Introduction to Gauge Theory

Andriy Haydys

11th July 2019

This an incomplete DRAFT of lecture notes given at the PCMI Summer School "Quantum Field Theory and Manifold Invariants" (July 1 – July 5, 2019). If you spot any mistakes/typos please let me know at andriy[dot]haydys{at}math.uni-freiburg.de (please clearly point the place your refer to).

Contents

1 Introduction

Gauge theory by now is a vast subject with many connections in geometry, analysis, and physics. In these notes I focus on gauge theory as it is used in the construction of manifolds invariants, other uses of gauge theory remain beyond the scope of these notes.

The basic scheme of construction invariants of manifolds via gauge theory is quite simple. To be more concrete, let me describe some details. Thus, let $\mathcal{G}$ be a Lie group acting on a manifold $\mathcal{A}$. The common convention, which I will follow, is that $\mathcal{G}$ acts *on the right*, although this is clearly nonessential. The quotient $\mathcal{B} := \mathcal{A}/\mathcal{G}$ may fail to be a nice space[1] in a few ways, for example due to the presence of points with nontrivial stabilizers. Denote by $\mathcal{A}^* \subset \mathcal{A}$ the subspace consisting of all points with the trivial stabilizer. Then $\mathcal{G}$ acts on $\mathcal{A}^*$ and the quotient $\mathcal{B}^* := \mathcal{A}^*/\mathcal{G}$ is better behaved. Let me assume that $\mathcal{B}^*$ is in fact a manifold.

Pick a $\mathcal{G}$–representation V and a smooth $\mathcal{G}$–equivariant map $F\colon \mathcal{A} \to V$, where the equivariancy means the following:

$$F(a \cdot g) = g^{-1} \cdot F(a).$$

Here V is thought of as a left $\mathcal{G}$–module.

More to the point, assuming that $0 \in V$ is a regular value of F, we obtain a submanifold $\mathcal{M}^* := F^{-1}(0)/\mathcal{G} \subset \mathcal{B}^*$ of a dimension d, say. This will be referred to as the 'moduli space', although the terminology may seem odd at this moment. If $\mathcal{M}^*$ is compact and oriented, it has the fundamental class $[\mathcal{M}^*] \in H_d(\mathcal{B}^*; \mathbb{Z})$. In particular, for any cohomology class $\eta \in H^d(\mathcal{B}^*; \mathbb{Z})$ we obtain an integer

$$\big\langle [\mathcal{M}^*],\, \eta \big\rangle = \int_{\mathcal{M}^*} \eta,$$

where η is thought of as a closed form of degree d on $\mathcal{B}^*$. This is the 'invariant' we are interested in.

One way to construct cohomology classes on $\mathcal{B}^*$ is as follows. Assume there is a normal subgroup $\mathcal{G}_0 \subset \mathcal{G}$ so that $\mathrm{G} := \mathcal{G}/\mathcal{G}_0$ is a Lie group. Then the 'framed moduli space' $\hat{\mathcal{M}}^* := F^{-1}(0)/\mathcal{G}_0$ is equipped with an action of $\mathcal{G}/\mathcal{G}_0 = \mathrm{G}$ such that $\hat{\mathcal{M}}^*/\mathrm{G} = \mathcal{M}^*$. In other words, $\hat{\mathcal{M}}^*$ can be viewed as a principal G bundle, whose characteristic classes yield the cohomology classes we are after. More details on this is provided in Section 3.

To be continued.

2 Bundles and connections

2.1 Vector bundles

2.1.1 Basic notions

Roughly speaking, a vector bundle is just a family of vector spaces parametrized by points of a manifold (or, more generally, of a topological space).

More formally, the notion of a vector bundle is defined as follows.

Definition 1. Choose a non-negative integer k. A real smooth vector bundle of rank k is a triple (π, E, M) such that the following holds:

(i) E and M are smooth manifolds, $\pi\colon E \to M$ is a smooth submersion (the differential is surjective at each point);

(ii) For each $m \in M$ *the fiber* $E_m := \pi^{-1}(m)$ has the structure of a vector space and $E_m \cong \mathbb{R}^k$;

(iii) For each $m \in M$ there is a neighborhood $U \ni m$ and a smooth map ψ_U such that the following diagram Diagram

[1] Ideally, one wishes the quotient to be a smooth manifold.

commutes. Moreover, ψ_U is a fiberwise linear isomorphism.

The following terminology is commonly used: E is the total space, M is the base, π is the projection, and ψ_U is the local trivialization (over U).

Example 1.

(a) The product bundle: $M \times \mathbb{R}^k$;

(b) The tangent bundle TM of any smooth manifold M.

Let E and F be two vector bundles over a common base M. A *homomorphism* between E and F is a smooth map $\varphi\colon E \to F$ such that the diagram diagram

commutes and φ is a fiberwise linear map.

Two bundles E and F are said to be isomorphic, if there is a homomorphism φ, which is fiberwise an isomorphism.

A bundle E is said to be *trivial*, if E is isomorphic to the product bundle.

2.1.2 Operations on vector bundles

Let E and F be two vector bundles over a common base M. Then we can construct new bundles E^*, $\Lambda^p E$, $E \oplus F$, $E \otimes F$, and $\mathrm{Hom}(E, F)$ as follows:

($*$) $(E^*)_m = (E_m)^*$;

(Λ) $(\Lambda^p E)_m = \Lambda^p(E_m)$;

($\oplus$) $(E \oplus F)_m := E_m \oplus F_m$;

($\otimes$) $(E \otimes F)_m := E_m \otimes F_m$;

(Hom) $\mathrm{Hom}(E, F)_m := \mathrm{Hom}(E_m, F_m)$.

If $f\colon M' \to M$ is a smooth map, we can define the *pull-back* of $E \to M$ via

$$(f^*E)_{m'} := E_{f(m')}.$$

For example, if M' is an open subset of M and ι is the inclusion, then $E|_{M'} := \iota^* E$ is just the restriction of E to M'.

The reader should check that the families of vector spaces defined above satisfy the properties required by Definition 1.

Exercise 2. Prove that $E^* \otimes F$ is isomorphic to $\mathrm{Hom}(E, F)$.

Exercise 3. Prove that the tangent bundle of the 2-sphere is non-trivial. (Hint: Apply the hairy ball theorem).

2.1.3 Sections

Definition 4. A smooth map $s\colon M \to E$ is called a section, if $\pi \circ s = \mathrm{id}_M$.

In other words, a section assigns to each point $m \in M$ a vector $s(m) \in E_m$ such that $s(m)$ depends smoothly on m.

Sections of the tangent bundle TM are called vector fields. Sections of $\Lambda^p T^*M$ are called differential p-forms.

Example 2. If f is a smooth function on M, then the differential df is a 1-form on M. More generally, given functions $f_1, \dots, f_p$ we can also construct a p-form $\omega := df_1 \wedge \dots \wedge df_p$.

Exercise 5. Let $E \to M$ be a vector bundle of rank k and $U \subset M$ be an open subset. Prove that E is trivial over U if and only if there are k sections $e = (e_1, \dots, e_k)$, $e_j \in \Gamma(U;\, E)$, such that $e(m)$ is a basis of E_m for each $m \in U$. More precisely, given e show that ψ_U can be constructed according to the formula

$$\psi_U^{-1}\colon U \times \mathbb{R}^k \longrightarrow E|_U, \qquad (m, x) \mapsto e(m) \cdot x.$$

In fact this establishes a one-to-one correspondence between k-tuples of pointwise linearly independent sections and local trivializations of E.

We denote by $\Gamma(E) = \Gamma(M;\, E)$ the space of all smooth sections of E. Clearly, $\Gamma(E)$ is a vector space, where the addition and multiplication with a scalar are defined pointwise. In fact, $\Gamma(E)$ is a $C^\infty(M)$-module.

Given a local trivialization e over U (cf. Exercise 5) and a section s, we can write

$$s(m) = \sum_{j=1}^{k} \sigma_j(m) e_j(m)$$

for some functions $\sigma_j\colon U \to \mathbb{R}$. Thus, locally any section of a vector bundle can be thought of as a map $\sigma\colon U \to \mathbb{R}^k$.

It is important to notice that σ depends on the choice of a local trivialization. Indeed, if e' is another local trivialization of E over U', then there is a map

$$g\colon U \cap U' \longrightarrow \mathrm{GL}_n(\mathbb{R}) \quad \text{such that} \quad e = e' \cdot g. \tag{6}$$

If $\sigma'\colon U' \to \mathbb{R}^k$ is a local representation of s with respect to e', we have

$$s = e'\sigma' = eg^{-1}\sigma' = e\sigma \qquad \Longrightarrow \qquad \sigma' = g\sigma.$$

2.1.4 Covariant derivatives

The reader surely knows from the basic analysis course that the notion of the derivative is very useful. It is natural to ask whether there is a way to differentiate sections of bundles too.

To answer this question, recall the definition of the derivative of a function $f\colon M \to \mathbb{R}$. Namely, choose a smooth curve $\gamma\colon (-\varepsilon, \varepsilon) \to M$ and denote $m := \gamma(0)$, $\mathrm{v} := \dot\gamma(0) \in T_mM$. Then

$$df(\mathrm{v}) = \lim_{t \to 0} \frac{f(\gamma(t)) - f(m)}{t}. \tag{7}$$

Trying to replace f by a section s of a vector bundle, we immediately run into a problem, namely the difference $s(\gamma(t)) - s(m)$ is ill-defined in general since these two vectors may lie in different vector spaces.

Hence, instead of trying to mimic (7) we will define the derivatives of sections axiomatically, namely asking that the most basic property of the derivative—the Leibniz rule—holds.

Definition 8. Let $E \to M$ be a vector bundle. *A covariant derivative* is an $\mathbb{R}$-linear map $\nabla\colon \Gamma(E) \to \Gamma(T^*M \otimes E)$ such that

$$\nabla(fs) = df \otimes s + f\nabla s \tag{9}$$

holds for all $f \in C^\infty(M)$ and all $s \in \Gamma(E)$.

Example 10. Let $M \subset \mathbb{R}^N$ be an embedded submanifold. Then the tangent bundle TM is naturally a subbundle of the product bundle $\underline{\mathbb{R}}^N := M \times \mathbb{R}^N$. In particular, any section s of TM can be regarded as a map $M \to \mathbb{R}^N$. With this at hand we can define a connection on TM as follows

$$\nabla s := \mathrm{pr}(ds),$$

where pr is the orthogonal projection onto TM. A straightforward computation shows that this satisfies the Leibniz rule, i.e., ∇ is a connection indeed.

Theorem 11. *For any vector bundle $E \to M$ the space of all connections $\mathcal{A}(E)$ is an affine space modelled on $\Omega^1(\operatorname{End} E) = \Gamma\big(T^*M \otimes \operatorname{End}(E)\big)$.*

To be somewhat more concrete, the above theorem consists of the following statements:

(a) $\mathcal{A}(E)$ is non-empty.

(b) For any two connections ∇ and $\hat\nabla$ the difference $\nabla - \hat\nabla$ is a 1-form with values in $\operatorname{End}(E)$;

(c) For any $\nabla \in \mathcal{A}(E)$ and any $a \in \Omega^1(\operatorname{End} E)$ the following

$$(\nabla + a)s := \nabla s + as$$

is a connection.

For the proof of Theorem 11 we need the following elementary lemma, whose proof is left as an exercise.

Lemma 12. *Let $A\colon \Gamma(E) \to \Omega^p(F)$ be an $\mathbb{R}$-linear map, which is also $C^\infty(M)$-linear, i.e.,*

$$A(fs) = fA(s) \qquad \forall f \in C^\infty(M) \quad \textit{and} \quad \forall s \in \Gamma(E).$$

Then there exists $a \in \Omega^p\big(\operatorname{Hom}(E, F)\big)$ such that $A(s) = a \cdot s$. □

Proof of Theorem 11. Notice first that $\mathcal{A}(E)$ is convex, i.e., for any $\nabla, \hat\nabla \in \mathcal{A}(E)$ and any $t \in [0,1]$ the following $t\nabla + (1-t)\hat\nabla$ is also a connection.

If ψ_U is a local trivialization of E over U, then we can define a connection ∇_U on $E|_U$ by declaring

$$\nabla_U s := \psi_U^{-1}\, d(\psi_U(s)).$$

Using a partition of unity and the convexity property, a collection of these local covariant derivatives can be sewed into a global covariant derivative just like in the proof of the existence of Riemannian metrics on manifolds, cf. This proves (a). cite Thomas-Barden?

