monodromy.

**Equivalence of flatness with trivial local holonomy** If a fibre bundle with connection has trivial local holonomy group, then of course the connection is flat (a manifold is locally simply connected, so the curvature can be obtained as a limit of parallel transport around contractible loops, and by assumption those have no effect). It is an important fact that the reverse implication holds as well. We will sketch a proof. Let $\gamma : [0,1] \to X$ be a loop based at $x$ and let $F : [0,1] \times [0,1] \to X$ be a homotopy from the constant loop at $x$ to $\gamma$ (we may take it to be a smooth homotopy). Consider the parallel transport around the path $F(\cdot, t)$. For $t = 0$, this does nothing, since $F(\cdot, 0)$ is the constant loop. As we move from $F(\cdot, t)$ to $F(\cdot, t + dt)$, the difference between these two paths can be written as the concatenation of many infinitesimal rectangles with corners $F(t, s)$, $F(t + dt, s)$, $F(t + dt, s + ds)$ and $F(t, s + ds)$ (this is appreciated best using a drawing). Parallel transport around each of these rectangles is the identity (because the curvature vanishes by assumption), so the difference between parallel transport around $F(\cdot, t)$ and $F(\cdot, t + dt)$ is also the identity (it
is just the continuous composition of the parallel transport around each of the infinitesimal rectangles); by continuous composition of all those differences, we see that parallel transport around $\gamma(\cdot) = F(\cdot, 1)$ is the identity as well.

This result can be interpreted as a justification of the definition of curvature: curvature can detect all local anholonomy.

### 2.5 Covariant derivative

An important reason to consider connections is that they facilitate the differentiation of sections. Let $\xrightarrow{\theta} E \xrightarrow{F} X$ be a fibre bundle, as always. Given a (local) section $s : U \to E$ on an open subset $U \subseteq X$, a point $x \in U$ and a vector $\xi \in T_xX$, we can take the derivative $ds(\xi) \in T_{s(x)}E$. This is perfectly legitimate, but we end up with a tangent vector to the total space $E$. That is usually not what you want. We already know how the horizontal component of $s$ changes – but we cannot even talk about the vertical component of a section without a connection.

When we do have the disposal of a connection, we can take the vertical projection of the derivative: $\pi^v ds(\xi) \in V_{s(x)}E$. This is called the **covariant derivative** of $s$ at $x$ in direction $\xi$ and denoted by $D_\xi s$. More generally, if $\mathcal{X}$ is a vector field on $U$, we can take the covariant derivative at every point of $U$ and obtain a function $D_\mathcal{X}s : U \to VE : x \mapsto D_{\mathcal{X}(x)}s$, a section over $U$ of the vertical tangent bundle.

If furthermore the bundle is a vector bundle, we can identify the vertical tangent space $V_{s(x)}E$ with the fibre $p^{-1}(x)$ in the usual way. The covariant derivative $D_\xi s$ is then simply an element of the fibre over $x$, and the covariant derivative $D_\mathcal{X}s$ of a section $s$ is again a section of the vector bundle.

In local coordinates, we can use formula (2.2) for the vertical projection to obtain an explicit formula for the covariant derivative. Using the same notation as above, let $h : p^{-1}(U) \to U \times F$ be a trivialisation, let $s_U : U \to F$ be the local representative of the section $s$ (given by $(h \circ s)(x) = (x, s_U(x))$) and let $\Gamma_U$ be the local representative of the connection. Substituting (2.2) in the definition of covariant derivative above, we see that the local representative of $D_\xi s$ is given by

$$ds_U(\xi) + \Gamma_U(x, s_U(x))\xi.$$  

The first term is just $\mathcal{X}(s_U)(x)$; we can write the second term as $\Gamma_U(\mathcal{X})(x, s_U(x))$ if we let $\Gamma_U$ act on the vector field $\mathcal{X}$ in the logical way (emphasizing the 1-form character of $\Gamma_U$). So in terse point-free notation, we can say that the local form of $D_\xi s$ is

$$\mathcal{X}(s_U) + \Gamma_U(\mathcal{X})(s_U).$$  

**Remark 2.5.1.** The complementary description, using differential forms instead of vector fields, is closer to the usual physical notation. We consider the case of a vector bundle with a linear connection. Let $t : X \to E$ be a section (written in index notation as $t^a$, the components of the vertical part of $t$ (according to some implicit trivialisation)). Then $\pi^v \circ dt : TX \to VE$ is called the covariant derivative of $t$, denoted by $D_t$ – this is the same definition as above, except that here we have not yet chosen in which direction we will differentiate: $D_\xi t = Dt(\xi)$. Since the bundle is a vector bundle, we can identify $VE$ with $E$, so
obtaining a $D t : TX \to E$, or – one more identification – a section of the tensor product bundle $T^*X \otimes E$.

This last object is written in index notation as $D_\mu t^a$; from (2.36), we see that

$$D_\mu = \partial_\mu + A_\mu$$  \hspace{1cm} (2.38)

, where $A$ is the connection 1-form, an $\text{End}(F)$-valued 1-form, given by $A(\xi)(t) = \Gamma_U(x,t)\xi$. Intuitively, the connection form of a linear connection (with respect to some trivialisation) gives the difference between the connection and the trivial connection (induced by the trivialisation).

---

6This ‘connection form’ of a linear connection is similar, but not directly related, to the ‘connection form’ of an invariant connection (from section 2.1.2).
Chapter 3
Applications

To give a taste of how the theory we introduced can be put to work, we will now review some applications.

3.1 Electromagnetism as a gauge theory

Classically, electromagnetism is described by two vector fields: the electric field and the magnetic field. From a special-relativistic point of view, those are really two manifestations of the same object, known by physicists as the field strength tensor, typically written in index notation as $F_{\mu\nu}$ (see [1, Chapter 12]). It is an antisymmetric $(0, 2)$-tensor: a 2-form on the 4-dimensional spacetime manifold (let’s call that $X$). We will denote the field strength by $F$.

Maxwell’s equations of electrodynamics take a particularly simple form in terms of differential forms. The two homogeneous equations translate to the condition that $F$ is a closed 2-form:

$$dF = 0$$  \hspace{1cm} (3.1)

(where $d$ is the ordinary exterior derivative of differential forms). The two inhomogeneous equations can be written as\(^1\)

$$\ast d(\ast F) = J,$$  \hspace{1cm} (3.2)

where $\ast$ is the so-called Hodge star (given a $p$-form on an $n$-dimensional pseudo-Riemannian manifold, it produces an $(n-p)$-form, making explicit use of the metric; physicists might know it under the name of “contracting with the Levi-Civita tensor”) and $J$ is a 1-form called the 4-current (the source of the electromagnetic field: the temporal component is (minus) the electric charge distribution $\rho$ and the spacial component is the ordinary current $J$). We will not go through this in detail; this is done for example in [2, Chapter 7, sections 4.1-4]

Just as in the field formulation, we can exploit the homogeneous equation $dF = 0$ to write the field strength $F$ in terms of a potential: $F = dA$ (so $A$ is a 1-form)\(^2\). (In index notation, this is written as $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$, and

\(^1\)We use natural units, so that $4\pi\epsilon_0 = 1$.

\(^2\)Note that for this to work, spacetime must be contractible (or at least its second de Rham cohomology group must vanish). To simplify a bit, we will assume a contractible spacetime in this section, so in particular all closed forms are exact.
\(A_\mu\) has (minus) the electric scalar potential \(V\) as temporal component and the magnetic vector potential \(A\) as spatial component. This potential formulation automatically takes care of the homogeneous Maxwell equation \((dF = ddA = 0)\); the only remaining equation is

\[ *d(*dA) = J. \tag{3.3} \]

The description of the electromagnetic fields in terms of the potential \(A\) has an obvious redundancy: if we add a closed 1-form \(\alpha (= d\lambda)\) to it, \(A \mapsto A + \alpha\), the field strength \(F\) does not change. (In the tensor formulation, this is written as \(A_\mu \mapsto A_\mu + \frac{\partial \lambda}{\partial x^\mu}\) with \(\lambda\) a scalar function on \(X\).) This local symmetry (local, because we can change \(A\) independently at different points of spacetime) is called a gauge symmetry; changing to a different but equivalent \(A\) (by adding a closed form) is called a gauge transformation.

Now comes the starting point of gauge theory: using this potential \(A\), we can interpret the electromagnetic fields as a connection: specifically, an invariant connection on the trivial principal \(U(1)\)-bundle over spacetime \(X\) (\(U(1) \cong S^1\) is the group of unitary complex numbers). To do so, we let \(A\) be the local principal gauge potential of our connection. This makes sense, since \(A\) is a real-valued 1-form, and the Lie algebra of \(U(1)\) is isomorphic to \(\mathbb{R}\). Furthermore, transformations of \(A\) due to a change of trivialisation correspond to gauge transformations of \(A\) in the sense defined above: let \(h_1\) and \(h_2\) be two different global trivialisations, and let \(g : X \to U(1)\) be the difference \(((h_2 \circ h_1^{-1})(x, f) = (x, g(x)f))\). According to formula (2.18), the transformation of \(A\) is given by

\[ A_2 = g \cdot A_1 \cdot g^{-1} - dg \cdot g^{-1}. \tag{3.4} \]

\(U(1)\) is abelian, so the first term on the right-hand side is just \(A_1\). The multiplication with \(g^{-1}\) in the second term is just the identification of the tangent space at \(g(x)\) with the Lie algebra, so the second term is a \(u(1)\)-valued one-form, say \(\alpha\):

\[ A_2 = A_1 - \alpha; \tag{3.5} \]

this is exactly what we called a gauge transformation! Reversely, starting with such a gauge transformation, we can exponentiate it to get a change of trivialisation. This is a nice development. We have incorporated the redundancy of the potential in a natural way: different gauges (equivalent \(A\)s) turn out to be different “chartings” of the same underlying object (the connection).