By (9), the difference $\nabla - \hat\nabla$ is $C^\infty(M)$–linear. Hence, (b) follows by Lemma 12.

The remaining step, namely (c), is straightforward. This finishes the proof of this theorem. □

While Theorem 11 answers the question of the existence of connections, the reader may wish to have a more direct way to put his hands on a connection. One way to do this is as follows.

Let e be a local trivialization. Since e is a pointwise base we can write

$$\nabla e = e \cdot A, \tag{13}$$

where $A = A(\nabla, e)$ is a $k \times k$-matrix, whose entries are 1-forms defined on U. A is called the connection matrix of ∇ with respect to e.

If σ is a local representation of a section s, then

$$\nabla s = \nabla(e\sigma) = \nabla(e)\sigma + e \otimes d\sigma = e\big(A\sigma + d\sigma\big).$$

Hence, it is common to say that locally

$$\nabla = d + A,$$

which means that $d\sigma + A\sigma$ is a local representation of ∇s. In particular, ∇ is uniquely determined by its connection matrix over U and any $A \in \Omega^1(U; \mathfrak{gl}_k(\mathbb{R}))$ appears as a connection matrix of some connection (cf. Theorem 11).

Just like σ, the connection matrix also depends on the choice of e. If $e' = eg$, we have

$$\nabla e' = \nabla(eg) = (\nabla e)g + e \otimes dg = e\big(Ag + dg\big) = e'\big(g^{-1}Ag + g^{-1}dg\big).$$

Hence, the connection matrix A' of ∇ with respect to e' can be expressed as follows:

$$A' = g^{-1}Ag + g^{-1}dg. \tag{14}$$

2.1.5 The curvature

While Definition 8 yields a way to differentiate sections, some properties very well known from multivariable analysis are *not* preserved. One of the most important cases is that covariant derivatives with respect to two variables do not need to commute. The failure of the commutativity of partial covariant derivatives is closely related to the notion of curvature, which is described next.

Denote $\Omega^p(E) := \Gamma\big(\Lambda^p T^*M \otimes E\big)$. We can extend the covariant derivative ∇ to a map $d_\nabla \colon \Omega^p(E) \to \Omega^{p+1}(E)$ as follows. If $s \in \Gamma(E)$ and $\omega \in \Omega^p(M)$ we declare

$$d_\nabla(\omega \otimes s) := d\omega \otimes s + (-1)^p \omega \wedge \nabla s,$$

which yields a unique $\mathbb{R}$-linear map d_∇ defined on all of $\Omega^p(E)$. This satisfies the following variant of the Leibniz rule

$$d_\nabla(\alpha \wedge \omega) = d\alpha \wedge \omega + (-1)^q \alpha \wedge d_\nabla \omega, \qquad \text{for all } \alpha \in \Omega^q(M) \text{ and } \omega \in \Omega^p(M;\, E).$$

Thus, we obtain a sequence

$$0 \to \Omega^0(E) \xrightarrow{d_\nabla = \nabla} \Omega^1(E) \xrightarrow{d_\nabla} \cdots \to \Omega^n(E) \to 0,$$

where $n = \dim M$. However, unlike in the case of the de Rham differential, the above does not need to be a complex.

Proposition 15. *There is a 2-form F_∇ with values in* $\operatorname{End} E$ *such that*

$$d_\nabla \circ d_\nabla = F_\nabla. \tag{16}$$

The above equality means that for any $\omega \in \Omega^p(E)$ we have $d_\nabla\big(d_\nabla(\omega)\big) = F_\nabla \wedge \omega$, where the right hand side of the letter equality is a combination of the wedge-product and the natural contraction $\operatorname{End}(E) \otimes E \to E$.

Proof of Proposition 15. We prove the proposition for $p = 0$ first. Applying the Leibniz rule twice, we see that $d_\nabla \circ \nabla \colon \Omega^0(E) \to \Omega^2(E)$ is $C^\infty(M)$–linear. Hence, by Lemma 12 we obtain a 2-form F_∇ such that

$$d_\nabla(\nabla s) = F_\nabla s$$

holds for any $s \in \Omega^0(E)$.

It remains to consider the case $p > 0$. For any $\eta \in \Omega^p(M)$ and $s \in \Gamma(E)$ we have

$$\begin{aligned} d_\nabla\big(d_\nabla(\eta \otimes s)\big) &= d_\nabla\big(d\eta \otimes s + (-1)^p \eta \wedge \nabla s\big) \\ &= 0 + (-1)^{p+1} d\eta \wedge \nabla s + (-1)^p d\eta \wedge \nabla s + (-1)^{2p} \eta \wedge d_\nabla\big(\nabla s\big) \\ &= \eta \wedge F_\nabla s \\ &= F_\nabla \wedge (\eta \otimes s). \end{aligned}$$

Here the last equality holds because F_∇ is a differential form of even degree. □

Definition 17. The 2-form F_∇ defined by (16) is called the curvature form of ∇.

Our next aim is to clarify somewhat the meaning of the curvature form. If v is a tangent vector at some point $m \in M$, we call

$$\nabla_{\mathrm{v}}\, s := \imath_{\mathrm{v}} \nabla s$$

the covariant derivative of s in the direction of v.

Choose local coordinates $(x_1, \dots, x_n)$ and denote by $\partial_i = \frac{\partial}{\partial x_i}$ the tangent vector of the curve

$$\gamma(t) = (x_1^0, \dots, x_{i-1}^0,\ t,\ x_{i+1}^0, \dots, x_n^0).$$

We may call

$$\nabla_i\, s := \nabla_{\partial_i}\, s$$

"partial covariant derivatives". With these notations at hand we have the expression

$$\nabla s = \sum_{i=1}^n dx_i \otimes \nabla_i\, s,$$

which is just a form of the familiar expression for the differential of a function: $df = \sum \frac{\partial f}{\partial x_i} dx_i$.

Furthermore, we have

$$\begin{aligned} d_\nabla(\nabla s) &= -\sum_{i=1}^n dx_i \wedge \nabla\big(\nabla_i\, s\big) = -\sum_{i,j=1}^n dx_i \wedge dx_j \otimes \nabla_j(\nabla_i s) \\ &= \sum_{i,j=1}^n dx_i \wedge dx_j \otimes \nabla_i(\nabla_j s). \end{aligned}$$

From this we conclude

$$F_\nabla\big(\partial_i,\,\partial_j\big) = \nabla_i(\nabla_j s) - \nabla_j(\nabla_i s),$$

i.e., the curvature form measures the failure of partial covariant derivatives to commute as mentioned at the beginning of this section.

Sometimes it is useful to have an expression of the curvature form in a local frame. Thus, pick a local frame e of E and let A be the connection form of ∇ with respect to e. If σ is a local representation of some $s \in \Gamma(E)$, then we have

$$\begin{aligned} d_\nabla\big(\nabla(e\sigma)\big) &= d_\nabla\big(e(d\sigma + A\sigma)\big) = \nabla e \wedge (d\sigma + A\sigma) + e(dA\,\sigma - A\wedge d\sigma) \\ &= e\Big(A\wedge d\sigma + A\wedge A\,\sigma + dA\,\sigma - A\wedge d\sigma\Big) \\ &= e\big(dA + A\wedge A\big)\,\sigma. \end{aligned}$$

Hence, we conclude that locally

$$F_\nabla = dA + A\wedge A. \tag{18}$$

In particular, the curvature form is a first order non-linear operator in terms of the connection form.

Remark 19. It turns out to be useful to think of A as a 1–form with values in the Lie algebra $\mathfrak{gl}_k(\mathbb{R}) = \mathrm{End}(\mathbb{R}^k)$. From this perspective it is more suitable to write (18) in the following form

$$F_\nabla = dA + \frac{1}{2}[A\wedge A], \tag{20}$$

where the last term is a combination of the wedge product and the Lie brackets.

If e' is another local frame such that $e = e'\cdot g$, then using (14) one can show that

$$F'_\nabla = dA' + A'\wedge A' = g^{-1}F_\nabla\, g.$$

The reader is strongly encouraged to check the details of this computation.

Local expression (18) implies immediately the following.

Proposition 21. *If $a \in \Omega^1(\mathrm{End}\,E)$, then the curvature forms of ∇ and $\nabla + a$ are related by the equality:*

$$F_{\nabla+a} = F_\nabla + d_\nabla a + a\wedge a. \qquad \square$$

2.1.6 The gauge group

Pick a vector bundle E and consider

$$\mathcal{G} = \mathcal{G}(E) := \Big\{ g\in\Gamma\big(\mathrm{End}(E)\big) \mid \forall m\in M\ g(m)\in \mathrm{GL}(E_m)\Big\},$$

which is endowed with the C^∞–topology. If E is endowed with an extra structure, for example an orientation or a scalar product, we also require that gauge transformations respect this structure.

Clearly, $\mathcal{G}$ is a topological group, where the group operations are defined pointwise. $\mathcal{G}$ is called the group of gauge transformations of E or simply *the gauge group*.

If ∇ is a connection on E and $g\in\mathcal{G}$, we can define another connection as follows

$$\nabla^g s := g^{-1}\nabla(gs).$$

This yields a right action of $\mathcal{G}$ on $\mathcal{A}(E)$.

Definition 22. Two connections are called gauge equivalent, if there is a gauge transformation that transforms one of the connections into the other one.

Insert: local connection matrix and gauge transformations.

2.2 Principal bundles

The computations we met in the previous sections on the dependence of our objects on the choice of a local frame become formidable quite soon. The notion of a principal bundle is useful in dealing with this and turns out to have other advantages as we will see below. The idea is to consider all possible frames at once rather than choosing local trivializations when needed.

2.2.1 The frame bundle and the structure group

Let G be a Lie group.

Definition 23. A principal bundle with the structure group G is a triple (P, M, π), where

(i) P and M are smooth manifolds and $\pi\colon P \to M$ is a surjective submersion;

(ii) G acts on P on the right such that $\pi(p \cdot g) = \pi(p)$ for all $p \in P$ and all $g \in G$;

(iii) G acts freely and transitively on each fiber $\pi^{-1}(m)$;

(iv) For each $m \in M$ there is a neighborhood $U \ni m$ and a map ψ_U such that the diagram Diagram

commutes. Moreover, ψ_U is G–equivariant $\psi_U(p \cdot g) = \psi_U(p) \cdot g$, where G acts on $U \times G$ by the multiplication on the right on the second factor.

A fundamental example of a principal bundle is the frame bundle of a vector bundle $E \to M$. We take a moment to describe the construction in some detail.

Thus, for a fixed $m \in M$ let $\mathrm{Fr}(E_m)$ denote the set of all bases of E_m. The group $\mathrm{GL}_k(\mathbb{R})$ acts freely and transitively on $\mathrm{Fr}(E_m)$ so that we can in fact identify $\mathrm{Fr}(E_m)$ with $\mathrm{GL}_k(\mathbb{R})$ even though in a non-canonical way. In any case, $\mathrm{Fr}(E_m)$ can be viewed as an open subset of $\mathbb{R}^{k^2}$.

Consider
$$\mathrm{Fr}(E) := \bigsqcup_{m\in M} \mathrm{Fr}(E_m).$$
Clearly, there is a well-defined projection $\pi\colon \mathrm{Fr}(E) \to M$ determined uniquely by the property: $\pi(p) = m$ if and only if $p \in \mathrm{Fr}(E_m)$.

We introduce a smooth structure on $\mathrm{Fr}(E)$ as follows. Pick a chart U on M. By shrinking U if necessary we can assume that there is a local frame e of E defined on U. Then we have the bijective map
$$\Psi_U\colon U \times \mathrm{GL}_k(\mathbb{R}) \longrightarrow \pi^{-1}(U), \qquad (m, h) \mapsto e(m) \cdot h. \tag{24}$$
Using this we can think of $\pi^{-1}(U)$ as an open subset of an Euclidean space so that we can declare $\pi^{-1}(U)$ to be a chart on $\mathrm{Fr}(E)$ with an obvious choice of coordinates.

Let U' be another chart on M such that there is a local frame e' defined on U'. A straightforward computation yields

$$\Psi_{U'}^{-1} \circ \Psi_U(m, h) = \big(m,\, g(m)h\big),$$

where g is defined by (6). This implies that the transition maps between $\pi^{-1}(U)$ and $\pi^{-1}(U')$ are smooth, i.e., we have constructed a smooth atlas on $\mathrm{Fr}(U)$. The rest of the properties required in the definition of the principal bundle are clear from the construction.