The curvature of the connection introduced above is given by (see (2.33))

\[ R = dA + [A, A] = dA \tag{3.6} \]

(the second term vanishes, because the Lie algebra of \(U(1)\) is abelian), so the curvature of the connection is just the field strength \(F\)!

We state without proof that the sourceless \((J = 0)\) Maxwell equations are the Euler-Lagrange equations emanating from the Lagrangian \(F \wedge *F\) (in index notation) (note that if \(F\) is a 2-form, then \(*F\) is a \((n-2)\)-form, so \(F \wedge *F\) is an \(n\)-form that can be integrated over \(n\)-dimensional spacetime). This is an instance of the general Yang-Mills Lagrangian

\[ -k tr(F \wedge *F) \tag{3.7} \]
, where $k$ is some constant and $\text{tr}$ denotes the trace (of matrices) on the matrix Lie algebra. The study of this Yang-Mills action is an important part of gauge theory.

Maxwell’s equations (or, if you prefer, the Yang-Mills Lagrangian) only describe the evolution of the electromagnetic field, not how it influences particles. To do that, you need to add some sort of interaction term to the Yang-Mills Lagrangian. There is a nice procedure to derive such a term. First you must choose an irreducible unitary representation of the gauge group, $U(1)$ in our case – this will turn out be the charge of the particle$^3$. Use this representation to construct the associated bundle from the principal bundle. The wave function of the particle is assumed to be a section of this associated bundle. Next, the connection representing the gauge field can be transferred in a natural way to a connection on the associated bundle. Finally, we obtain an interaction term by taking the usual ‘free’ Lagrangian of the particle and replacing all occurrences of the derivative by the covariant derivative, with respect to the connection we just obtained. This procedure is called minimal coupling.

3.1.1 The abelian Aharonov-Bohm effect

In quantum electrodynamics, there is the following phenomenon, first noted by Ehrenberg and Siday in 1949, and later by Aharonov and Bohm ([3]), after whom it is called. Suppose that we have an infinitely long solenoid. Inside, there is a nonzero magnetic field. Outside, the magnetic field is zero (we assume for simplicity that the configuration is static and that there are no electric fields). We send a beam of electrically charged particles – electrons, say – towards the solenoid, split it into two beams, let the beams go past the solenoid on different sides, and then rejoin the beams. If you think about this experiment classically, you wouldn’t expect anything special to happen: the electrons stay away from the solenoid, so they never encounter any magnetic field. Quantum mechanically, however, it turns out that there is an electromagnetic effect: electrons going past the solenoid on different sides acquire a relative phase factor, thereby creating interference. Let’s see if we can find out what happens. To simplify things, we constrain the problem to 2 dimensions (the electrons stay in the $z = 0$-plane).

In electrodynamics the Hamiltonian is given by$^4$

$$H = \frac{(P - qA)^2}{2m}. \quad (3.8)$$

Note that, to avoid confusion with the base of the natural logarithm, we have denoted the the electron charge by $-q$. Now to better exploit the symmetry of our situation it is better to express this in cylindrical coordinates. Substituting $P \to -i\nabla$ and realising that in cylindrical coordinates the nabla operator is

$^3$Note that the unitary irreducible representations of $U(1)$ are (complex) one-dimensional and are parameterised by an integer $q$, sending $e^{i\theta}$ to multiplication by $e^{iq\theta}$; this nicely incorporates the quantised nature of electrical charge.

$^4$In other words, it turns out that using this Hamiltonian gives the correct predictions with respect to experiments and reduces to classical electrodynamics in an appropriate limit. One way to obtain this Hamiltonian is by minimal coupling: take the Hamiltonian $H = \frac{P^2}{2m}$ of a free neutral particle and replace the partial derivative $\nabla$ (in $P$) with the covariant derivative $\nabla - iqA$. 

33
given by $\nabla = \frac{\partial}{\partial r} \hat{r} + \frac{1}{r} \frac{\partial}{\partial \theta} \hat{\theta}$ we can rewrite the Hamiltonian with respect to these coordinates to give

$$H = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left( \frac{\partial}{\partial \theta} + i\alpha \right)^2 .$$

(3.9)

Here we used a gauge\(^5\) in which $A_r = 0$ and $A_\theta = \frac{\phi}{2\pi r}$, where $\phi$ is the magnetic flux through the solenoid, and where we have defined $\alpha := -\frac{q\phi}{\hbar}$. Putting this in the time-independent Schrödinger equation $H \psi = E \psi$ and setting\(^6\) $E =: -k^2$ we get the following wave equation:

$$\left( \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left( \frac{\partial}{\partial \theta} + i\alpha \right)^2 + k^2 \right) \psi = 0 .$$

(3.10)

The solution to this equation is given by:

$$\psi(r,\theta) = \sum_{m=-\infty}^{\infty} e^{im\theta} \left[ a_m J_{m+\alpha}(kr) + b_m J_{-(m+\alpha)}(kr) \right] ,$$

(3.11)

where $a_m$ and $b_m$ are arbitrary constants and $J_{\pm(m+\alpha)}$ is a Bessel function of order $\pm(m+\alpha)$ (indeed, a common approach is to define these Bessel functions as the solutions of this very differential equation). Note that this solution holds only outside the solenoid.

From now on, we will work in the limit where the radius of the solenoid goes to zero while the total flux $\phi$ remains fixed. We do so because it will simplify the calculations a lot, and while this means that we will be working in an idealized situation that does not describe a real physical system, it still

\(^5\)It is easy to see that this gauge indeed describes our situation outside the solenoid: the curl in cylindrical coordinates of $A$ is $\nabla \times A = (\frac{1}{r} \frac{\partial A_r}{\partial \theta} - \frac{\partial A_\theta}{\partial r}) \hat{r} + (\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}) \hat{\theta} + \frac{1}{r} \frac{\partial}{\partial \theta} (r A_\theta) - \frac{\partial A_z}{\partial \theta} \hat{z} = 0 = B_{\text{outside}}$. Furthermore, the path integral of $A$ along a path that encloses the solenoid once is $\phi$, the flux through the enclosed surface.

\(^6\)Note that $k$ will also be the wave number of the incoming electron beam as given in (3.13). This makes perfect sense, since far to the right of the solenoid, where the magnetic influence is negligible, a solution with energy $-k^2$ should look like a plain wave with wavenumber $k$. 

34
demonstrates an interesting and important quantum phenomenon, which has no classical counterpart.

The concrete advantage of working in this limit has to do with the Bessel functions of negative order. Of course, the wave function must be defined on all of space, and it must also be continuous. This means that we have to extend this solution (3.11) to region inside the solenoid. To do this, in the limit where the radius of the solenoid goes to zero, we only have to specify a value at the origin. The Bessel functions of negative order, however, have a pole at the origin, so there is no continuous way to extend them there. Therefore they cannot be part of the solution, in this limit, and we have

$$\psi(r, \theta) = \sum_{m=-\infty}^{\infty} a_m J_{|m+\alpha|} (kr) e^{im\theta}. \quad (3.12)$$

Remember that the situation we want to study is as follows. A beam of electrons comes in from the right, splits up, passes the solenoid on both sides and is rejoined on the other side. This means that we have the initial condition that for large $r$ to the right of the solenoid (where the effect of the solenoid is not yet felt), $\psi$ must represent such an incoming wave; in that limit, we must have

$$\psi = \psi_{\text{inc}} = e^{-i(\alpha \theta + kx)}. \quad (3.13)$$

We mention two ways to see that this $\psi$ indeed represents the desired incoming wave. The first is an intuitive argument. Let’s say we are working in a different gauge, where we have gauged away the potential on the right side of the solenoid. This is possible since the curl of $A$ is equal to zero here, as we have shown above (in footnote 5 on page 34). So locally, we can write $A$ as the gradient of some scalar function $f$ and gauge it away. In fact, we can easily write down this scalar function, by solving the equation $\nabla f = A$, which is a system of partial differential equations with solution (as the reader will be able to check easily): $f = \frac{\phi}{2\pi} \theta = -\frac{\alpha}{2} \theta$. Since the potential in the region of the incoming wave is then zero, we know what the incoming wave should look like; it should be the normal free particle wavefunction $\psi_{\text{inc}} = e^{-ikx}$. Now return to the original gauge $A_\theta = \frac{\phi}{2\pi}$. When performing this gauge transformation, the wave function is multiplied by $e^{iqf}$, so after gauging back to our original situation, the incoming wave function becomes $\psi_{\text{inc}} = e^{-i\alpha \theta} e^{-ikx}$, as we claimed in (3.13).

The second reason to want (3.13) as the incoming wave appears once we rephrase the initial condition as the demand that far to the right of the solenoid, the wave function should have a constant current density in the $-x$ direction. If we take the gauge covariant form of the current density (with covariant derivatives instead of ordinary ones), then this implies that $\psi_{\text{inc}}$ must have the given form (see also [3]).

---

7 This transformation behaviour of $\psi$ is just that of a section of the (electromagnetic) bundle associated to a particle with charge $-q$. A compelling reason to demand this behaviour is that it is necessary to make the Schrödinger equation gauge invariant: if we changed only the potential, the Hamiltonian alone would change; we need to change the wave function simultaneously with the potential. Of course, a gauge transformation cannot change the physics of a situation, so the only thing we can do to the wave function is tag on a phase factor. For a gauge transformation $A \rightarrow A + \nabla f$ the right phase factor is $e^{iqf}$, as the reader can check by direct calculation.
It turns out that the correct choice for $a_m$ for this initial condition is given by $a_m = (-i)^{m+\alpha}$, thus giving

$$\psi(r, \theta) = \sum_{m=-\infty}^{\infty} (-i)^{m+\alpha} J_{m+\alpha}(kr)e^{im\theta}. \quad (3.14)$$

In [3] it is shown that this indeed gives the right $\psi_{inc}$ by calculating that, in the limit of large $r$, the wave function is given by

$$\psi \rightarrow e^{-i(\alpha\theta+kx)} + \frac{e^{ikr}}{(2\pi ikr)^{\frac{1}{2}}} \frac{\sin(\pi\alpha)}{\cos(\theta/2)} e^{-i\theta/2}. \quad (3.15)$$

We recognise the first term to be the incoming beam because it does not decrease with growing $r$, as a reflected term in a multi dimensional problem should (to keep the total probability to find the electron somewhere equal to one).

The scattering cross section can now be calculated by writing

$$\psi = \psi_{inc} + e^{ikx}g(\theta)$$

and using $^8 \frac{d\sigma}{d\Omega} = |g(\theta)|^2$. This results in:

$$\frac{d\sigma}{d\Omega} = \frac{\sin^2(\pi\alpha)}{2\pi k} \frac{1}{\sin^2(\theta/2)}. \quad (3.17)$$

We see that if $\alpha$ is integer, there is no scattering at all. However, if $\alpha$ is some arbitrary real number, an interference pattern emerges; particles passing the solenoid on one side interfere with particles that took the other route! This is interesting, because neither of the particles directly encounters the magnetic field. So either the field influences the particle over a finite distance, which would be incompatible with relativity, or there is something more to the potential than just being a mathematical aid; the potential itself influenced the particle. This is what Aharonov and Bohm argued in their article [3], and what makes this effect an important one for understanding electromagnetism.

**Gauge-theoretic interpretation** From the gauge-theoretic point of view, this effect is not so strange – or at least we can describe it neatly: in the region outside of the solenoid the electromagnetic field strength is zero, so the connection $A$ is flat there. We can conclude that there is no local holonomy, but not that there is no monodromy (especially since the region has non-trivial fundamental group: you can walk around the solenoid). Therefore the connection may be non-trivial, and the electrons can feel this as an electromagnetic influence. The reason the Aharonov-Bohm effect may seem so strange at first is that we are used to being able to describe electromagnetism entirely in terms of the curvature (field strength) $F$ (since space(time) is simply connected). We now see that it may be more appropriate to view the connection as the fundamental physical object, not its curvature.

$^8$We will actually take this as the definition of the differential cross section $\frac{d\sigma}{d\Omega}$. This is not an illogical thing to do, since it is intuitively clear that $g(\theta)$ represents some measure of the probability to have a particle scattering off in direction $\theta$. 

36
3.2 Chern-Simons theory

To define the Yang-Mills action (3.7) we employed the Hodge star, which makes explicit use of the metric. This is not a flaw of this particular formulation: the Yang-Mills action does depend on the metric. Chern-Simons theory is a gauge theory that does not need a metric; it is therefore called a topological gauge theory. The starting point of the theory is a smooth manifold \( M \) together with a Lie group \( G \) (the “gauge group”) (usually some properties are demanded of \( G \), but we will not be concrete enough to need them; we only require the Lie algebra of \( G \) to be a matrix Lie algebra); it then considers the trivial principal \( G \)-bundle over \( M \). We will look at the special case where the underlying manifold is 3-dimensional (there are generalisations to all odd dimensions). Given an invariant connection (represented by the 1-form \( A \)), we can form the Lagrangian

\[
\mathcal{L} = \frac{k}{4\pi} \text{tr} \left( A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right)
\]  

(3.18)

\( k \) is some constant). This is a real-valued (because of the trace) 3-form, so we can integrate it over \( M \) to obtain an action. Unfortunately, this Lagrangian is not gauge invariant: a different choice of trivialisation (a gauge transformation) can give a different \( \mathcal{L} \). However, this dependence is not too wild: the difference depends only on the homotopy type of the gauge transformation, and even then \( \mathcal{L} \) changes only by an integer multiple of some constant. If we choose \( k \) in the right way, then \( \mathcal{L} \) changes only by integer multiples of \( 2\pi \) (the normalisation above is conveniently chosen, so that in fact \( k \) must be integer). When the theory is quantised, this means that the expectation value \( e^{i\mathcal{L}} \) does not change at all under gauge transformations, which is all you need to end up with a proper gauge quantum field theory. Now, the so-called Feynman path integral of our quantum field theory can be calculated (this, and actually all we hint at in this section, is done properly in [4]). The result is a number that can be used as an invariant of the manifold \( M \).

This Chern-Simons theory has an unexpected application: it can help to classify and understand knots\(^9\). Very roughly, this is done by multiplying the expectation value \( e^{i\mathcal{L}} \) with the holonomy (parallel transport) around the knot \( K \) we wish to examine. Actually we multiply with the trace of this holonomy: the holonomy itself is an element of \( G \), and it is well-defined only up to conjugation (picking a different basepoint for the holonomy results in conjugation); taking the trace (that is, computing the matrix that corresponds to the element of \( G \) under some representation, and taking the trace of the resulting matrix) gives a number that does not change under change of basepoint. This extra factor (known as a Wilson line\(^10\)) is clearly gauge invariant, since we defined it in terms of (parallel transport around) the connection itself and we made no reference to any trivialisation or representative \( A \). The new Feynman path integral can be calculated, and (for fixed \( M \)) can be used to classify the knot \( K \).

\(^9\)A knot is an embedding of the circle \( S^1 \) in space \((\mathbb{R}^3)\).

\(^10\)This is analogous to electromagnetism: if you view the connection as a physical field on the manifold \( M \), the action (3.18) can be read as the action of free evolution of this field. If you see \( K \) as the world line of a ‘charged’ particle, the Wilson line factor then couples the particle to the field.
After all this interesting hand-waving, we will do a small concrete calculation: we derive the Euler-Lagrange equation from the Chern-Simons Lagrangian (3.18). We can leave out the factor $\frac{k}{4\pi}$ for this purpose. In index notation, the Lagrangian is

$$L = \left\{ A^b_{a\mu} \partial_\nu A^b_{\alpha\rho} + \frac{2}{3} A^a_{b\mu} A^b_{c\nu} A^c_{a\rho} \right\} dx^\mu \wedge dx^\nu \wedge dx^\rho$$  \hspace{1cm} (3.19)$$

(where $A^b_a$ is the $b$-entry of the matrix of real-valued 1-forms $A$, and $A^a_{b\mu}$ is its $\mu$-component), and the Euler-Lagrange equation reads $\frac{\partial L}{\partial A^b_{a\mu}} - \partial_\nu \frac{\partial L}{\partial (\partial_\nu A^b_{a\mu})} = 0$. So let’s first compute:

$$\frac{\partial L}{\partial A^b_{a\mu}} = \left\{ \partial_\nu A^b_{a\rho} + \frac{2}{3} (A^b_{c\nu} A^c_{a\rho} - A^c_{a\nu} A^b_{c\rho} - A^b_{c\rho} A^c_{a\nu}) \right\} dx^\mu \wedge dx^\nu \wedge dx^\rho$$

$$= (dA + 2A \wedge A)^b_a \wedge dx^\mu$$

(3.20)

and

$$\frac{\partial L}{\partial (\partial_\nu A^b_{a\mu})} = A^b_{a\lambda} dx^\lambda \wedge dx^\nu \wedge dx^\mu, \text{ so}$$

$$\partial_\nu \frac{\partial L}{\partial (\partial_\nu A^b_{a\mu})} = \partial_\nu A^b_{a\lambda} dx^\lambda \wedge dx^\nu \wedge dx^\mu = -\partial_\nu A^b_{a\lambda} dx^\nu \wedge dx^\lambda \wedge dx^\mu = -dA^b_a \wedge dx^\mu$$

(3.21)

Plugging this in the Euler-Lagrange equation, we get

$$0 = \frac{\partial L}{\partial A^b_{a\mu}} - \partial_\nu \frac{\partial L}{\partial (\partial_\nu A^b_{a\mu})} = (dA + 2A \wedge A)^b_a \wedge dx^\mu - (-dA^b_a \wedge dx^\mu)$$

$$= 2 (dA + A \wedge A)^b_a \wedge dx^\mu.$$  \hspace{1cm} (3.22)$$

This is true for all $b\mu$ if and only if $dA + A \wedge A = 0$. This says that the curvature of the connection must vanish, i.e. that the connection must be flat!

In more physical language: fields (i.e. connections) that satisfy the equations of motion have zero field strength. Intuitively, this means that this Chern-Simons action is ‘sourceless’: we included no charge that generates a field.

### 3.3 $S^2$ is no Lie group

As a small mathematical application of the fibre-bundle theory we introduced, we show that the sphere $S^2$ cannot be made into a Lie group. This follows directly from the next two facts:

1. The tangent bundle of a Lie group $G$ is trivial. By multiplying from the right with $g^{-1}$, the tangent space above a point $g \in G$ is sent to the tangent space of the identity element, i.e. the Lie algebra $\mathfrak{g}$ of $G$. We therefore can take the following global trivialisation: $h: TG \rightarrow G \times \mathfrak{g}$ : $v \mapsto (p(v), v \cdot (p(v))^{-1})$ (where $p: TG \rightarrow G$ is the projection). This $h$ is indeed a trivialisation: a smooth bijection that commutes with projection.
2. The tangent bundle of $S^2$ is not trivial. If it were, there would be a global section that is nowhere zero (simply compose a constant section $x \mapsto (x, f_0)$ (for some nonzero $f_0$) with (the inverse of) a global trivialisation). The ‘hairy ball theorem’ of algebraic topology asserts that such non-vanishing vector fields do not exist on $S^2$. 

Chapter 4

Characteristic classes

4.1 Introduction

In mathematics, the following situation is quite common: we have some class of objects (say, topological spaces or knots) and we have some equivalence relation on this class (homeomorphism, homotopy). Now, we often want to know if two given objects in this class are equivalent, but this is not always easy to find out. In particular, it can be quite hard to prove that two objects are not equivalent, because this entails proving that something does not exist. This section is about a tool which can help you in such a situation; this is the notion of an invariant, which can help you discover whether two spaces are equivalent.