Exercise 25. Show that local triviality of a principal bundle, i.e., Property (iv) of Definition 23, is equivalent to the existence of local sections. More precisely, if P admits a trivialization over U, then there is a section of $P|_U$ and conversely, if $P|_U$ admits a section, then $P|_U$ is also trivializable over U. In particular, show that the frame bundle of TS^2 does not admit any global sections.

Often vector bundles come equipped with an extra structure, for example orientations of each fiber and/or scalar product on each fiber. In the language of principal bundles this corresponds to the notion of a G-structure.

Definition 26. Let G be a Lie subgroup of $\mathrm{GL}_k(\mathbb{R})$. A G-structure on E is a G–subbundle P of the frame bundle. In this case G is called the structure group of E.

To illustrate this notion, let us consider the following example. Assume E is an Euclidean vector bundle, which means that each fiber E_m is equipped with an Eucliden scalar product $\langle \cdot, \cdot \rangle_m$, which depends smoothly on m. Here the dependence is said to be smooth if for any two smooth sections s_1 and s_2 the function $\langle s_1, s_2 \rangle$ is also smooth.

It is natural to consider the subset

$$\mathrm{O}(E) := \big\{ e \in \mathrm{Fr}(E) \mid e \text{ is orthonormal} \big\}.$$

The restriction of π yields a surjective map $\mathrm{O}(E) \to M$, which is still denoted by π. If e is any local frame of E over an open subset U, the Gram-Schmidt orthogonalization process shows that there is also a smooth pointwise orthonormal frame e_O defined on U. Just like in the case of $\mathrm{Fr}(E)$ we can cover $\mathrm{O}(E)$ by open subsets $\pi^{-1}(U)$ such that

$$\Psi_U \colon U \times \mathrm{O}(k) \to \pi^{-1}(U), \qquad (m, h) \mapsto e_O(m) \cdot h$$

is a bijection. While $\mathrm{O}(k)$ is not an open subset of an Euclidean space, it is a manifold, and therefore we can cover $\mathrm{O}(k)$—hence, also $\pi^{-1}(U)$—by a collection of charts. The same argument as in the case of the frame bundle shows that the transition functions are smooth so that $\mathrm{O}(E)$ is a principal $\mathrm{O}(k)$–bundle.

We see that an Euclidean structure on E determines an $\mathrm{O}(k)$–structure on E. Conversely, an $\mathrm{O}(k)$-structure $P \subset \mathrm{Fr}(E)$ determines an Euclidean structure on E. Indeed, pick any $p \in P_m$ and any two vectors $\mathrm{v}_1,\, \mathrm{v}_2 \in E_m$. Since p is a basis of E_m, we can write $\mathrm{v}_i = p \cdot \mathrm{x}_i$, where $\mathrm{x}_i \in \mathbb{R}^k$. We define a scalar product on E_m by

$$\langle \mathrm{v}_1, \mathrm{v}_2 \rangle = \mathrm{x}_1^t \mathrm{x}_2.$$

It is straightforward to check that this does not depend on the choice of p. Moreover, the scalar product defined in this way depends smoothly on m.

To summarize, an Euclidean structure on a vector bundle is equivalent to an $\mathrm{O}(k)$–structure.

Exercise 27. A fiberweise volume form is by definition a nowhere vanishing section of $\Lambda^k E^*$, where $k = \operatorname{rk} E$. Show that there is a one-to-one correspondence between fiberwise volume forms and $\mathrm{SL}_k(\mathbb{R})$–structures.

Let V be a complex vector space of complex dimension k. One can view V as a real vector space of dimension $2k$ equipped with an endomorphism $I \in \operatorname{End}_{\mathbb{R}}(V),\ I\mathrm{v} = i\mathrm{v}$ so that $I^2 = -\mathrm{id}$. Conversely, given a real vector space equipped with an endomorphism $I \in \operatorname{End}_{\mathbb{R}}(V)$ such that $I^2 = -\mathrm{id}$ we can regard V as a complex vector space, where $i \cdot \mathrm{v} := I\mathrm{v}$. In this case $\dim_{\mathbb{R}} V$ is necessarily even. The map I is called a complex structure.

With the above understood, a complex vector bundle is just a real vector bundle equipped with $I \in \Gamma\big(\operatorname{End} E\big)$ such that $I^2 = -\mathrm{id}$. In particular, each fiber E_m is endowed with a complex structure $I(m)$. Thus, a complex vecor bundle is essentially a family of complex vector spaces parametrized by points of the base.

Exercise 28.

(i) Show that a complex vector bundle can be defined as a locally trivial family of complex vector spaces akin to Definition 1;

(ii) Show that there is a one-to-one correspondence between complex structures on a real vector bundle E or rank $2k$ and $\mathrm{GL}_k(\mathbb{C})$–structures. Here $\mathrm{GL}_k(\mathbb{C})$ is viewed as a subgroup of $\mathrm{GL}_{2k}(\mathbb{R})$

$$\mathrm{GL}_k(\mathbb{C}) = \big\{ A \in \mathrm{GL}_{2k}(\mathbb{R}) \mid A \circ I_{st} = I_{st} \circ A \,\big\},$$

where I_{st} is the standard complex structure on $\mathbb{R}^{2k}$:

$$I_{st}(x_1, y_1, \dots, x_k, y_k) := (-y_1, x_1, \dots, -y_k, x_k).$$

Exercise 29. A Hermitian structure on a complex vector bundle is a smooth family of Hermitian scalar products on each fiber. Prove that the following holds:

(i) Any complex vector bundle admits a Hermitian structure;

(ii) There is a one-to-one correspondence between Hermitian structures and $\mathrm{U}(k)$–structures.

2.2.2 The associated vector bundle

Let $\pi\colon P \to M$ be a principal G–bundle. For any representation $\rho\colon G \to \mathrm{GL}(V)$, where V is a vector space, we can construct a vector bundle over M as follows.

Define the right G–action on the product bundle $P \times V$ via

$$(p, \mathrm{v}) \cdot g = \big(p \cdot g,\ \rho(g^{-1})\mathrm{v}\big).$$

Clearly, this action is free and properly discontinuous so that the quotient

$$P \times_\rho V := \big(P \times V\big)/G \tag{30}$$

is a smooth manifold. The map π yields a well-defined projection

$$P \times_\rho V \to M, \qquad [p, \mathrm{v}] \mapsto \pi(p)$$

which we still denote by the same letter π. Its fibers are isomorphic to V, hence these have a canonical structure of vector spaces. Moreover, given a local section $s \in \Gamma(U;\, P)$ we can construct a local trivialization of E via the map

$$\psi_U^{-1}\colon U \times V \to \pi_E^{-1}(U), \qquad (m, \mathrm{v}) \mapsto [s(m),\, \mathrm{v}].$$

Hence, if P is locally trivial, so is E.

Definition 31. The vector bundle $E = E(P, \rho, V)$ defined by (30) is called the vector bundle associated with (P, ρ), or simply *the associated bundle*.

Example 32. Let $P = \mathrm{Fr}(E)$, $V = \mathbb{R}^k$, and $\rho = \mathrm{id}$ be the tautological representation of $\mathrm{GL}_k(\mathbb{R})$. Then the map

$$\mathrm{Fr}(E) \times \mathbb{R}^k \longrightarrow E, \qquad (e, x) \mapsto \sum_{i=1}^n x_i e_i$$

induces an isomorphism $E\big(\mathrm{Fr}(E), \mathrm{id}\big) \cong E$.

This example shows that the construction of an associated bundle allows one to recover a vector bundle from its frame bundle. However, by varying the representation ρ we can obtain other bundles as well. The following example illustrates this.

Example 33. Consider $P = \mathrm{Fr}(E)$ again, however this time we take $V = \Lambda^p(\mathbb{R}^k)^*$ and ρ the natural representation of $\mathrm{GL}_k(\mathbb{R})$ on $\Lambda^p(\mathbb{R}^k)^*$, i.e.,

$$\rho\colon \mathrm{GL}_k(\mathbb{R}) \times \Lambda^p(\mathbb{R}^k)^* \to \Lambda^p(\mathbb{R}^k)^*, \qquad (g, \alpha) \mapsto \alpha\big(g^{-1}\cdot\,, \dots, g^{-1}\cdot\,\big). \tag{34}$$

We have the map

$$\mathrm{Fr}(E) \times \Lambda^p(\mathbb{R}^k)^* \to \Lambda^p E^*, \qquad (e, \alpha) \mapsto \alpha\big(e^{-1}\cdot\,, \dots, e^{-1}\cdot\,\big)$$

where we think of a frame e as an isomorphism $\mathbb{R}^k \to E_{\pi(e)}$. This induces an isomorphism of vector bundles

$$\mathrm{Fr}(E) \times_\rho \Lambda^p(\mathbb{R}^k)^* \cong \Lambda^p E^*.$$

Thus, $\Lambda^p E^*$ can be recovered as the vector bundle associated with $(\mathrm{Fr}(E), \rho)$, where ρ is given by (34).

Exercise 35. Let V be the space of $k \times k$-matrices $M_k(\mathbb{R}) \cong \mathrm{End}(\mathbb{R}^k)$ viewed as a $\mathrm{GL}_k(\mathbb{R})$–representation as follows:

$$\rho\colon (g, A) \mapsto gAg^{-1}.$$

Show that $\mathrm{Fr}(E) \times_\rho \mathrm{End}(\mathbb{R}^k)$ is isomorphic to $\mathrm{End}(E)$.

Denote

$$C^\infty(P;\, V)^G := \big\{\hat{s}\colon P \to V \mid s(p\cdot g) = \rho(g^{-1})\hat{s}(p) \quad \forall p \in P \text{ and } g \in G\,\big\}.$$

Pick any $\hat{s} \in C^\infty(P;\, V)^G$ and denote by $\varpi\colon P \times V \to P \times_\rho V$ the natural projection. Consider the map

$$\varpi \circ (\mathrm{id}, \hat{s})\colon P \longrightarrow P \times_\rho V, \qquad p \mapsto [p,\, \hat{s}(p)].$$

The letter map is G-invariant, where G acts trivially on the target space. Hence, there is a unique map $s\colon M \to P\times_\rho V$ such that

$$s\circ\pi = \varpi\circ(\mathrm{id},\hat s). \tag{36}$$

In other words, s is defined by requiring that the diagram

Diagram!

commutes.

Proposition 37. *The map*

$$C^\infty(P;\,V)^G \to \Gamma\big(P\times_\rho V\big), \qquad \hat s\mapsto s,$$

where s is defined by (36), is a bijection.

Proof. We only need to construct the inverse map. Thus, let s be a section of $P\times_\rho V$. For any $p\in P$ there is a unique $\hat s(p)\in V$ such that

$$s\big(\pi(p)\big) = [p,\ \hat s(p)].$$

It is straightforward to check that $\hat s$ is equivariant. □

It will be useful below to have a description of differential forms with values in an associated bundle in the spirit of Proposition 37. Before stating the claim, we need a few notions.

For $g\in G$ denote $R_g\colon P\to P,\ R_g(p) = p\cdot g$. The infinitesimal action of the Lie algebra $\mathfrak g$ is given by the vector field

$$K_\xi(p) := \frac{d}{dt}\Big|_{t=0}\big(p\cdot\exp(t\xi)\big), \qquad \xi\in\mathfrak g.$$

Since G acts freely on the fibers, we have

$$\mathcal V_p := \ker\pi_*|_p = \{K_\xi(p)\mid \xi\in\mathfrak g\}\cong\mathfrak g.$$

$\mathcal V_p$ is called *the vertical subspace*.

Definition 38.

- A q–form ω with values in some G-representation V is said to be *G-equivariant*, if $R_g^*\omega = \rho(g^{-1})\omega$;
- A q–form ω with values in some G-representation V is said to be *basic*, if $\omega(\mathrm v_1,\dots,\mathrm v_q) = 0$ whenever one of the arguments belongs to the vertical subspace.

Denote by $\Omega^q_{bas}(P;\,V)^G$ the space of all basic and G-equivariant q-forms on P with values in V.

Proposition 39. *For any $a\in\Omega^q\big(M;\,P\times_\rho V\big)$ the pull-back π^*a can be viewed as an element of $\Omega^q_{bas}(P;\,V)^G$. Moreover, the map $a\mapsto\pi^*a$ establishes a bijective correspondence between $\Omega^q\big(M;\,P\times_\rho V\big)$ and $\Omega^q_{bas}(P;\,V)^G$.*

Proof. For any tangent vectors $\hat{\mathrm{v}}_1, \dots, \hat{\mathrm{v}}_q$ to P the equality

$$a\big(\pi_*\hat{\mathrm{v}}_1, \dots, \pi_*\hat{\mathrm{v}}_q\big) = [p,\ \hat{a}_p(\hat{\mathrm{v}}_1, \dots, \hat{\mathrm{v}}_q)]$$

determines uniquely $\hat{a} \in \Omega^q(P; V)$. Since by definition the vertical space is $\ker \pi_*$, the fact that $\pi^* a$ is basic is clear. The equivariancy of $\hat{a}$ follows from a straighforward computation.