Example 4.1.1. In topology two spaces are seen as being equivalent if there exists a homeomorphism between the two. To check whether two spaces are equivalent it is often convenient to calculate the fundamental groups of those spaces; if these are not the same, then neither are the topological spaces.

Example 4.1.2. In knot theory we regard two knots as being the same if one can be continuously deformed into the other. When we want to know if two given knots are in fact the same, we can for example calculate the 3-coloring for both knots, and if they don’t come out the same, the knots are not equivalent either.

The fundamental group of a topological space and the 3-coloring of a knot are both example of invariants. An invariant is a way of associating something to each object of the class (the fundamental group to a topological space) in the above situation, such that if two objects are equivalent, then the same thing will be associated to them. For those readers who have seen a little category theory, a nice example of an invariant is given by a functor from the category of objects you want to be able to separate to some other category.

There are two properties of an invariant which determines its usefulness. First is of course calculizability; it should be easier to compute the invariant than to directly check whether the original objects are the same. Second is strictness; a very simple invariant is given by associating the number ‘0’ to each of our objects. While very easy to calculate, it obviously does not help us in any way to check for equivalence of objects. The stricter an invariant, the more objects it can separate, but this usually means that it is more difficult to calculate...
The reason we are going through all this is of course that we ourselves are in the above situation. In chapter one we introduced the concept of a bundle and with it that of a bundle isomorphism. We regard two bundles as being equivalent if there exist some bundle isomorphisms between those two, and we would like to have a way to check for this equivalence. As we explained, this is exactly what an invariant is, so what we are going to look for in the coming chapter is a vector bundle invariant. The reason we are limiting ourselves to vector bundles at this point is the fact that there is a whole class of invariants which can be easily defined for vector bundles.

4.2 Characteristic classes

In this chapter we will often sloppily denote a vector bundle $F \to V \xrightarrow{\pi} M$ by $V$. Let us begin by just giving a definition:

**Definition 4.2.1.** A characteristic class $c$ is a natural transformation associating to all vector bundles $V$ over a manifold $M$ some element $c(V)$ of the cohomology group $H^* M$, such that if $V_1 \cong V_2$ then $c(V_1) = c(V_2)$.

This needs some explaining, of course. First we need to know what is meant by a natural transformation. This is a concept from category theory and a detailed explanation can be found in any book on this subject. In short, if we have two categories $\mathcal{A}$ and $\mathcal{B}$ and two functors $F, G : \mathcal{A} \to \mathcal{B}$ between those categories, then a natural transformation between those functors is a way to relate them to each other. More concrete, for each object $A$ of $\mathcal{A}$ it is a morphism $c(A) : F(A) \to G(A)$ in $\mathcal{B}$, such that for every morphism $f \in \mathcal{A}(A_1, A_2)$, the following diagram commutes:

\[
\begin{array}{ccc}
F(A_1) & \xrightarrow{c(A_1)} & G(A_1) \\
F(f) \downarrow & & \downarrow G(f) \\
F(A_2) & \xrightarrow{c(A_2)} & G(A_2)
\end{array}
\]

How does this apply to our situation? Well, we have two contravariant functors:

$H^* : \text{Man} \to \text{Set}$, the cohomology functor, here seen as a functor to set

$\text{Vect} : \text{Man} \to \text{Set}$, to be explained below.

Here, $\text{Vect}$ associates to a manifold $M$ the set of all isomorphism classes of vector bundles over $M$, and to a map $f : M \to N$ the map between the two corresponding sets induced by the pullbacks $f^*$ of $f$ from vector bundles over $N$ to vector bundles over $M$. It is thus a contravariant functor.

Note that the condition from the definition, that $V_1 \cong V_2$ should imply $c(V_1) = c(V_2)$, is exactly what is necessary to make $c$ act on isomorphism classes of vector bundles, instead of vector bundles.
Characteristic classes provide us with a nice invariant for vector bundles, as $V_1 \simeq V_2$ implies $c(V_1) = c(V_2)$. However, we have not yet seen any examples, and our first job will be to construct some. There is more than one way to do so; we will take a differential-geometric approach, using connections, curvature and invariant polynomials. In the next section, we will recast some of the earlier material on connections and curvature into a form more appropriate to this chapter. After that, we will define what invariant polynomials are and use them to construct characteristic classes.

### 4.3 Covariant derivative revisited

In this section we take another look at the covariant derivative induced by a linear connection on a vector bundle. First we prove a few of its properties. Next we show how to recover the connection from a given differential operator satisfying those properties. This allows us to view the covariant derivative as an alternative presentation of the geometric information contained in a connection. Finally we re-interpret curvature in this new context.

**Definition 4.3.1.** Let $F \rightarrow E \rightarrow M$ be a vector bundle. A CD on $E$ is an $\mathbb{R}$-linear operator $\nabla : \Gamma(E) \rightarrow \Gamma(T^* M \otimes E)$ satisfying the following Leibniz-like property: for all sections $s \in \Gamma(E)$ and functions $f \in C^\infty(M)$,

$$\nabla(fs) = f \nabla s + df \otimes s \quad (4.1)$$

**Proposition 4.3.2.** Given a linear connection on a vector bundle $E$, the covariant derivative $\nabla : \Gamma(E) \rightarrow \Gamma(T^* M \otimes E)$ (defined in section 2.5) is a CD.

**Proof.** From the definition $\nabla = \pi^v \circ d$ we immediately see that $\nabla$ is $\mathbb{R}$-linear.

Proof of (4.1):

$$\nabla(fs) = \pi^v(df \otimes s + ds) \quad (\text{usual Leibniz rule for } d)$$

$$= \pi^v(df \otimes s) + \pi^v(ds) \quad (\text{linearity of the projection})$$

$$= df \otimes s + f \pi^v(ds) \quad (\text{linearity of the connection})$$

$$= df \otimes s + f \nabla s \quad (df \otimes s \text{ is already vertical})$$

(To see that $df \otimes s$ is vertical, recall the various identifications involved: $df \otimes s$ stands for the bundle map $\xi_x \mapsto (df(\xi_x))s(x)$, and $s(x)$ is identified with a vector in the vertical tangent space $VE_x$.)

**Theorem 4.3.3.** Given a vector bundle $E$ with a CD $\nabla$, there is a unique connection on $E$ that induces $\nabla$.

**Proof.** Choose a local trivialisation of $E$, say over an open subset $U \subseteq X$. Transfer the trivial connection on $U \times F$ to a connection on $p^{-1}(U) \subseteq E$, with induced CD $\tilde{\nabla}$. Consider $A := \nabla - \tilde{\nabla}$. By the Leibniz rule for $\nabla$ and $\tilde{\nabla}$, we have

$$A(fs) = f \tilde{\nabla}s + df \otimes s - f \tilde{\nabla}s - df \otimes s = f \tilde{\nabla}s - f \tilde{\nabla}s = fAs \quad (4.2)$$

42
for any section $s$ and function $f$, so the value of $A_s$ at $x$ depends only on the value of $s$ at $x$. Because the CDs are $\mathbb{R}$-linear, $A$ is $\mathbb{R}$-linear, so we can write $(A_s)(x) = L(x)(s(x))$, where $L(x) : F_x \to (T_xM)^* \otimes F_x$ is a linear map – or equivalently, using a familiar identification, $L$ is a $\text{End}(F_x)$-valued 1-form. In local coordinates, this looks like

$$
\nabla = \tilde{\nabla} + A \cong \partial_\mu + A_\mu \quad (4.3)
$$

, where $A_\mu$ is the local representative of $L$. Comparing this with (2.38), we see that $\nabla$ is in fact the covariant derivative induced by the connection on $p^{-1}(U)$ with connection form $A_\mu$.

We must show that this local inducing connection does not depend on the choice of trivialisation, in order to glue different such local inducing connections together to form a global inducing connection. Indeed, let $s$ be a section over $U \subseteq M$, $x \in U$ a point $(s(x) \cong (x,f))$ in our trivialisation) and $\xi \in T_xM$. Then for the tangent vector $\alpha := ds(\xi) \in T_{s(x)}E$ (locally represented by $(\xi,y)$), we have

$$
\alpha \text{ is horizontal} \iff y = -A(\xi)(f) \quad \text{(property of connection form)}
$$
$$
\iff y = \tilde{\nabla}(s)(\xi) - \nabla(s)(\xi) \quad \text{(our choice of connection form)}
$$
$$
\iff y = y - \nabla(s)(\xi) \quad \text{(definition of } \tilde{\nabla})
$$
$$
\iff \nabla(s)(\xi) = 0
$$

. The last statement is independent of the trivialisation. Now, given any tangent vector $\alpha$, we can clearly find a local section $s$ such that $ds(\xi) = \alpha$ (where $\xi = dp(\alpha)$), and we can apply the above criterion. Therefore “being horizontal”, and hence the connection, are independent of trivialisation.

We have proved the ‘existence’ part of the theorem. Uniqueness follows from the fact (see section 2.1.1) that a connection is fully determined by its local connection forms, which in turn are determined by our construction (they are the difference between $\nabla$ and the CD induced by the respective trivial connection).

This result allows us to identify a linear connection with its induced covariant derivative. In fact, what we call a CD, many authors call a connection.

Remarks.

1. From the above proof, we see that the set of linear connections on $E$ is a $T^*M \otimes \text{End}(F)$-torsor: the difference between two connections is an $\text{End}(F)$-valued 1-form (because it does not involve differentiation: it is of order zero or ‘tensorial’), but the connections themselves are not, and there is no canonical choice of ‘zero’ or ‘trivial’ connection.

2. Let’s rephrase the trivialisation-independent description of the connection inducing a given CD: “a section is horizontal at a point iff its covariant derivative there vanishes”. This means that the covariant derivative measures the deviation from being horizontal – not so strange, considering its definition as the vertical projection of the derivative.

\[\text{One says that } A \text{ is an operator of order zero. Similarly, a CD is an operator of order one, i.e. it depends only on the behaviour of its argument up to first order (value and derivative) (this follows from the Leibniz rule).}\]

43
4.3.1 Curvature

Suppose that we have a vector bundle $E$ with a linear connection. The covariant derivative $\nabla$ sends sections of $E$ to $E$-valued 1-forms. Another way to look at sections is as $E$-valued 0-forms. This suggests that $\nabla : \Gamma(E) \to \Gamma(\Lambda^1(T^*M) \otimes E)$, so we see that $\nabla^2$ reduces to the case $\nabla$. First of all, let $R$ be a section, and for convenience denote $\nabla s =: \alpha \otimes s$, for a section and $p$-form $\alpha$:

$$\nabla(\alpha \otimes s) := d\alpha \otimes s + (-1)^p \alpha \wedge \nabla s \quad \text{(definition (4.4))}$$

$$= df \wedge s + f d\alpha \otimes s + f(-1)^p \alpha \wedge \nabla s \quad \text{(distribution of $d$ over $\wedge$)}$$

$$= df \wedge s + f \nabla(\alpha \otimes s) \quad \text{(definition (4.4))}$$

$$= df \wedge s + df \wedge \nabla \rho \quad . \text{ The general case follows because both sides are $\mathbb{R}$-linear in $\rho$.}$$

Now, let’s investigate whether $\nabla$ forms a cochain – that is, whether $\nabla^2 = 0$. First of all, let $\alpha$ be a $p$-form and $s$ a section, and for convenience denote $\nabla s =: \beta^j \otimes t_j$ (for 1-forms $\beta^j$ and section $t_j$); then

$$\nabla^2(\alpha \otimes s) = \nabla (d\alpha \otimes s + (-1)^p \alpha \wedge \nabla s)$$

$$= d^2 \alpha \otimes s + (-1)^{p+1} d\alpha \wedge \nabla s + (-1)^p d(\alpha \wedge \beta^j) \otimes t_j - \alpha \wedge \beta^j \wedge \nabla t_j$$

$$= -(-1)^p d\alpha \wedge \beta^j \otimes t_j + (-1)^p d(\alpha \wedge \beta^j) \otimes t_j - \alpha \wedge \beta^j \wedge \nabla t_j$$

$$= \alpha \wedge \nabla(\beta^j \otimes t_j)$$

$$= \alpha \wedge \nabla^2 s \quad (4.5)$$

, so we see that $\nabla^2$ acts only on the $E$-part of an $E$-valued $p$-form, and we can reduce to the case $p = 0$. In that case we see from the lemma above and (4.4) (and $d^2 = 0$) that

$$\nabla^2(f s)$$

$$= \nabla(df \otimes s + f \nabla s) = d^2 f \otimes s - df \wedge \nabla s + df \wedge \nabla s + f \nabla^2 s \quad (4.6)$$

$$= f \nabla^2 s$$

---

A general section of the tensor product bundle is a sum of such pure terms $\alpha \otimes s$. Because the right-hand side is $\mathbb{R}$-linear in $\alpha$ and $s$, (4.4) is in fact well-defined.

44
so at least $\nabla^2$ is an operator of order zero, and we can identify it with a linear map from $E$ to $\Lambda^2(T^*M) \otimes E$, i.e. an $\text{End}(E)$-valued 2-form. However, this need not be the zero form! In fact, it is equal to that other $\text{End}(E)$-valued 2-form we met before:

**Proposition 4.3.5.** For any $E$-valued $p$-form $\rho$, we have $\nabla^2 \rho = R \wedge \rho$, where $R$ is the curvature of the connection associated to $\nabla$.

**Proof.** We already saw that $\nabla^2 (\alpha \otimes s) = \alpha \wedge \nabla^2 s$, so it suffices to prove the case $p = 0$. Now, let $s$ be a section. To prove that the forms $\nabla^2 s$ and $Rs$ are equal, it suffices to prove that they are equal at every point. Choose a local trivialisation of $E$; we will treat the trivialisation as an equality. Denote the connection form by $\Lambda$, and write $\nabla s = dx^i \otimes t_j$, where thus $t_j = \partial_j s + \Lambda_j s$. We compute

\[
\nabla^2 s = \nabla(dx^i \otimes t_j) = d^2 x^i \otimes t_j - dx^i \wedge \nabla t_j
\]

\[
= -dx^i \wedge dx^j (\partial_i (\partial_j s + \Lambda_j s) + \Lambda_i (\partial_j s + \Lambda_j s))
\]

\[
= dx^i \wedge dx^j (\partial_i (\Lambda_j s) + \Lambda_i (\partial_j s + \Lambda_j s))
\]

\[
= dx^i \wedge dx^j ((\partial_i \Lambda_j) s + \Lambda_j \partial_i s + \Lambda_i \partial_j s + \Lambda_i \Lambda_j s)
\]

\[
= dx^i \wedge dx^j ((\partial_i \Lambda_j) s + \Lambda_i \Lambda_j s)
\]

\[
= (d\Lambda) s + [\Lambda, \Lambda] s
\]

which was to be proved. 

Concisely said: curvature is the obstruction to the covariant derivative being a cochain map.

### 4.4 Invariant polynomials

Again, let’s start with a definition.

**Definition 4.4.1.** Let $\mathfrak{gl}_m(\mathbb{C})$ denote the Lie algebra of $m \times m$ matrices over $\mathbb{C}$. A **invariant polynomial** on $\mathfrak{gl}_m(\mathbb{C})$ is a polynomial function $P : \mathfrak{gl}_m(\mathbb{C}) \to \mathbb{C}$ such that for all $X, Y \in \mathfrak{gl}_m(\mathbb{C})$, $P(XY) = P(YX)$.

As above, we have some additional explaining to do about this definition; we have to make clear what we mean by a polynomial function from $\mathfrak{gl}_m(\mathbb{C})$ to $\mathbb{C}$, for the first definition for polynomials on $\mathfrak{gl}_m(\mathbb{C})$ that comes to mind would take values in $\mathfrak{gl}_m(\mathbb{C})$ itself, not in $\mathbb{C}$. What we mean in this case is a map $P : \mathfrak{gl}_m(\mathbb{C}) \to \mathbb{C}$ that is a polynomial in the entries of the matrix in the argument.

As a first example, the trace and determinant of matrices are invariant polynomials.

Now, let’s see how this notion helps us to construct characteristic classes. Let $V$ be a complex vector bundle, over some manifold $M$, and equipped with a connection. We denote by $R$ the curvature of the connection; it is a 2-form on $\mathfrak{gl}_m(\mathbb{C})$ to $\mathbb{C}$, for the first definition for polynomials on $\mathfrak{gl}_m(\mathbb{C})$ that comes to mind would take values in $\mathfrak{gl}_m(\mathbb{C})$ itself, not in $\mathbb{C}$. What we mean in this case is a map $P : \mathfrak{gl}_m(\mathbb{C}) \to \mathbb{C}$ that is a polynomial in the entries of the matrix in the argument.

As a first example, the trace and determinant of matrices are invariant polynomials.

Now, let’s see how this notion helps us to construct characteristic classes. Let $V$ be a complex vector bundle, over some manifold $M$, and equipped with a connection. We denote by $R$ the curvature of the connection; it is a 2-form on $\mathfrak{gl}_m(\mathbb{C})$ to $\mathbb{C}$, for the first definition for polynomials on $\mathfrak{gl}_m(\mathbb{C})$ that comes to mind would take values in $\mathfrak{gl}_m(\mathbb{C})$ itself, not in $\mathbb{C}$. What we mean in this case is a map $P : \mathfrak{gl}_m(\mathbb{C}) \to \mathbb{C}$ that is a polynomial in the entries of the matrix in the argument.
$M$ with values in $\text{End}(F)$ (recall that $V$ was actually short for $F \to V \xrightarrow{\pi} M$). Choosing a basis for the fibre above some point $x \in M$, $R$ locally (i.e. at $x$) becomes a matrix-valued 2-form. An equivalent way to look at this, is to regard $R$ locally as a matrix of ordinary 2-forms. If now $P$ is an invariant polynomial, we can apply it to this matrix and get an even-dimensional form, which we will denote by $P(R)$. Of course, $R$ is not an ordinary $\mathfrak{gl}_n(\mathbb{C})$ matrix, but a matrix of 2-forms, so we have to state what we mean by applying $P$ to it. In order to make sure the outcome is again a form, the only logical way to do this is to let $P$ act on the entries of the matrix as it would in the ordinary case, but replace the product of two entries with the wedge product of differential forms.\(^4\)

For example, if $F$ is a two-dimensional vector space, $P_1$ is the trace, $P_2$ is the determinant, and $\alpha, \beta, \gamma$ and $\delta$ are differential forms, then:

\[
P_1 \left( \begin{array}{cc}
\alpha & \beta \\
\gamma & \delta 
\end{array} \right) = \alpha + \delta \\
\end{array}
\]

\[
P_2 \left( \begin{array}{cc}
\alpha & \beta \\
\gamma & \delta 
\end{array} \right) = \alpha \wedge \delta - \beta \wedge \gamma
\]

The first is a two-form, the second a four-form.

Finally we note that, since $P$ is invariant, $P(R)$ does not depend on the choice of basis we made earlier.\(^5\) Therefore, we can define the differential form $P(R)$ globally.

So what have we done? We have defined the notion of an invariant polynomial, and we have shown that this provides us with a way to associate a differential form to the curvature on a vector bundle. (Since curvature comes from a connection, we could just as well say that it is a way to associate a differential form to the connection on a vector bundle.) While this may be a nice thing to have, it is not what we were looking for in the first place; we wanted to associate a differential form to the vector bundle itself, rather than to a connection defined upon it.

Luckily for us, there is a very nice theorem to save the day. It tells us that, although our construction of a differential form very much uses the curvature (and thus also the connection), the cohomology class of the result actually does not depend upon the choice of connection at all! As we shall see, this means that any invariant polynomial provides us with a characteristic class by the prescription “apply it to the curvature of any connection on your bundle". Let us now state and prove this theorem.