Conversely, let $\hat{a}$ be given. Pick any $m \in M$ and any $p \in \pi^{-1}(m)$. Pick also any $\mathrm{v}_1, \dots, \mathrm{v}_q \in T_m M$ and choose $\hat{\mathrm{v}}_1, \dots, \hat{\mathrm{v}}_q \in T_p P$ such that $\pi_*(\hat{\mathrm{v}}_j) = \mathrm{v}_j$ for all $j \in \{1, \dots, q\}$. These lifts do exist because π_* is surjective, however these need not be unique. With this at hand, we can define a by the equality

$$a_m\big(\mathrm{v}_1, \dots, \mathrm{v}_q\big) = [p,\ \hat{a}_p(\hat{\mathrm{v}}_1, \dots, \hat{\mathrm{v}}_q)].$$

Since $\hat{a}$ is basic and equivariant, a does not depend on the choices involved. □

2.2.3 Connections on principal bundles

The Lie algebra $\mathfrak{g}$ can be viewed as a G-representation, where the action is the adjoint one: $\mathrm{ad}\colon G \to \mathrm{GL}(\mathfrak{g})$. For example, if G is a subgroup of $\mathrm{GL}_k(\mathbb{R})$, then

$$\mathrm{ad}_g\, \xi = g\, \xi\, g^{-1}.$$

Definition 40. *A connection form*, or simply *a connection*, on a principal G-bundle P is a G–equivariant 1–form a with values in the Lie algebra $\mathfrak{g}$ such that

$$a(K_\xi) = \xi \qquad \forall \xi \in \mathfrak{g}.$$

Denote

$$\mathrm{ad}\, P = P \times_{\mathrm{ad}} \mathfrak{g}.$$

Theorem 41. *For any principal bundle the space of all connections $\mathcal{A}(P)$ is an affine space modelled on $\Omega^1(\mathrm{ad}\, P)$.*

This theorem can be proved in the same manner as Theorem 11. Instead of going through the details, we describe the only essential modification of the argument used in the proof of Theorem 11. Namely, given $a,\ a' \in \mathcal{A}(P)$ the difference $b := a - a'$ is basic and G-invariant so that by applying Proposition 39 one can think of b as a 1–form on M with values in $\mathrm{ad}\, P$.

Notice that Theorem 41 states in particular, that $\mathcal{A}(P)$ is non-empty.

Example 42. $\mathrm{U}(1)$ acts freely on

$$S^{2n+1} := \big\{ z \in \mathbb{C}^{n+1} \mid |z_0|^2 + \dots + |z_n|^2 = 1 \big\}$$

in the diagonal manner. This represents S^{2n+1} as the total space of the $\mathrm{U}(1)$-bundle

$$\pi\colon S^{2n+1} \to \mathbb{C}P^n, \qquad z \mapsto [z], \tag{43}$$

which is sometimes called the *Hopf fibration*.

The infinitesimal action of $U(1)$ on S^{2n+1} is given by the vector field

$$v(z) = \big(iz_0, \dots, iz_n\big).$$

There is a unique connection $a \in \Omega^1\big(S^{2n+1}; \mathbb{R}i\big)$ such that $\ker a = v^\perp$. Explicitly,

$$a_z(u) = \langle v(z), u\rangle\, i, \qquad \text{where } u \in T_z S^{2n+1}.$$

Since the action of $\mathrm{U}(1)$ preserves the Riemannian metric of the sphere, a is an invariant 1-form. It remains to notice that for abelian groups the notions of equivariant and invariant forms coincide. Thus, a is a connection 1-form on (43).

Exercise 44. Prove that the associated bundle $S^{2n+1} \times_{\mathrm{U}(1)} \mathbb{C}$ is (canonically) isomorphic to *the tautological line bundle:*

$$\mathcal{O}(-1) := \Big\{\big([z], w\big) \in \mathbb{C}P^n \times \mathbb{C}^{n+1} \mid w \in [z] \cup \{0\}\Big\}.$$

The definitions of a connection on a vector and principal bundles differ significantly and the reader may wonder what is the relation between these two notions. The following result yields an answer to this question.

First notice that by differentiating $\rho\colon G \to \mathrm{GL}(V)$ we obtain an action of $\mathfrak{g}$ on V, i.e., a Lie algebra homomorphism

$$\rho_*|_{\mathbb{1}} \colon \mathfrak{g} \to \mathrm{End}(V). \tag{45}$$

This way one can think of a connection a as a 1-form with values in $\mathrm{End}(V)$ provided a representation V is given.

Theorem 46.

(i) Let E be a vector bundle. Any connection ∇ on E determines a unique connection a on the frame bundle such that

$$e^* a = A(\nabla, e).$$

Here we think of a local frame e as a local section of $\mathrm{Fr}(E)$ and $A(\nabla, e)$ is the local connection 1–form of ∇ with respect to e.

(ii) Let P be a principal G–bundle. Any connection a on P induces a unique connection ∇ on any associated vector bundle $P \times_\rho V$ such that

$$\pi^*\nabla s = d\hat{s} + a \cdot \hat{s}. \tag{47}$$

Here the right hand side is a basic and G-invariant 1–form on P and the equality is understood in the sense of Proposition 39.

Proof. The proof consists of a number of the following steps.

Step 1. *Denote*

$$\hat{R}\colon P \times G \longrightarrow P, \qquad \hat{R}(p, g) = p \cdot g.$$

Then the differential $\hat{R}_*$ at (p, g) satisfies:

$$\hat{R}_*(\mathrm{v}, \mathrm{w}) = (R_g)_*\mathrm{v} + K_{(L_{g^{-1}})_*\mathrm{w}}(p \cdot g),$$

where $L_{g^{-1}} \colon G \to G,\ h \mapsto g^{-1}h$, $\mathrm{v} \in T_pP$, and $\mathrm{w} \in T_gG$.

The proof of this step is a simple computation, which is left as an exercise.

Step 2. *Let e be a local frame of E over U. For any 1-form $A \in \Omega^1(U;\, \mathfrak{gl}_k(\mathbb{R}))$ there is a unique connection a on $\mathrm{Fr}(E)|_U$ such that*

$$e^*a = A.$$

Since e is a section of $\mathrm{Fr}(E)|_U$, we have $\pi_* \circ e_* = \mathrm{id}_U$ and therefore e_* is injective. Hence, by the dimensional reasons, we have $T_{e(m)}P = \mathcal{V}_{e(m)} \oplus \operatorname{Im} e(m)_*$. Therefore, we can define

$$a_{e(m)}\big(K_\xi + e_*\mathrm{v}\big) = \xi + A(\mathrm{v}).$$

This extends to a unique G-invariant 1-form on $\mathrm{Fr}(E)|_U$. The equality

$$(R_g)_* K_\xi = K_{\mathrm{ad}_{g^{-1}}\xi},$$

which can be checked by a straightforward computation, implies that a is a connection.

Step 3. *We prove (i).*

Choose two local frames e and e', which we may assume to be defined on the same open set U (restrict to the intersection of the corresponding domains if necessary). The equalities

$$a_{e(m)}\big(e_*\cdot\,\big) = A_m(\cdot) \qquad \text{and} \qquad a'_{e'(m)}\big(e'_*\cdot\,\big) = A'_m(\cdot)$$

determine unique $\mathrm{GL}_k(\mathbb{R})$–invariant 1-forms a and a' on $\mathrm{Fr}(E)|_U$ by Step 2.

We have

$$\begin{aligned} e'^*a(\cdot) &= a_{e'}\big(e'_*\,\cdot\,\big) = a_{e\cdot g}\Big(\big(\hat R \circ (e,g)\big)_*\,\cdot\,\Big) = a_{e\cdot g}\Big(\big(R_g\big)_* \circ e_*\,\cdot\,\Big) + a_{e\cdot g}\big(K_{g^{-1}dg}(e\cdot g)\big) \\ &= g^{-1}a_e(e_*\cdot)g + g^{-1}dg = A'(\cdot). \end{aligned}$$

Here the second equality follows by Step 1 and the last one by (14). Thus, we conclude that $a = a'$ on the intersection of the domains of local frames e and e'. Thus, a is globally well-defined.

Step 4. *We prove (ii).*

By the equivariancy of $\hat s\colon P \to V$ we have

$$\big(d\hat s + a\cdot \hat s\big)(K_\xi) = -\xi\cdot\hat s + a(K_\xi)\cdot\hat s = 0,$$

i.e., $d\hat s + a\cdot\hat s$ is a basic 1-form, which is G-invariant as well. Hence, there is a 1-form ∇s on M such that (47) holds. Moreover, for any G-invariant function f on P we have the equality

$$d(f\cdot\hat s) + a\cdot(f\cdot\hat s) = df\otimes\hat s + f\big(d\hat s + a\cdot\hat s\big),$$

which shows that ∇ is a connection. This finishes the proof of this theorem. □

One immediate corollary of this theorem is that a connection on E induces a connection on E^*, $\mathrm{End}(E)$ ans so on. Indeed, these bundles can be interpreted as associated bundles for $P = \mathrm{Fr}(E)$ and the corresponding $\mathrm{GL}_k(\mathbb{R})$–representations. A more direct characterization of the induced connections is given in the following.

Exercise 48. Let ∇ be a connection on E. Show that the following holds.

(i) There is a unique connection on E^*, still denoted by ∇, such that

$$d\langle \alpha, s\rangle = \langle \nabla\alpha, s\rangle + \langle \alpha, \nabla s\rangle \qquad \forall \alpha\in\Gamma(E^*) \quad \text{and} \quad \forall s\in\Gamma(E),$$

where $\langle\cdot,\cdot\rangle$ denotes the natural pairing $E^*\otimes E\to\mathbb{R}$. This connection coincides with the induced one.

(ii) There is a unique connection on $\mathrm{End}(E)$, still denoted by ∇, such that

$$\nabla\big(\varphi(s)\big) = (\nabla\varphi)(s) + \varphi(\nabla s) \qquad \forall \varphi\in\Gamma(\mathrm{End}(E)) \quad \text{and} \quad \forall s\in\Gamma(E),$$

This connection coincides with the induced one.

2.2.4 The curvature of a connection on a principal bundle

Local expression (20) of the curvature form suggests the following construction. Let a be a connection on a principal G–bundle P. The 2-form $da + \frac{1}{2}[a\wedge a]$ with values in $\mathfrak{g}$ is clearly G-invariant. This form is also basic as the following computation shows:

$$\begin{aligned}
\imath_{K_\xi}\Big(da + \frac{1}{2}[a\wedge a]\Big) &= \mathcal{L}_{K_\xi}a - d\big(\imath_{K_\xi}a\big) + \frac{1}{2}\big[a(K_\xi), a(\cdot)\big] - \frac{1}{2}\big[a(\cdot), a(K_\xi)\big]\\
&= -[\xi, a] + 0 + \frac{1}{2}[\xi, a] + \frac{1}{2}[\xi, a]\\
&= 0.
\end{aligned}$$

Here $\mathcal{L}_K$ denotes the Lie derivative with respect to a vector field K and the first equality uses Cartan's magic formula

$$\mathcal{L}_K = d\,\imath_K + \imath_K d.$$

Hence, by Proposition 39 we obtain that there is some $F_a\in\Omega^2(M;\mathrm{ad}\ P)$ such that

$$\pi^*F_a = da + \frac{1}{2}[a\wedge a]. \tag{49}$$

Definition 50. The 2–form F_a defined by (49) is called *the curvature form* of a.

Example 51. Let us compute the curvature of the connection a that was constructed in Example 42 in the simplest case $n=1$. On the 3-sphere we have the following vector fields

$$v_1 := (-x_1, x_0, -x_3, x_2), \quad v_2 := (-x_2, x_3, x_0, -x_1), \quad \text{and} \quad v_3 := (-x_3, -x_2, x_1, x_0),$$

which at each point yield an orthonormal oriented basis of the tangent space to the sphere. It is worthwhile to notice that v_1 coincides with the vector field v from Example 42.