**Theorem 4.4.2.** For any invariant polynomial $P : \mathfrak{gl}_n(\mathbb{C}) \to \mathbb{C}$, the differential form $P(R)$ (where $R$ is the curvature of some vector bundle $V$) is closed and its de Rahm cohomology class is independent of the choice of connection.

**Proof.** To show that $P(R)$ is closed means to show that $dP(R) = 0$, where $d$ denotes the exterior derivative. We use a few tricks to accomplish this. First, let us assume that $P$ is homogeneous of degree $k$. This can be done without loss

\(^4\)By definition, $P$ is an ordinary, commutative polynomial in the entries of a matrix, i.e. the different ‘variables’ (entries) commute. Therefore, we can only substitute elements of a commutative ring for the variables of $P$ (otherwise, $P(R)$ would not be well-defined). Because the curvature matrix $R$ is a matrix of two-forms, this is no problem; two-forms commute with each other.

\(^5\)A matrix with respect to any basis can be written with respect to any other basis by a transformation of the form $X \to UXU^{-1}$, and we have $P(UXU^{-1}) = P(U^{-1}UX) = P(X)$. 

46
of generality, because d is a linear operator, so if \(dP_1(R) = 0\) and \(dP_2(R) = 0\), then \(d(P_1(R) + P_2(R)) = 0\).

Now that \(P\) is a homogeneous polynomial, we can use the polarization \(\hat{P}\) of \(P\). This is defined as follows: if we view \(P(t_1X_1 + \cdots + t_kX_k)\) as a polynomial in \(t_1, \ldots, t_k\), then \(k!P(X_1, \ldots, X_k)\) is the coefficient of \(t_1 \cdots t_k\) in its expansion, seen as a polynomial in \(X_1, \ldots, X_k\). To get a feeling for what is going on here, note that \(\hat{P}\) is a polynomial in \(m^2k\) variables (remember that, even though we write \(P(X)\), \(P\) is a polynomial in the \(m^2\) entries of the matrix \(X\)). Furthermore, \(\hat{P}\) is symmetric; since the \(t_1, \ldots, t_k\) commute, each permutation of \(X_1 \cdots X_k\) enters in the coefficient of \(t_1 \cdots t_k\). To relate \(P\) and \(\hat{P}\) we have \(P(X) = \hat{P}(X, \ldots, X)\).

- This is best appreciated by an example: take \(P(x, y) = xy + x^2\) as your homogeneous polynomial (of degree 2); then
  \[
P(t_1x_1 + t_2x_2, t_1y_1 + t_2y_2) = \cdots + (x_1y_2 + x_2y_1 + x_1x_2 + x_2x_1)t_1t_2
  \]

  

  so

  \[
  2!\hat{P}(x_1, x_2; y_1, y_2) = x_1y_2 + x_2y_1 + x_1x_2 + x_2x_1
  \]

  and we see that

  \[
  2!\hat{P}(x, x; y, y) = xy + xy + xx + xx = 2P(x, y).
  \]

- To give a sketch of the proof of the general fact, we first note that we only need to look at monomials, since if \(P(X) = \hat{P}(X, \ldots, X)\) holds for two polynomials, it will also hold for the sum. Next, we see, for a monomial \(P\), that the terms of \(P(t_1X_1 + \cdots + t_kX_k)\) are obtained by picking for each factor of degree one\(^6\) a term \(t_iX_i\) from \(t_1X_1 + \cdots + t_kX_k\) and substituting the appropriate entry of \(X_k\). However, \(\hat{P}\) is the coefficient of \(t_1 \cdots t_k\) in this expansion, so only the terms where we picked a different \(i\) for each factor of degree one enter in \(\hat{P}\). These terms individually all look like the \(\hat{P}\), only each factor of degree one has an extra index; furthermore, these indices are all different within a term. So, if we replace \(X_i\) by \(X\) in \(\hat{P}\) for all \(i\), each term will be equal to \(P\), and since there are \(k!\) possible orders in which to pick the \(i\)'s, we exactly compensate the \(\frac{1}{k!}\) in the definition of \(\hat{P}\), and we conclude: \(\hat{P}(X, \ldots, X) = P(X)\).

In the same way as we did above for \(P\), we can interpret \(\hat{P}\) as a polynomial in matrices of two-forms instead of a polynomial in ordinary matrices. Even better, since two-forms also commute with three-forms and one-forms, we may substitute a one- or three-form in one of the entries of \(\hat{P}\) as long as all the other arguments are two-forms. We will use this later on for \(\hat{P}(dR, R, \ldots, R)\) and \(\hat{P}(\theta, R, \ldots, R)\), where \(\theta\) will be some one-form.\(^7\)

To complete the proof of the first part of the theorem, choose a point \(x_0 \in M\) and work in geodesic coordinates based at this point, so that the connection

\(^6\)Example: one of the \(x\)'s, \(y\)'s or \(z\)'s in the monomial given by \(P(x, y, z) = xxyzzzz\).

\(^7\)For all this to work, we also need that \(\hat{P}\) is again invariant in some way; we want the matrix form of \(\hat{P}(R_1, \ldots, R_k)\) to be independent of the choice of local basis. As a matter of fact, we have:

\[
P(h_1X_1h^{-1} + \cdots + h_kX_kh^{-1}) = P(h(t_1X_1 + \cdots + t_kX_k)h^{-1}) = P(t_1X_1 + \cdots + t_kX_k)
\]

, so since \(\hat{P}\) is the coefficient of \(t_1 \cdots t_k\) in the above expansion, this is okay.
one-form and (therefore) the exterior derivative of the curvature vanish at \(x_0\): 
\[ A(x_0) = dR(x_0) = 0. \] (Such coordinates can always be found; see [5, chapter 1].) Then we have (to be explained below):

\[ dP(R)(x_0) = d\bar{P}(dR, R, \ldots, R)(x_0) = 0 \quad (4.10) \]

where we understand \(dP(X)\) to mean \(d(P(X))\) (as is the only logical interpretation), and the same for \(\bar{P}\).

- The first equality is just the property of \(\bar{P}\) we saw above.
- The second follows from the Leibniz rule for the exterior derivative: since each term in the expansion of \(\bar{P}(R, \ldots, R)\) is a product of \(k\) entries from \(R\), it will create \(k\) terms in \(d\bar{P}(R, \ldots, R)\) – one for every possible position of \(d\) in this term. From the symmetry of \(\bar{P}\), they will all appear \(k!\) times. On the other hand, \(\bar{P}(dR, R, \ldots, R)\) is almost the same as \(\bar{P}(R, \ldots, R)\); only in the \((k-1)!\) terms where the first entry in \(\bar{P}\) is in first place, we get a \(d\) in front of the first factor; for the \((k-1)!\) terms where it is in second place we get a \(d\) in front of the second factor, etc. As you see, this is the same as with \(d\bar{P}(R, \ldots, R)\), with the difference that there we had a factor \(k!\) in front. This amounts to exactly the factor \(k\) in the second equality. If you do not follow this argument, just try it out for yourself; it will become a lot clearer.
- The third equality follows from the linearity of \(\bar{P}\) and the fact that, by our choice of frame, \(dR(x_0) = 0\).

Because \(P\) is invariant, \(P(R)\) does not depend on our choice of frame. This implies that (4.10) holds in any frame. Since the point \(x_0\) was chosen arbitrarily, it also holds at any point. Thus, \(dP(R) = 0\), or, in other words, \(P(R)\) is closed.

Next, we need to show that the cohomology class of \(P(R)\) does not depend on the connection chosen. This means that for two connections with curvature \(R_0\) and \(R_1\) the difference \(P(R_1) - P(R_0)\) should be exact; there should be some form \(TP\) such that \(dTP = P(R_1) - P(R_0)\). We will do this by explicitly constructing the form \(TP\).

Let two connections \(\nabla_0\) and \(\nabla_1\) be given. We define a family of linear operators \(\nabla_t\) on the space of sections \(C^\infty(V)\) on \(V\) by:

\[ \nabla_t = t\nabla_1 + (1-t)\nabla_0 \]

We claim that for all \(t\), \(\nabla_t\) defines a connection on \(V\). Since \(\nabla_t\) is obviously a linear differential operator, we only have to check that it obeys equation (4.1). Let us do this directly:

\[ \nabla_t(fs) = t\nabla_1(fs) + (1-t)\nabla_0(fs) \\
= tf \otimes s + tf \nabla_1s + (1-t)df \otimes s + (1-t)f\nabla_0s \\
= df \otimes s + f(t\nabla_1 + (1-t)\nabla_0)s = df \otimes s + f\nabla_t s \]

Note that for this to work, we really need the fact that \(t\) and \(1-t\) sum to 1; otherwise the first term would get some factor in front.

We also want to find the connection one-form for \(\nabla_t\). As might be expected, this is given by

\[ A_t = tA_1 + (1-t)A_0 = A_0 + t\theta, \quad (4.11) \]

48
where \( \theta = A_1 - A_0 \). This follows directly from the fact (see the proof of Theorem 4.3.3) that the connection one-form is the difference between the connection and the trivial connection: \( A_1 = \nabla_t - \nabla \).

Furthermore, we will want to be able to differentiate and integrate differential forms with respect to the parameter \( t \), so we have to state what we will mean by that. If \( \alpha_t \) is a family of \( k \)-forms, then for a \( k \)-tuple \( \xi \) of tangent vectors we can regard \( \alpha (\xi) \) as a map \( \mathbb{R} \to \mathbb{R} \), sending \( t \) to \( \alpha_t(\xi) \). This map can of course be differentiated in the usual sense, providing us with a new map \( \frac{d}{dt}(\alpha_t(\xi)) : \mathbb{R} \to \mathbb{R} \) for every \( k \)-tuple \( \xi \) of tangent vectors. Now we define the \( k \)-form \( \frac{d\alpha}{dt} \), by stating that it sends \( \xi \) to \( \frac{d}{dt}(\alpha_t(\xi)) \). Integration is completely analogous: \( \int \alpha_t dt \) sends \( \xi \) to \( \int \alpha_t(\xi) dt \), where \( \alpha (\xi) \) is again seen as a function \( \mathbb{R} \to \mathbb{R} \) for all \( \xi \).