By the definition of the connection form a, we have

$$a = \big(-x_1\,dx_0 + x_0\,dx_1 - x_3\,dx_2 + x_2\,dx_3\big)\,i.$$

Hence,

$$\pi^*F_a = da = 2\big(dx_0\wedge dx_1 + dx_2\wedge dx_3\big)\,i.$$

In particular, we have

$$F_a(\pi_*v_2, \pi_*v_3) = \pi^*F_a(v_2, v_3) = 2\,i.$$

It can be shown that the quotient metric on $S^3/\mathrm{U}(1)$ yields the round metric on the sphere of radius $1/2$. Explicitly, the corresponding isometry is given by

$$(z_0, z_1) \mapsto \left(z_0\bar{z}_1, \tfrac{1}{2}(|z_0|^2 - |z_1|^2)\right).$$

Here we think of S^3 as a subset of $\mathbb{C}^2$.

Since $(\pi_* v_2, \pi_* v_3)$ is an oriented orthonormal basis of the tangent space of $S^2_{1/2}$, we conclude that $F_a = 2\,\mathrm{vol}_{S^2_{1/2}}\, i$, where $\mathrm{vol}_{S^2_{1/2}}$ is the volume form of the standard round metric on $S^2_{1/2}$.

Notice that

$$\int_{S^2} F_a = 2\,\mathrm{Vol}(S^2_{1/2})\, i = 2\pi\, i.$$

The proof of the following proposition is left as an excercise.

Proposition 52. *Let a be a connection on a principal bundle P and let V be a G–representation. Using (45) we can think of F_a as a 2-form with values in $\mathrm{End}(P \times_\rho V)$. With these identifications in mind, the curvature of the induced connection ∇ on $P \times_\rho V$ equals F_a.* □

Given a local trivialization of P, i.e., a section σ over an open subset $U \subset M$, we say that $A := \sigma^* a \in \Omega^1(U; \mathfrak{g})$ is a local representation of a with respect to σ. Then over U we have

$$F_a = \sigma^*\pi^* F_a = \sigma^*\left(da + \frac{1}{2}[a \wedge a]\right) = dA + \frac{1}{2}[A \wedge A].$$

One of the most basic properties of the curvature form is the so called Bianchi identity.

Proposition 53 (Bianchi identity)**.** *Let a be a connection on P. Then the curvature form F_∇ satisfies*

$$d_{\nabla^a} F_a = 0,$$

where ∇^a is the connection on $\mathrm{ad}\, P$ induced by a.

Proof. Let A be a local representation of a as above. We have

$$\begin{aligned} d_{\nabla^a} F_a &= d\Big(dA + \frac{1}{2}[A \wedge A]\Big) + \Big[A \wedge \big(dA + \frac{1}{2}[A \wedge A]\big)\Big] \\ &= \frac{1}{2}[dA \wedge A] - \frac{1}{2}[A \wedge dA] + [A \wedge dA] + \frac{1}{2}\big[A \wedge [A \wedge A]\big] \\ &= 0. \end{aligned}$$

Here the first three summands sum up to zero because $[\omega, \eta] = -[\eta, \omega]$ for any $\omega \in \Omega^2(U; \mathfrak{g})$ and $\eta \in \Omega^1(U;\, \mathfrak{g})$; The vanishing of the last summand follows from the Jacobi identity. □

Remark 54. Thinking of $\hat{F}_A := \pi^* F_A = dA + \frac{1}{2}[A \wedge A]$ also as a curvature form of A, the computation in the proof of Proposition 53 yields the following equivalent form of the Bianchi identity:

$$d\hat{F}_A = [\hat{F}_A \wedge A]. \tag{55}$$

Indeed, this can be obtained by the computation: $d\hat{F}_A = [dA \wedge A] = [\hat{F}_A \wedge A]$.

Given a bundle $P \to M$ and a map $f\colon N \to M$ we can construct the pull-back bundle $f^*P \to N$ just like in the case of vector bundles. Informally speaking, we have the equality of fibers $(f^*P)_n = P_{f(n)}$ and this defines f^*P. More formally, consider

$$f^*P := \big\{(p, n) \in P \times N \mid f(n) = \pi(p)\big\}.$$

It is easy to see that f^*P is a submanifold of $P \times N$. Moreover, we have a projection $\varpi\colon f^*P \to N$, which is just the restriction of the natural projection $(p, n) \mapsto n$. The structure group G clearly acts on P such that the action on the fibers $\varpi^{-1}(n) = P_{f(n)}$ is transitive. Moreover, we also have a natural G-equivariant map $\hat{f}\colon f^*P \to P$, which covers f, i.e., the diagram

Insert the diagram!

commutes.

For future use we note the following statement, whose proof is left as an exercise.

Proposition 56. *Let A be a connection on P. Then $f^*A := \hat{f}^*A$ is a connection on f^*P and $F_{f^*A} = f^*F_A$.* □

2.2.5 The gauge group

2.3 The Levi–Civita connection

2.4 Classification of $\mathrm{U}(1)$ and $\mathrm{SU}(2)$ bundles

It is more convenient in this section to work in a topological category rather than the smooth one. The reader will have no difficulties to adopt the corresponding notions to this setting.

Definition 57. Let G be a compact Lie group. A topological space E equipped with an action of G is said to be *a classifying bundle* for G, if E is contractible and the G–action is free.

Denoting $B := E/G$, we obtain a natural projection $E \to B$ so that E could be thought of as a principal[2] G–bundle over B. If E exists, it is easy to see that E is unique up to a homotopy equivalence.

One can prove that for a compact Lie group a classifying space always exist. An interested reader may wish to consult [GS99, Sect. 1.2]. Also, we take the following result as granted.

Theorem 58 ([GS99, Thm. 1.1.1 and Rem. 2])**.** *Let $P \to M$ be a (topological) principal G–bundle over a manifold M. Then there exists a continuous map $f\colon M \to B$ such that P is isomorphic to f^*E. In fact f is unique up to a homotopy so that the map*

$$f \mapsto f^*E$$

yields a bijective correspondence between the set of isomorphism classes of principal G-bundles over M and the set $[M;\, B]$ of homotopy classes of maps $M \to B$. □

[2]The notations E and B are traditional in this context so I will keep to this tradition. The reader should not be confused by the fact that here E denotes a principal rather than vector bunlde.

In some sense, the above theorem yields a classification of principal—hence, also vector—bundles. In praxis it is not always easy to describe the set $[M;\, B]$ though. One way to deal with this problem is via the so called characteristic classes, which we describe next.

Definition 59. Let $f\colon M \to B$ be a continuous map. Pick any $c \in H^\bullet(B;\, R)$, where R is a ring. Then $f^*c \in H^\bullet(M;\, R)$ is called a characteristic class of f^*E.

It is worth pointing out that characteristic classes depend on the isomorphism class of the bundle only.

Usually, characteristic classes are easier to deal with than the set of homotopy classes of maps. The most common choices for the ring R are $\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$.

In some cases the classifying bundle can be constructed fairly explicitly. I restrict myself to the following two cases, namely $G = \mathrm{U}(1)$ and $G = \mathrm{SU}(2)$, which are most commonly used in gauge theoretic problems.

2.4.1 Complex line bundles

It follows from Exercise 29 that the classification problems for complex line bundles and principal $\mathrm{U}(1)$-bundles are equivalent. Even though what follows below can be described in terms of vector bundles only, the language of $\mathrm{U}(1)$–bundles has certain advantages and will be mainly used below.

Consider the following commutative diagram

where the horizontal arrows are natural inclusions. For example, the inclusion of the spheres is given by

$$\mathbb{C}^n \supset S^{2n-1} \ni \quad (z_0, \dots, z_{n-1}) \mapsto (z_0, \dots, z_{n-1}, 0) \quad \in S^{2n+1} \subset \mathbb{C}^{n+1},$$

which is a $\mathrm{U}(1)$–equivariant map.

The direct limit construction yields a CW-complex S^∞ equipped with a free $\mathrm{U}(1)$-action. S^∞ can be shown to be contractible. Furthermore, we have also a CW-complex $\mathbb{C}P^\infty$ so that the natural quotient map

$$S^\infty \to \mathbb{C}P^\infty$$

is the classifying bundle for the group $\mathrm{U}(1)$.

Example 60 (Classification of line bundles on 2–manifolds)**.** Let M be an oriented two–manifold. A continuous map $f\colon M \to \mathbb{C}P^\infty$ is homotopic to a map, which takes values in the 2-skeleton $\mathbb{C}P^1 \subset \mathbb{C}P^\infty$ so that we have the equality $[M;\, \mathbb{C}P^\infty] = [M;\, \mathbb{C}P^1]$. Topologically, $\mathbb{C}P^1$ is just the 2-sphere so that we have a well–defined degree-map

$$[M;\, S^2] \to \mathbb{Z}, \qquad [f] \mapsto \deg f,$$

which is in fact a bijection. Thus, a complex line bundle L on an oriented two–manifold is classified by an integer d, which is called the degree of L.

It is easy to see that the cohomology ring $H^\bullet\big(\mathbb{C}P^\infty;\,\mathbb{Z}\big)$ is generated by a single element $a \in H^2\big(\mathbb{C}P^\infty;\,\mathbb{Z}\big)$. We can fix the choice of the generator by requiring

$$\langle a, [\mathbb{C}P^1]\rangle = 1, \tag{61}$$

where $\langle\cdot,\cdot\rangle$ is the pairing between the homology and cohomology groups.

With this understood, to any principal $\mathrm{U}(1)$-bundle $P \to M$ (equivalently, to any complex line bundle $L \to M$) we can associate a cohomology class as follows. If $f\colon M \to \mathbb{C}P^\infty$ is a map such that P is the pull-back of the bundle $S^\infty \to \mathbb{C}P^\infty$, then

$$c_1(P) := -f^*a \in H^2(M;\,\mathbb{Z}). \tag{62}$$

The minus sign in this definition is a convention.

Definition 63. The class $c_1(P)$ defined by (62) is called *the first Chern class* of P.

If $L \to M$ is a complex line bundle, we can choose a Hermitian scalar product, or, in other words, we can choose a $\mathrm{U}(1)$-structure $P \subset \mathrm{Fr}(L)$. Then

$$c_1(L) := c_1(P)$$

is also called the first Chern class of L.

Exercise 64. Check that the first Chern class of L does not depend on the choice of the Hermitian scalar product on L.

Theorem 65. *The first Chern class of the complex line bundle has the following properties:*

(i) $c_1(\underline{\mathbb{C}}) = 0$*, where* $\underline{\mathbb{C}}$ *is the product bundle;*

(ii) $c_1(L_1 \otimes L_2) = c_1(L_1) + c_1(L_2)$ *for all line bundles* L_1 *and* L_2 *over the same base* M*;*

(iii) $c_1(L^\vee) = -c_1(L)$*, where* $L^\vee := \mathrm{Hom}(L;\mathbb{C})$ *is the dual line bundle;*

(iv) $c_1(f^*L) = f^*c_1(L)$ *for all line bundles* $L \to M$ *and all (continuous) maps* $f\colon N \to M$*.* □

2.4.2 Quaternionic line bundles

Let $\mathbb{H}$ denote the $\mathbb{R}$-algebra of quaternions. One can think of the (compact) symplectic group

$$\mathrm{Sp}(1) := \big\{q \in \mathbb{H} \mid |q|^2 = q\bar{q} = 1\big\}$$

as a quaternionic analogue of $\mathrm{U}(1)$. It is easy to see that $\mathrm{Sp}(1)$ is isomorphic to $\mathrm{SU}(2)$. Indeed, it is easy to write down an isomorphism explicitly:

$$\mathrm{Sp}(1) \to \mathrm{SU}(2), \qquad q = z + wj \mapsto \begin{pmatrix} z & w \\ -\bar{w} & \bar{z}\end{pmatrix}.$$

Just like $\mathrm{U}(1)$ acts freely on S^{2n+1}, $\mathrm{Sp}(1)$ acts freely on $S^{4n+3} = \big\{(h_0,\cdots,h_n) \in \mathbb{H}^n \mid |h_0|^2 + \cdots + |h_n|^2 = 1\big\}$ in the diagonal manner so that we have a natural principal $\mathrm{Sp}(1)$-bundle:

$$S^{4n+3} \to \mathbb{HP}^n.$$

The commutative diagram diagram!

leads to the construction of the classifying bundle for $\mathrm{Sp}(1)$:

$$S^\infty \to \mathbb{HP}^\infty.$$

Notice that $\mathbb{HP}^\infty$ is (an infinite-dimensional) CW-complex, which has exactly one cell of dimension $4n$. This implies in particular that the cohomology ring $H^\bullet(\mathbb{HP}^\infty; \mathbb{Z})$ is generated by a single generator b of degree 4. The choice of b can be fixed for example by requiring

$$\langle b, [\mathbb{HP}^1] \rangle = 1.$$

Proposition 66. *Let M be a manifold of dimension at most 3. Any $\mathrm{Sp}(1)$-bundle over M is trivial.*

Proof. Given a principal $\mathrm{Sp}(1)$-bundle P, by Theorem 58 we can find a map $f\colon M \to \mathbb{HP}^\infty$ such that f^*S^∞ is isomorphic to P. Furthermore, f is homotopic to a map f_1 that maps into the 3-skeleton of $\mathbb{HP}^\infty$, which is a point. Hence, P is trivial. $\square$

Let M be a manifold of arbitrary dimension. To any map $f\colon M \to \mathbb{HP}^\infty$ we can associate a class $f^*b \in H^4(M; \mathbb{Z})$, where b is a generator of $H^\bullet(\mathbb{HP}^\infty; \mathbb{Z})$ as above.

Definition 67. Let $P \to M$ be a principal $\mathrm{Sp}(1)$–bundle. If $f\colon M \to \mathbb{HP}^\infty$ is a map such that $f^*S^\infty \cong P$, then sign?

$$c_2(P) := -f^*b \in H^4(M; \mathbb{Z})$$

is called *the second Chern class* of P.

Remark 68. The terminology may seem to be somewhat strange at this point. The reason is that for a principal $\mathrm{U}(r)$–bundle P one can define r characteristic classes $c_1(P), c_2(P), \dots, c_r(P)$, where $c_j(P) \in H^{2j}(M; \mathbb{Z})$. In the particular cases described above, this yields the constructions of the first and the second Chern classes.

3 The Chern–Weil theory

3.1 The Chern–Weil theory

In this section we could equally well work with both $\mathbb{R}$ and $\mathbb{C}$ as ground fields. I opt for $\mathbb{C}$ mainly for the sake of definiteness. The modifications needed for the case of $\mathbb{R}$ as a ground field are straightforward.

Let $p\colon \mathfrak{g} \to \mathbb{C}$ be an ad–invariant homogeneous polynomial of degree d. This means the following:

- Given a basis $\xi_1, \dots, \xi_n$ of $\mathfrak{g}$, the expression $p\big(x_1\xi_1 + \dots + x_n\xi_n\big)$ is a polynomial of degree d in $x_1, \dots, x_n$;
- $p(ad_g\, \xi) = p(\xi)$ for all $g \in G$ and all $\xi \in \mathfrak{g}$;
- $p(\lambda\xi) = \lambda^d p(\xi)$ for all $\lambda \in \mathbb{R}$ and all $\xi \in \mathfrak{g}$.

Example 69.

(a) For $\mathfrak{g} = \mathfrak{u}(r)$ the map $p_d(\xi) = i\,\mathrm{tr}\,\xi^d$ is an ad-invariant polynomial of degree d.

(b) Choose $\mathfrak{g} = \mathfrak{u}(r)$ again and define polynomials $c_1, \dots, c_r$ of degrees $1, \dots, r$ respectively by the equality

$$\det\Bigl(\lambda \mathbb{1} + \frac{i}{2\pi}\xi\Bigr) = \lambda^r + c_1(\xi)\lambda^{r-1} + \dots + c_r(\xi).$$

For example, $c_r(\xi) = \frac{i^r}{(2\pi)^r}\det\xi$ and $c_1(\xi) = \frac{i}{2\pi}\,\mathrm{tr}\,\xi$. Notice also that the equality

$$\overline{\det\Bigl(\lambda \mathbb{1} + \frac{i}{2\pi}\xi\Bigr)} = \det\Bigl(\bar\lambda \mathbb{1} + \frac{i}{2\pi}\xi\Bigr)$$

implies that each c_j takes values in $\mathbb{R}$.

Let $P \to M$ be a principal G–bundle equipped with a connection $a \in \Omega^1(P; \mathfrak{g})$. Think of the curvature form $\pi^*F_a = da + \frac{1}{2}[a \wedge a]$ as a matrix, whose entries are 2-forms on P. Since forms of even degrees commute, the expression $p(\pi^*F_a)$ makes sense as an $\mathbb{R}$-valued differential form of degree at most $2d$ on P. Since each entry of π^*F_a is basic, so is $p(\pi^*F_a)$. Moreover, the ad–invariancy of p implies that $p(\pi^*F_a)$ is G–invariant. By Proposition 39 applied in the case of the trivial G–representation we obtain that there is a form $p(F_a)$ on M of degree $2d$ such that

$$\pi^* p(F_a) = p(\pi^*F_a).$$

Lemma 70. *The following holds:*

(i) $p(F_a)$ is closed;

(ii) The de Rham cohomology class of $p(F_a)$ does not depend on the choice of connection a.

Proof. The proof consists of the following steps.

Step 1. *We prove (i).*

Pick any $\xi, \xi_1, \dots, \xi_d \in \mathfrak{g}$. Thinking of $\mathfrak{g}$ as a matrix Lie algebra, I write temporarily $ad_{e^{t\xi}}\xi_j = e^{t\xi}\xi_j e^{-t\xi}$. Slightly abusing notations, denote by $p\colon Sym^d(\mathfrak{g}) \to \mathbb{R}$ the d-multilinear function whose restriction to the diagonal yields the original polynomial p. Then, differentiating the equality

$$p\bigl(e^{t\xi}\xi_1 e^{-t\xi}, \dots, e^{t\xi}\xi_d e^{-t\xi}\bigr) = p(\xi_1, \dots, \xi_d)$$

with respect to t, yields

$$p\bigl([\xi,\xi_1], \xi_2, \dots, \xi_d\bigr) + p\bigl(\xi_1, [\xi,\xi_2], \dots, \xi_d\bigr) + \dots + p\bigl(\xi_1, \xi_2, \dots, [\xi,\xi_d]\bigr) = 0. \tag{71}$$

Denote $\hat F_A := \pi^* F_A$. Then (71) implies

$$p\bigl([\hat F_A \wedge A], \hat F_A, \dots, \hat F_A\bigr) + p\bigl(\hat F_A, [\hat F_A \wedge A], \dots, \hat F_A\bigr) + \dots + p\bigl(\hat F_A, \hat F_A, \dots, [\hat F_A \wedge A]\bigr) = 0.$$

Hence,

$$\begin{aligned} d\bigl(p(\hat F_A)\bigr) &= p(d\hat F_A, \hat F_A, \dots, \hat F_A) + p(\hat F_A, d\hat F_A, \dots, \hat F_A) + \dots + p(\hat F_A, \hat F_A, \dots, d\hat F_A) \\ &= p\bigl([\hat F_A \wedge A], \hat F_A, \dots, \hat F_A\bigr) + p\bigl(\hat F_A, [\hat F_A \wedge A], \dots, \hat F_A\bigr) + \dots + p\bigl(\hat F_A, \hat F_A, \dots, [\hat F_A \wedge A]\bigr) \\ &= 0. \end{aligned}$$

Here the second equality uses (55). This finishes the proof of *(i)*.

Step 2. *Let $I = [0,1]$ be the interval and $\imath_0, \imath_1 \colon M \to M \times I$ be the natural inclusions correspodnding to the endpoints of the interval. There exist linear maps $Q\colon \Omega^k(M \times I) \to \Omega^{k-1}(M)$ such that for any $\omega \in \Omega^k(M \times I)$ we have*

$$\imath_1^*\omega - \imath_0^*\omega = d\,Q\omega - Q\,d\omega.$$

Provide a reference.

Step 3. *We prove (ii).*

Pick any two connection A_0 and A_1 and think of $A_t := (1-t)A_0 + tA_1$ as a connection on $\varpi^*P \to M \times I$, where $\varpi\colon M \times I \to M$ is the natural projection. Then

$$p(F_{A_1}) - p(F_{A_0}) = \imath_1^* p(F_{A_t}) - \imath_0^* p(F_{A_t}) = d\,Qp(F_{A_t})$$

by the previous step. This proves *(ii)*. □

3.1.1 The Chern classes

Let c_j be the polynomial of degree j from Example 69.

Let $P \to M$ be a principal $\mathrm{U}(r)$–bundle with a connection A. By Lemma 70, $c_j(F_A)$ is closed, real valued, and the de Rham cohomology class of $c_j(F_A)$ does not depend on the choice of the connection.

Definition 72. The class $c_j(P) := [c_j(F_A)] \in H^{2j}_{dR}(M;\mathbb{R})$ is said to be the jth Chern class of P and

$$c(P) := 1 + c_1(P) + \dots + c_r(P) \in H^\bullet(M;\mathbb{R})$$

is called the total Chern class of P.

Remark 73. The above definition yields Chern classes as elements of the de Rham cohomology groups only. In fact, one can show that they lie in the image of $H^\bullet(M;\mathbb{Z}) \to H^\bullet_{dR}(M;\mathbb{R})$. I will discuss this briefly in the case of the first two Chern classes below. Also, at this point we have two a priori unrelated definitions of the first two Chern classes. We will see below that in fact they agree.

Done?

Remark 74. Let E be a complex vector bundle of rank r. Choosing a fiberwise Hermitian structure on E, we obtain a principal $\mathrm{U}(r)$–bundle Fr_U so that we can define $c_j(E) := c_j(\mathrm{Fr}_\mathrm{U})$. It is easy to show that this does not depend on the choice of the Hermitian structure.

Theorem 75. *The Chern classes satisfy the following properties*

(i) $c_0(E) = 1$ *for any vector bundle E;*

(ii) $c(f^*E) = f^*c(E)$ *for all vector bundles $E \to M$ and all maps $f\colon N \to M$;*

(iii) $c(E_1 \oplus E_2) = c(E_1) \cup c(E_2)$*;*

(iv) $c\big(\mathcal{O}(-1)\big) = 1 - a$*, where $\mathcal{O}(-1)$ is the tautological line bundle over $\mathbb{P}^1$ and a is the generator of the cohomology group of $\mathbb{P}^1$ such that (61) holds.*

The first property above is jut the definition, the remaining properties are called naturality, Whitney sum formula, and normalization respectively.

Proof. The naturality follows immediately from Proposition 56. The normalization is equivalent to $\frac{i}{2\pi}\int_{\mathbb{P}^1} F_a = -1$, where a is a unitary connection on the tautological line bundle. This was established in Example 51.

Thus, we only need to prove the Whitney sum formula. If ∇_1 and ∇_2 are unitary connections on E_1 and E_2 respectively, then the curvature of the corresponding connection on the Whitney sum is a block diagonal matrix. More precisely, this means that the curvature is a 2-form with values in $\mathrm{End}(E_1) \oplus \mathrm{End}(E_2)$. If A and B are any square matrices, we have

$$\det\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \det A \, \det B \qquad \Longrightarrow \qquad \begin{aligned} c(F_{\nabla_1 \oplus \nabla_2}) &= \det\begin{pmatrix} \mathbb{1} + \frac{i}{2\pi} F_{\nabla_1} & 0 \\ 0 & \mathbb{1} + \frac{i}{2\pi} F_{\nabla_2} \end{pmatrix} \\ &= c(F_{\nabla_1}) \wedge c(F_{\nabla_2}). \end{aligned}$$

The latter equality clearly implies the Whitney sum formula. $\square$

Exercise 76. Let E be a vector bundle. Prove that the following holds:

(a) The Chern classes depend on the isomorphism class of E only;

(b) $c_j(E^\vee) = (-1)^j c_j(E)$ for all j;

(c) If E is trivial, then $c(E) = 1$;

(d) If $E \cong E_1 \oplus \underline{\mathbb{C}}^k$, then $c_j(E) = 0$ for $j > \operatorname{rk} E - k$.

Exercise 77. Show that the tangent bundle of S^2 is non-trivial.

Theorem 78. *Let L be a complex line bundle. Then the first Chern class in the sense of Definition 72 coincides with the image in $H^2_{dR}(M;\mathbb{R})$ of the first Chern class in the sense of Definition 63.*

Sketch of proof. Let $L \to M$ be a complex line bundle. It is not too hard to show that there is $N < \infty$ and a smooth map $f\colon M \to \mathbb{P}^N$ such that $f^*\mathcal{O}(-1)$ is isomorphic to L, where $\mathcal{O}(-1)$ is the tautological bundle of $\mathbb{P}^N$. Notice that $H^2_{dR}(\mathbb{P}^N;\mathbb{R})$ is one dimensional and generated by the class Poincaré dual to $[\mathbb{P}^1]$, where $\imath\colon \mathbb{P}^1 \subset \mathbb{P}^N$ is a standard embedding.