Now we have, by the main theorem of calculus:

\[
P(R_1) - P(R_0) = \int_0^1 \frac{d}{dt} \tilde{P}(R_t, \ldots, R_t) dt,
\]

where \( R_t \) is the curvature of the connection \( \nabla_t \).

We would like to be able to pull the same trick for the time derivative as we did above for the exterior derivative; pull it through \( \tilde{P} \) so that instead of working on the whole thing, it is now only in front of the first argument (all at the cost of a factor \( k \)). The only property of the exterior derivative we used in our argument there, was that it obeys some sort of Leibniz rule. If we can show that \( \frac{d}{dt} \) obeys this same property, then the rest of the argument is just the same. Let us therefore look at (for families of \( k_1 \)- and \( k_2 \)-forms \( \alpha_t \) and \( \beta_t \), and \( \alpha_{k_1} \) and \( \alpha_{k_2} \)-vectors \( \xi_1 \) and \( \xi_2 \)):

\[
\frac{d}{dt}(\alpha_t \wedge \beta_t)(\xi_1, \xi_2) = \frac{d}{dt} \left( \alpha_t(\xi_1) \beta_t(\xi_2) - \alpha_t(\xi_2) \beta_t(\xi_1) \right)
= \frac{d\alpha_t}{dt}(\xi_2) \beta_t(\xi_1) + \alpha_t(\xi_1) \frac{d\beta_t}{dt}(\xi_2) - \frac{d\alpha_t}{dt}(\xi_1) \beta_t(\xi_2) - \alpha_t(\xi_1) \frac{d\beta_t}{dt}(\xi_1)
= \left( \frac{d\alpha_t}{dt}(\xi_1) \beta_t(\xi_2) - \frac{d\alpha_t}{dt}(\xi_2) \beta_t(\xi_1) \right) + \left( \alpha_t(\xi_1) \frac{d\beta_t}{dt}(\xi_2) - \alpha_t(\xi_2) \frac{d\beta_t}{dt}(\xi_1) \right)
= \left( \frac{d\alpha_t}{dt} \wedge \beta_t \right)(\xi_1, \xi_2) + (\alpha_t \wedge \frac{d\beta_t}{dt})(\xi_1, \xi_2) = \left( \frac{d\alpha_t}{dt} \wedge \beta_t + \alpha_t \wedge \frac{d\beta_t}{dt} \right)(\xi_1, \xi_2)
\]

So by definition

\[
\frac{d}{dt}(\alpha_t \wedge \beta_t) = \left( \frac{d\alpha_t}{dt} \wedge \beta_t + \alpha_t \wedge \frac{d\beta_t}{dt} \right),
\]

or in other words, the time derivative satisfies the Leibniz rule. Therefore, by the same reasoning as for the exterior derivative, we may conclude:

\[
\frac{d}{dt} \tilde{P}(R_t, \ldots, R_t) = k \tilde{P}(\frac{dR_t}{dt}, R_t, \ldots, R_t)
\]

and, by equation (4.12);

\[
P(R_t) - P(R_0) = k \int_0^1 \tilde{P}(\frac{dR_t}{dt}, R_t, \ldots, R_t)
\]

Before we go on, it is a good idea to look back a little at what we have done, and what we still need to accomplish. What we wanted to show was that the
cohomology class of the differential form $P(R)$ does not actually depend upon the connection (and thus curvature) chosen. In other words, for two different connections on our bundle, with curvatures $R_0$ and $R_1$, we want to show that $P(R_1) - P(R_0)$ is exact. Or, in yet other words, there exists some differential form $TP(R_0, R_1)$ such that $dTP(R_0, R_1) = P(R_1) - P(R_2)$.

We now claim that the following $TP$ does the job:

$$TP(R_0, R_1) = k \int_0^1 \tilde{P}(\theta, R_t, \ldots, R_t) \, dt \in \Lambda^{2k-1}(T^*M).$$

What remains to be shown is that

$$dTP(R_0, R_1) = d(k \int_0^1 \tilde{P}(\theta, R_t, \ldots, R_t) \, dt) = k \int_0^1 \tilde{P}(\frac{dR_t}{dt}, R_t, \ldots, R_t)$$

i.e.

$$d\tilde{P}(\theta, R_t, \ldots, R_t) = \tilde{P}(\frac{dR_t}{dt}, R_t, \ldots, R_t). \quad (4.13)$$

Since $\tilde{P}$ is invariant, both sides of equation (4.13) are independent of choice of frame. Let $x_0 \in M$ and $t_0 \in \mathbb{R}$. We use this freedom to set $dR_{t_0}(x_0) = A_{t_0}(x_0) = 0$ (i.e. we work in geodesic coordinates based at $x_0$). Working at the point $x_0$, we have:

$$\frac{dR_t}{dt} = \frac{d}{dt}(dA_t + A_t \wedge A_t) = \frac{d}{dt}(dA_0 + td\theta + A_t \wedge A_t)$$

$$= d\theta + \frac{dA_t}{dt} \wedge A_t + A_t \wedge \frac{dA_t}{dt}$$

$$\frac{dR_t}{dt}|_{x_0, t_0} = d\theta.$$

The first equality is just equation (2.32), the second is equation (4.11). The third follows from the fact that $dA_0$ and $d\theta$ do not depend on $t$, as well as from the Leibniz-rule we derived above for the time derivative. The last equality uses $A_{t_0}(x_0) = 0$ (by choice of coordinates).

On the other hand:

$$d\tilde{P}(\theta, R_t, \ldots, R_t)|_{x_0, t_0} = \tilde{P}(d\theta, R_t, \ldots, R_t)|_{x_0, t_0},$$

which follows from the Leibniz rule for the exterior derivative and the fact that in these coordinates $dR_{t_0}(x_0) = 0$. Putting these last two together results in:

$$d\tilde{P}(\theta, R_t, \ldots, R_t)|_{x_0, t_0} = \tilde{P}(\frac{dR}{dt}, R_t, \ldots, R_t)|_{x_0, t_0},$$

which, noting that the points $x_0$ and $t_0$ were entirely arbitrary, proves equation (4.13). With that, we now have

$$P(R_1) - P(R_0) = d(TP(R_0, R_1))$$

completing the proof of the theorem. □

Let us once more review what has been going on. We began by defining the concept of a characteristic class; loosely speaking, it is a way to associate a cohomology class on the base manifold to a vector bundle. We wanted to
have this concept because it can help us distinguish different vector bundles; it is an example of a vector-bundle invariant. Next, we defined the notion of an invariant polynomial. We showed that we could use such a polynomial as a way to associate a differential form to the curvature of the connection on a vector bundle. We noted, however, that this was not what we originally set out to find; we wanted to associate one cohomology class to a vector bundle; not one for every possible connection on that bundle.

In the end, the solution was found in the theorem we just proved; it shows that, although the differential form associated to the curvature by an invariant polynomial may differ according to the connection chosen, the cohomology class of this differential form does not. Therefore, invariant polynomials provide us with a very easy recipe to create characteristic classes; if $P$ is an invariant polynomial, then we can define a characteristic class $c$ by stating that it associates to a vector bundle $V$ the cohomology class of $P(R)$, where $R$ is the curvature of an arbitrary connection on $V$.

4.5 Chern classes

We will now study an important set of characteristic classes on a complex vector bundle; the Chern classes. They come from a special set of invariant polynomials, and it turns out that any characteristic class of complex vector bundles coming from an invariant polynomial can be written as a polynomial in the Chern classes.

**Definition 4.5.1.** Let $k \in \mathbb{N}$, then define $c_k : \text{gl}_m(\mathbb{C}) \to \mathbb{C}$ to be the invariant polynomial given by $c_k(X) = (-2\pi i)^{-k} \text{tr}(\Lambda^k X)$. Here we mean by $\Lambda^k X : \Lambda^k \mathbb{C}^m \to \Lambda^k \mathbb{C}^m$ the linear transformation given for any pure $k$-form $v_1 \wedge \cdots \wedge v_k$ by

$$\Lambda^k X(v_1 \wedge \cdots \wedge v_k) = (Xv_1) \wedge \cdots \wedge (Xv_k).$$

Now the $k^{\text{th}}$ **Chern class** is defined to be the characteristic class associated to $c_k$.

As has been the case a couple of times before, we need to justify this definition. What we did is define some maps $c_k : \text{gl}_m(\mathbb{C}) \to \mathbb{C}$, and call them invariant polynomials, while in fact it is not clear at all that they are invariant, or even polynomials. Let us therefore check this now. First invariance, because this is easy. On any pure $k$-form $\alpha = v_1 \wedge v_2 \wedge \cdots \wedge v_k \in \Lambda^k \mathbb{C}^m$,

$$\Lambda^k(\alpha) = \Lambda^k(Xv_1 \wedge Xv_2 \wedge \cdots \wedge Xv_k) = \Lambda^k X(\Lambda^k(\alpha)),$$

so by linearity $\Lambda^k(\alpha) = \Lambda^k(X)\Lambda^k(\alpha)$, and hence

$$\text{tr}(\Lambda^k(\alpha)) = \text{tr}(\Lambda^k X \Lambda^k(\alpha)) = \text{tr}(\Lambda^k Y \Lambda^k X) = \text{tr}(\Lambda^k(YX)),$$

c_k is indeed invariant. To see that $c_k(X)$ is a polynomial in the entries of $X$, we will show that the entries of $\Lambda^k X$, when seen as a matrix, are all polynomials in the entries of $X$. Then, because the trace is a polynomial in the entries of $\Lambda^k X$, we will be able to conclude that $c_k(X)$ is a polynomial in the entries of $X$. 51
We write $e(i_1, \ldots, i_k)$ for the vector $e_{i_1} \wedge \cdots \wedge e_{i_k} \in \Lambda^k \mathbb{C}^m$, and take as a basis of $\Lambda^k \mathbb{C}^m$ the set of all $e_I$, where $I$ is a strictly increasing $k$-tuple of elements of $\{1, \ldots, m\}$. We compute the $I,J$-entry of the matrix of $\Lambda^k X$ with respect to this basis:

$$\Lambda^k X(e_I) = X e_{j_1} \wedge \cdots \wedge X e_{j_k} = \left( \sum_{i_1=1}^{m} X_{i_1,j_1} e_{i_1} \right) \wedge \cdots \wedge \left( \sum_{i_k=1}^{m} X_{i_k,j_k} e_{i_k} \right),$$

(4.14)

where $X_{a,b}$ is the $a,b$-entry of the matrix of $X$, with respect to the $e_i$-basis) and $(\Lambda^k X)_{I,J}$ is the the $I$-component of this:

$$(\Lambda^k X)_{I,J} = \sum_{\pi \in S_m} \epsilon(\pi) X_{\pi(i_1),j_1} X_{\pi(i_2),j_2} \cdots X_{\pi(i_k),j_k} \quad \text{(4.14)}$$

where $S_m$ is the permutation group on $\{1, \ldots, m\}$ and $\epsilon(\pi)$ is the sign of the permutation $\pi$. Thus we have shown that the entries of $\Lambda^k X$ are polynomials in the entries of $X$, and therefore $c_k$ is a polynomial in the entries of $X$. Since it was also invariant, we have justified the statement that $c_k$ is an invariant polynomial.

Next, we prove the statement made at the beginning of this paragraph:

**Theorem 4.5.2.** The polynomials $c_k$ generate the ring of invariant polynomials.

**Proof.** Let $P : \mathfrak{gl}_m(\mathbb{C}) \to \mathbb{C}$ be an invariant polynomial. The idea of the proof will be to use the invariance of $P$ to reduce to diagonal matrices, and then apply the algebraic fact that symmetric polynomials are generated by the elementary symmetric polynomials. We will need to extend this with a continuity argument, because not all matrices are diagonalisable.

In the remainder of the proof, let $G \subseteq \mathfrak{gl}_m(\mathbb{C})$ denote the subset of diagonalisable matrices, and $D \subseteq G$ the set of diagonal ones. Recall that $G$ is dense in $\mathfrak{gl}_m(\mathbb{C})$.

Let us look at the restriction $P|_D$ of $P$ to the diagonal matrices in $\mathfrak{gl}_m(\mathbb{C})$. Since $P$ is invariant, and since the diagonal entries of a diagonal matrix can be interchanged by conjugation, we see that it must be possible to write $P|_D$ as a symmetric polynomial in the diagonal entries.

If $X \in G$, we can diagonalise it by conjugation with some matrix $U$. As we know from linear algebra, it then has the eigenvalues of $X$ on the diagonal. This gives us:

$$P(X) = P(U X U^{-1}) = P|_D(U X U^{-1}) .$$

Since $P|_D$ was a symmetric polynomial in the diagonal entries, and since in this case they are the eigenvalues of $X$, we see that on diagonalisable matrices $P$ equals a symmetric polynomial in the eigenvalues. In other words, we can make the following commutative diagram:
where \( \omega : \mathfrak{gl}_m(\mathbb{C}) \to \mathbb{C}^m/S_m \) is the map sending a matrix to its eigenvalues (and dividing out by the permutations \( S_m \), because we want the eigenvalues to remain unordered), and where \( f \) is some map (a symmetric polynomial, as we argued above) making the diagram commutative. As a polynomial, \( f \) is equal to \( P|_D \), when the last is seen as a polynomial in the diagonal entries alone.

Now, \( P \) is continuous (being a polynomial), so \( P|_G : G \to \mathbb{C} \) is continuous on a dense subset of \( \mathfrak{gl}_m(\mathbb{C}) \), so if it can be extended to a continuous function on \( \mathfrak{gl}_m(\mathbb{C}) \), this extension is unique. But \( P \) obviously constitutes one such extension, while another is given by \( f \circ \omega \): \( \omega \) is continuous (the reader should prove this if it is not familiar) and so is \( f \) (it is a polynomial), and we argued above that \( f \circ \omega \) extends \( P|_G \) (this is expressed by the diagram). We conclude that \( P = f \circ \omega \).

We have now shown that \( P \) equals some symmetric polynomial in the eigenvalues of its argument. But the ring of symmetric polynomials in \( m \) variables is generated by the elementary symmetric polynomials of \( m \) variables. To prove the theorem, we just have to show that \( c_k \) constitutes the \( k \)th elementary symmetric polynomial in \( m \) variables, applied to the eigenvalues of its argument.

Since the trace of a matrix is nothing but the sum of the diagonal entries, we have to ask ourselves what the diagonal entries of the matrix \( \Lambda^k X \) are. We know the form of the entries of \( \Lambda^k X \): it is given by equation (4.14). To find the diagonal ones, we have to realise that those are just the ones where \( i_l = j_l \) for all \( l \in \{1, \ldots, k\} \). Furthermore, we want each one to count exactly once, so we should order them (remember that the basis of \( \Lambda^k \mathbb{C}^m \) is given by \( k \)-fold wedge products of basis vectors of \( \mathbb{C}^m \), where for each combination of basis vectors, only one ordering is taken). The sum of the diagonal entries then becomes:

\[
\text{tr}(\Lambda^k X) = \sum_{i_1 < \cdots < i_k} \sum_{\pi \in S_m} \epsilon(\pi) X_{\pi(i_1), i_1} \cdots X_{\pi(i_k), i_k} . \tag{4.15}
\]

Now we use the invariance of \( c_k \) to put \( X \) in its Jordan normal form. As we know from linear algebra, this is always possible. Now we know that \( X \) has its eigenvalues on the diagonal, and maybe has some ones directly above it. The rest of the entries is zero. With this in mind, let us look at equation (4.15) again. We see that, for each term in the inner sum, if we have some entry of \( X \) in it which is not on the diagonal of \( X \) (i.e., \( i_l \neq \pi(i_l) \) for some \( l \)), then there must be at least one other which is not either. Even better, if there is some \( l \) such that \( i_l < \pi(i_l) \), then there must be an \( l' \) such that \( i_{l'} > \pi(i_{l'}) \), and vice versa (because \( \pi \) is a permutation of a finite set).

But then we see, that if a term in the inner sum contains entries that are not on the diagonal, it will contain at least one entry from below the diagonal, and since \( X \) was a Jordan matrix, these are all zero. We can conclude that only the terms in the inner sum that contain only entries from the diagonal of \( X \) will remain. These are of course precisely the ones for which \( \pi \) is the identity. This leaves us with:

\[
\text{tr}(\Lambda^k X) = \sum_{i_1 < \cdots < i_k} X_{i_1, i_1} \cdots X_{i_k, i_k} = \sum_{i_1 < \cdots < i_k} \lambda_{i_1} \cdots \lambda_{i_k} ,
\]

where the last equality is true because a Jordan matrix has its eigenvalues on the diagonal.

This shows us that \( c_k \) is indeed the \( k \)th elementary symmetric polynomial in the eigenvalues of \( X \) (apart from the normalisation constant \( (-2\pi)^{-k} \), which
is not important since all we want to do is generate the ring of symmetric polynomials. They generate the ring of symmetric polynomials in the eigenvalues of $X$, which by our earlier arguments was equal to the ring of invariant polynomials, and so we have proved the theorem.

A quick recapitulation of what we have done in this chapter: first, we have defined characteristic classes to be special invariants of vector bundles. Next, we have seen that invariant polynomials can be used to generate such classes (this was essentially theorem 4.4.2). Finally, we defined a special set of invariant polynomials, which turned out to generate the ring of polynomials (this was theorem 4.5.2). This showed us that any characteristic class which comes from an invariant polynomial is a polynomial in the Chern classes, making them important for further study.

What we would have liked to do now is work out an example which shows the use of the Chern classes. It would have been nice, for example, to show that the tangent bundle to the sphere is non-trivial. We already noted in section 3.3 that this follows from a theorem of algebraic topology; we could now prove it using characteristic classes. However, we would then be discussing a real vector bundle, instead of a complex one; thus, Chern classes would not apply.

Instead, we would need to introduce the Pontryagin classes of a real vector bundle, which are, loosely speaking, the Chern classes of its complexification. We decided not to do this: we felt that it would not fit in that well, adding significant volume while not offering any new viewpoint. (We do recommend to the reader to look up the concept of Pontryagin classes and work out the example for him or herself; it is a nice way to have some practice with the material from this and earlier chapters.) Moreover, we have been working on our bachelor project for quite a while now, and it is time to finish.

Finally, we would like to summarise what we have done. We started by introducing fibre bundles, which we specialised to vector bundles on the one hand and principal bundles on the other. We defined what a connection on these bundles is (with linear and invariant connections being structure preserving kinds, belonging to vector and principal bundles respectively), and how such a connection gives rise to parallel transport and curvature. To connect all this with physics we explained how bundles can be used to describe electromagnetism as a gauge theory. As an explicit example, we discussed the abelian Aharanov-Bohm effect quite thoroughly, while skimming over the more involved subject of Chern-Simons theory.

Going back to mathematics, we introduced the concept of characteristic classes, which can be used as an invariant of vector bundles. We also showed how characteristic classes can be created from invariant polynomials. We specialised these to the Chern classes, which were specific characteristic classes associated to some special polynomials.

Now, it is obvious that one can travel a lot further along the paths we set out on, far beyond the scope of this paper. We encourage the reader (and ourselves) to make this trip, for it is sure to be an interesting one.
Bibliography