Pick a Hermitian structure on $\mathcal{O}(-1)$ and a Hermitian connection ∇. Then $\imath^*\mathcal{O}(-1)$ is the tautological bundle of $\mathbb{P}^1$, so that $\frac{i}{2\pi}\imath^* F_\nabla$ represents the first Chern class of $\mathcal{O}_{\mathbb{P}^1}(-1)$. Hence, by Example 51 we have

$$\left\langle \left[\tfrac{i}{2\pi}\imath^* F_\nabla\right],\ [\mathbb{P}^1] \right\rangle = -1$$

Hence, $c_1(\mathcal{O}(-1)) = -a$ so that $c_1(L) = c_1(f^*L) = -f^*a$. $\square$

Remark 79. One can prove arguing along similar lines that in the case of $\mathrm{SU}(2)$–bundles the two definitions of the second Chern class agree on the level of the de Rham cohomology groups. I leave the details to the readers. Moreover, one can also show that the infinite Grassmannian $\mathrm{Gr}_k(\mathbb{C}^\infty)$ is a classifying space for the group $\mathrm{U}(k)$. Thus one could also define the Chern classes as pull-backs of certain classes on $\mathrm{Gr}_k(\mathbb{C}^\infty)$.

Remark 80. One corollary of Definition 63 is follows. Let $L \to M$ be a Hermitian line bundle and $h\colon \Sigma \to M$ a smooth map, where Σ is a compact oriented two–manifold. Then we have

$$\frac{i}{2\pi}\int_M h^* F_\nabla \in \mathbb{Z},$$

where ∇ is any Hermitian line bundle. This property may be quite surprising if one's starting point is Definition 72.

In particular, if M is itself a compact oriented two-dimensional manifold (and h is the identity map), then

$$\frac{i}{2\pi}\int_M F_A$$

is an integer, which coincides with the degree of L, cf. Example 60.

Remark 81. A straightforward computation yields that for any matrix $\xi \in \mathfrak{su}(2)$ we have $\operatorname{tr}\xi^2 = -2\det\xi$. Hence, for an $\mathrm{SU}(2)$–bundle P we have

$$c_2(P) = \frac{1}{8\pi^2}\big[\operatorname{tr}(F_A \wedge F_A)\big] \in H^4_{dR}(M;\mathbb{R}).$$

In particular, if M is a closed oriented four-manifold, the integration yields an isomorphism $H^4_{dR}(M;\,\mathbb{R}) \cong \mathbb{R}$. In fact, just as in the case of line bundles above, we have

$$c_2(P) = \frac{1}{8\pi^2}\int_M \operatorname{tr}(F_A \wedge F_A) \in \mathbb{Z}. \tag{82}$$

3.2 The Chern–Simons functional

In this section I will restrict myself to dimension three and $G = \mathrm{SU}(2)$. Thus, let M be a three manifold equipped with an $\mathrm{SU}(2)$–bundle $P \to M$. Notice that P is trivial as we have seen in Proposition 66.

As a matter of fact, any closed oriented three-manifold is a boundary of a compact oriented four-manifold, say $\partial X = M$. Assume there is an extension of P to X, i.e., a bundle P_X such that $P_X|_M = P$. In this case any connection A on P can be extended to a connection A_X on P_X so that we can form the integral

$$\frac{1}{8\pi^2}\int_X \operatorname{tr}\big(F_{A_X}\wedge F_{A_X}\big).$$

If we take any other extension (X', P'_X, A'_X), we can glue X and X' along their common boundary to form a four-manifold without boundary. Strictly speaking, when performing the gluing we have to change the orientation of X' so that $\partial X'$ is equipped with the orientation opposite to that of M to have the resulting manifold oriented. This together with (82) yields, that the difference

$$\frac{1}{8\pi^2}\int_X \operatorname{tr}\big(F_{A_X}\wedge F_{A_X}\big) - \frac{1}{8\pi^2}\int_{X'} \operatorname{tr}\big(F_{A'_X}\wedge F_{A'_X}\big)$$

is an integer. Hence,

$$\vartheta(A) := \frac{1}{8\pi^2}\int_X \operatorname{tr}\big(F_{A_X}\wedge F_{A_X}\big) \tag{83}$$

is well-defined as a function with values in $\mathbb{R}/\mathbb{Z}$. This is called *the Chern–Simons functional*.

While this definition makes transparent the relation of the Chern–Simons functional with the Chern–Weil theory, it is possible to compute the value of the Chern–Simons functional directly without extending A to a four-manifold. In fact, choosing a trivialization of P we can think of A as a 1-form on M with values in $\mathfrak{su}(2)$. Then

$$\vartheta(A) = \frac{1}{8\pi^2}\int_M \operatorname{tr}\big(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\big).$$

Notice that this expression does not yield an $\mathbb{R}$–valued function. The reason is that by changing the trivialization of P the value of ϑ changes by an integer so that we obtain again a map to the circle.

Exercise 84.

(a) For $A \in \Omega^1(X)$, where X is a four-manifold, prove the equality

$$d\ \operatorname{tr}\Big(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\Big) = \operatorname{tr}\big(F_A\wedge F_A\big). \tag{85}$$

Notice the following: In the special case $X = M\times\mathbb{R}$ denote by A_t the pull-back of A to $M\times\{t\}$. Then (85) clearly implies

$$\vartheta(A_t) - \vartheta(A_{t_0}) = \int_{M\times[t_0,t]} \operatorname{tr}\big(F_A\wedge F_A\big).$$

(b) Prove that the two definition of the Chern–Simons functional agree.

(c) Prove that the values of the Chern–Simons functional with respect to two different trivializations differ by an integer.

(d) Let g be a gauge transformation, which can be though of as a map $M\to \mathrm{SU}(2)\cong S^3$. Show that

$$\vartheta(A\cdot g) = \vartheta(A) + \deg g.$$

Let us compute the differential of ϑ. For $a\in\Omega^1(M;\mathfrak{su}(2))$ we have

$$d\vartheta_A(a) = \frac{1}{8\pi^2}\int_M \operatorname{tr}\big(a\wedge dA + A\wedge da + 2a\wedge A\wedge A\big) = \frac{1}{4\pi^2}\int_M \operatorname{tr}\big(F_A\wedge a\big).$$

In the second equality the integration by parts is used. Hence, we conclude the following.

Proposition 86. *The critical points of the Chern–Simons functional are flat connections, i.e., connections A such that $F_A = 0$.* □

3.3 The modui space of flat connections

Even though in Section 3.2 I opted to work with three-manifolds and $G = \mathrm{SU}(2)$, the notion of a flat connection clearly makes sense for any background manifold and any structure group. Thus, we do not need to impose these restrictions in this section.

Let $P \to M$ be a principal G–bundle. Denote by $\mathcal{A}^\flat(P)$ the space of all flat connections on P. The gauge group $\mathcal{G}(P)$ acts on $\mathcal{A}^\flat(P)$ so that we can form *the moduli space of flat connections:*

$$\mathcal{M}^\flat(P) := \mathcal{A}^\flat(P)/\mathcal{G}(P).$$

Such moduli spaces are typical objects in gauge theory that we will meet many times below. The main questions we are interested in are the following: Is $\mathcal{M}^\flat(P)$ compact? Is $\mathcal{M}^\flat(P)$ a manifold?

An important point to notice is that $\mathcal{M}^\flat$ represents the space of all solutions of a non-linear PDE modulo an equivalence relation, so that in essence the question is to describe topological properties of the space of all solutions of a non-liner PDE. In general, this may be a hard question, however, in this particular case we will see below that this can be done with a little technology involved.

However, why could one be potentially interested in spaces like $\mathcal{M}^\flat$? The two main reasons are as follows: First, sometimes $\mathcal{M}^\flat$ encodes a subtle information about the background manifold M (as well as the bundle P) and thus can be used for instance as a tool in studies of the topology of M; Secondly, moduli spaces come often equipped with an extra structure, which may be of interest on its own. In these notes I will mainly emphasize the first point, while the second one will be only briefly mentioned below.

3.3.1 Parallel transport and holonomy

Let $E \to M$ be a vector bundle equipped with a connection ∇. For any (smooth) curve $\gamma\colon [0,1] \to M$, $\gamma^*\nabla$ is a connection on γ^*E. A section $s \in \Gamma(\gamma^*E)$ is said to be *parallel along* γ, if $(\gamma^*\nabla)(s) = 0$.

Remark 87. If γ is a simple embedded curve, then s can be thought of as a section of E defined along the image of γ.

Since any bundle over an interval is trivial, we can trivialize $\gamma^*E \cong \mathbb{R}^k \times [0,1]$ so that $\gamma^*\nabla$ can be written as $\frac{d}{dt} + B(t)\,dt$, where $B\colon [0,1] \to M_k(\mathbb{R})$ is a map with values in the space of $k \times k$-matrices. Thinking of s as a map $[0,1] \to \mathbb{R}^k$, we obtain that s is parallel along γ if and only if s is a solution of the equations:

$$\dot{s} + A(t)s(t) = 0.$$

By the main theorem of ordinary differential equations, the above equation has a unique solution for any initial value s_0 and this solution is defined on the whole interval $[0,1]$.

Definition 88. If $s \in \Gamma(\gamma^*E)$ is parallel along γ, then $s(1) \in E_{\gamma(1)}$ is called the parallel transport of $s(0) = s_0 \in E_{\gamma(0)}$ with respect to ∇.

The above consideration shows in fact that for any connection ∇ any curve γ we have a linear isomorphism

$$\mathrm{PT}_\gamma\colon E_{\gamma(0)} \to E_{\gamma(1)},$$

which is called *the parallel transport.*

In a special case, namely when γ is a loop, the parallel transport yields an isomorphism of the fiber. If we concatenate two loops, the parallel transport is the composition of the parallel transports corresponding to the initial loops. Hence, for a fixed connection the set of all parallel transports is in fact a group.

Definition 89. Pick a point $m \in M$. The group

$$\mathrm{Hol}_m(\nabla) := \left\{ \mathrm{PT}_\gamma \in \mathrm{GL}(E_m) \mid \gamma \text{ is a loop based at } m \right\}$$

is called the holonomy group of ∇ based at m.

Choosing a basis of E_m, we can think of $\mathrm{Hol}_m(\nabla)$ as a subgroup of $\mathrm{GL}_k(\mathbb{R})$. A standard argument shows that if m and m' lie in the same connected component, then the holonomy groups $\mathrm{Hol}_m(\nabla)$ and $\mathrm{Hol}_{m'}(\nabla)$ are conjugate, i.e., there is $A \in \mathrm{GL}_k(\mathbb{R})$ such that $\mathrm{Hol}_{m'}(\nabla) = A\,\mathrm{Hol}_m(\nabla)A^{-1}$. With this understood, we can drop the basepoint from the notation. Even though $\mathrm{Hol}(\nabla)$ is defined up to a conjugacy only, it is still commonly referred to as a subgroup of $\mathrm{GL}_k(\mathbb{R})$.

Exercise 90. Show that the following holds:

- ∇ is Euclidean $\implies$ PT_γ is orthogonal $\implies$ $\mathrm{Hol}(\nabla) \subset O(k)$.
- ∇ is complex $\implies$ PT_γ is complex linear $\implies$ $\mathrm{Hol}(\nabla) \subset \mathrm{GL}_{k/2}(\mathbb{C})$.
- ∇ is complex Hermitian $\implies$ PT_γ is unitary $\implies$ $\mathrm{Hol}(\nabla) \subset \mathrm{U}(k/2)$.

Remark 91. The concept of the parallel transport also makes sense for connections on principal bundles. The construction does not differ substantially from the case of vector bundles. The details are left to the readers.

3.3.2 The monodromy representation of a flat connection

Let ∇ be a flat connection on a vector bundle E of rank k. We can view the parallel transport as a map

$$\gamma \mapsto \mathrm{Hol}(\nabla; \gamma),$$

where γ is a loop based at some fixed point m. If A is flat, this map depends on the homotopy class of γ only so that effectively we obtain a representation of the fundamental group: $\rho_A \colon \pi_1(M) \to \mathrm{GL}_k(\mathbb{R})$. Ref?

Exercise 92. Show that a gauge-equivalent connection yields a conjugate representation.

Conversely, given a representation $\rho\colon \tilde{M} \to \mathrm{GL}_k(\mathbb{R})$ we can construct the bundle

$$E := \tilde{M} \times_{\pi_1(M),\,\rho} \mathbb{R}^k.$$

Here $\pi_1(M)$ acts on $\tilde{M}$ by the deck transformations. This means that $\tilde{M}$ can be viewed as a principal $\pi_1(M)$–bundle so that E is the associated bundle corresponding to the representation ρ.

This bundle is equipped with a natural flat connection. Indeed, interpreting a section s of E as a $\pi_1(M)$–equivariant map $\hat{s}\colon \tilde{M} \to \mathbb{R}^k$, we can define ∇s via

$$\pi^* \nabla s = d\hat{s},$$

cf. (47).

These constructions establish a bijective correspondence between the space of all flat connections $\mathcal{M}^\flat$ and the representation variety

$$\mathcal{R}(M; \mathrm{GL}_k(\mathbb{R})) := \{\rho\colon \pi_1(M) \to \mathrm{GL}_k(\mathbb{R}) \text{ is a group homomorphism }\}/\mathrm{Conj},$$

where two representations are considered to be equivalent if they are conjugate.

Remark 93. We could equally well consider flat connections on Euclidean or Hermitian vector bundles. This requires only cosmetic changes and the outcome is the representation space $\mathcal{R}(M;\, \mathrm{O}(k))$ and $\mathcal{R}(M;\, \mathrm{U}(k))$ respectively. Even more generally, the constructions above can be modified to the case of principal G–bundles so that the space of flat G–connections correspond to $\mathcal{R}(M;\, \mathrm{G})$. I leave the details to the readers.

Since the fundamental group of a manifold is finitely presented, we can choose a finite number of generators of $\pi_1(M)$, say $\gamma_1, \dots, \gamma_N$. Then any representation ρ is uniquely specified by the images of the generators $g_i = \rho(\gamma_i) \in G$, which satisfy a finite number of relations. This shows the inclusion Ref ?

$$\mathcal{R}(M;\, \mathrm{G}) \subset G^N/G,$$

where G acts by the adjoint action on each factor. This implies in particular the following.

Proposition 94. *Let M be a manifold. If* G *is a compact Lie group, then the space $\mathcal{M}^\flat$ of all flat* G*-connections is compact.* □

Example 95. For $M = \mathbb{T}^n$, we have clearly

$$\mathcal{R}(\mathbb{T}^n;\, \mathrm{U}(1)) = \mathrm{Hom}(\mathbb{T}^n, \mathrm{U}(1)) = \mathrm{U}(1)^n \cong \mathbb{T}^n.$$

Add a more interesting example.

4 Dirac operators

4.1 Spin groups and Clifford algebras

In this subsection I recall briefly the notions of Clifford algebra and spin group focusing on low dimensions. More details can be found for instance in [LM89].

Since $\pi_1(\mathrm{SO}(n)) \cong \mathbb{Z}/2\mathbb{Z}$ for any $n \geq 3$, there is a simply connected Lie group denoted by $\mathrm{Spin}(n)$ together with a homomorphism $\mathrm{Spin}(n) \to \mathrm{SO}(n)$, which is a double covering. This characterizes $\mathrm{Spin}(n)$ up to an isomorphism. The spin groups can be constructed explicitly with the help of Clifford algebras, however in low dimensions this can be done more directly with the help of quaternions.

Since $\mathrm{Sp}(1) \cong \mathrm{SU}(2)$ is diffeomorphic to the 3–sphere, this is a connected and simply connected Lie group. Identify $\mathrm{Im}\,\mathbb{H} = \{\bar h = -h\}$ with $\mathbb{R}^3$ and consider the homomorphism

$$\alpha\colon \mathrm{Sp}(1) \to \mathrm{SO}(3), \qquad q \mapsto A_q, \tag{96}$$

where $A_q h = qh\bar{q}$. It is easy to check that the corresponding Lie-algebra homomorphism is in fact an isomorphism. Since $\mathrm{SO}(3)$ is connected, α is surjective. Moreover, $\ker\alpha = \{\pm 1\}$. Hence, (96) is a non-trivial double covering, i.e., $\mathrm{Spin}(3) \cong \mathrm{Sp}(1)$.

To construct the group $\mathrm{Spin}(4)$, recall first that the Hodge operator $*$ yields the splitting $\Lambda^2(\mathbb{R}^4)^* = \Lambda^2_+(\mathbb{R}^4)^* \oplus \Lambda^2_-(\mathbb{R}^4)^*$, where $\Lambda^2_\pm(\mathbb{R}^4)^* = \{\omega \mid *\omega = \pm\omega\}$. Since $\mathfrak{so}(4) \cong \Lambda^2(\mathbb{R}^4)^* = \Lambda^2_+(\mathbb{R}^4)^* \oplus \Lambda^2_-(\mathbb{R}^4)^* = \mathfrak{so}(3) \oplus \mathfrak{so}(3)$, the adjoint representation yields a homomorphism $\mathrm{SO}(4) \to \mathrm{SO}(3) \times \mathrm{SO}(3)$.

Identify $\mathbb{R}^4$ with $\mathbb{H}$ and consider the homomorphism[3]

$$\beta\colon \mathrm{Sp}_+(1) \times \mathrm{Sp}_-(1) \to \mathrm{SO}(4), \qquad (q_+, q_-) \mapsto A_{q_+,q_-},$$

where $A_{q_+,q_-} h = q_+ h \bar{q}_-$. An explicit computation shows that the composition $\mathrm{Sp}_+(1) \times \mathrm{Sp}_-(1) \to \mathrm{SO}(4) \to \mathrm{SO}(3) \times \mathrm{SO}(3)$ is given by $(q_+, q_-) \mapsto (A_{q_+}, A_{q_-})$. Hence, the Lie algebra homomorphism corresponding to β is an isomorphism and $\ker\beta$ is contained in $\{(\pm 1, \pm 1)\}$. As it is readily checked, $\ker\beta = \{\pm(1,1)\} \cong \mathbb{Z}/2\mathbb{Z}$. Hence, $\mathrm{Sp}_+(1) \times \mathrm{Sp}_-(1) \cong \mathrm{Spin}(4)$.

Let U be an Euclidean vector space. Then the Clifford algebra $Cl(U)$ is the tensor algebra $TU = \mathbb{R} \oplus U \oplus U \otimes U \oplus \dots$ modulo the ideal generated by elements $u \otimes u + |u|^2 \cdot 1$. In other words, $Cl(U)$ is generated by elements of U subject to the relations $u \cdot u = -|u|^2$. For instance, $Cl(\mathbb{R}^1) \cong \mathbb{R}[x]/(x^2+1) \cong \mathbb{C}$. The algebra $Cl(\mathbb{R}^2)$ is generated by $1, e_1, e_2$ subject to the relations $e_1^2 = -1 = e_2^2$ and $e_1 \cdot e_2 = -e_2 \cdot e_1$, which follows from $(e_1 + e_2)^2 = -2$. In other words, $Cl(\mathbb{R}^2) \cong \mathbb{H}$. In general, $Cl(\mathbb{R}^n)$ is generated by $1, e_1, \dots, e_n$ subject to the relations $e_i^2 = -1$ and $e_i \cdot e_j = -e_j \cdot e_i$ for $i \neq j$.

It can be shown that the subgroup of $Cl(\mathbb{R}^n)$ generated by elements of the form $v_1 \cdot v_2 \cdot \ldots \cdot v_{2k}$ is isomorphic to $\mathrm{Spin}(n)$, where each $v_j \in \mathbb{R}^n$ has the unit norm. In particular, this shows that $\mathrm{Spin}(n)$ is a subgroup of $\mathrm{Cl}(\mathbb{R}^n)$.

It is convenient to have some examples of modules over Clifford algebras. Such module is given by a vector space V together with a map

$$U \otimes V \to V, \qquad u \otimes v \mapsto u \cdot v,$$

which satisfies $u \cdot (u \cdot v) = -|u|^2 v$ for all $u \in U$ and $v \in V$. An example of a $Cl(U)$–module is $V = \Lambda U^*$, where the $Cl(U)$–module structure is given by the map

$$u \otimes \varphi \mapsto \imath_u \varphi - \langle u, \cdot\rangle \wedge \varphi. \tag{97}$$

Let V be a quaternionic vector space. Then the quaternionic multiplication gives rise to the map $\operatorname{Im}\mathbb{H} \otimes V \to V$, $h \otimes v \mapsto h \cdot v$, which satisfies $h \cdot (h \cdot v) = -h\bar{h}v = -|h|^2 v$. Thus, any quaternionic vector space is a $Cl(\mathbb{R}^3)$–module. In particular, the fundamental representation $\$ \cong \mathbb{H}$ of $\mathrm{Sp}(1) \cong \mathrm{Spin}(3)$ with the action given by the left multiplication is a $Cl(\mathbb{R}^3)$–module.

The multiplication on the right by $\bar{i}$ endows $\$$ with the structure of a complex $\mathrm{Sp}(1)$–representation, which is in fact also Hermitian (this is just another manifestation of the isomorphism $\mathrm{Sp}(1) \cong \mathrm{SU}(2)$). It is then an elementary exercise in the representation theory to show the isomorphisms

$$\operatorname{Im}\mathbb{H} \otimes \mathbb{C} \cong \operatorname{End}_0(\$), \tag{98}$$

[3] We adopt the common convention $Sp_\pm(1) = Sp(1)$. The significance of the subscripts "$\pm$" will be clear below.

where the left hand side is viewed as an $\mathrm{Sp}(1)$–representation via the homomorphism α and $\mathrm{End}_0(\displaystyle{\not}S)$ denotes the subspace of traceless endomorphisms. Moreover, the real subspace $\mathrm{Im}\,\mathbb{H}$ can be identified with the subspace of traceless Hermitian endomorphisms.

Furthermore, for any quaternionic vector space V the space $V \oplus V$ is a $Cl(\mathbb{R}^4)$–module. Indeed, the $Cl(\mathbb{R}^4)$–module structure is induced by the map

$$\mathbb{H} \otimes_{\mathbb{R}} (V \oplus V) \to V \oplus V, \qquad h \otimes (v_1, v_2) \mapsto (hv_2, -\bar{h}v_1) = \begin{pmatrix} 0 & h \\ -\bar{h} & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \tag{99}$$

In particular, the $\mathrm{Sp}_+(1) \times \mathrm{Sp}_-(1)$–representation $\not{S}^+ \oplus \not{S}^-$ is a $Cl(\mathbb{R}^4)$–module. Here, as the notation suggests, $\not{S}^\pm$ is the fundamental representation of $\mathrm{Sp}_\pm(1)$.

Just like in the case of dimension three, we have an isomorphism of $\mathrm{Spin}(4)$–representations

$$\mathbb{H} \otimes \mathbb{C} \cong \mathrm{Hom}(\not{S}^+; \not{S}^-),$$

where the left hand side is viewed as a $\mathrm{Spin}(4)$–representation via the homomorphism β.

4.2 Dirac operators

Let M be a Riemannian oriented manifold of dimension n. The tautological action of $\mathrm{SO}(n)$ on $\mathbb{R}^n$ extends to an action on $Cl(\mathbb{R}^n)$ so that we can construct the associated bundle

$$Cl(M) := \mathrm{Fr}_{\mathrm{SO}} \times_{\mathrm{SO}(n)} Cl(\mathbb{R}^n).$$

This can be thought of as the bundle, whose fiber at a point $m \in M$ is $Cl(T_mM) \cong Cl(T_m^*M)$. Notice that the Levi–Civita connection yields a connection on $Cl(M)$. This is denoted by the same symbol ∇^{LC}.

Let $E \to M$ be a bundle of $Cl(M)$–modules, i.e., there is a morphism of vector bundles

$$Cl\colon TM \otimes E \to E, \qquad (v, e) \mapsto v \cdot e,$$

such that $v \cdot (v \cdot e) = -|v|^2 e$. Then E is called a Dirac bundle if it is equipped with an Euclidean scalar product and a connection ∇ such that the following conditions hold:

- ∇ is Euclidean;
- $\langle v \cdot e_1, v \cdot e_2 \rangle = |v|^2 \langle e_1, e_2 \rangle$ for any $v \in T_mM$ and $e_1, e_2 \in E_m$;
- $\nabla(\varphi \cdot s) = (\nabla^{LC}\varphi) \cdot s + \varphi \cdot \nabla s$ for any $\varphi \in \Gamma(Cl(M))$ and $s \in \Gamma(E)$.

Definition 100. If E is a Dirac bundle, the operator

$$\mathcal{D}\colon \Gamma(E) \xrightarrow{\nabla} \Gamma(T^*M \otimes E) \xrightarrow{Cl} \Gamma(E)$$

is called the Dirac operator of E.

In other words, if $e_1, \dots, e_n$ is a local orthonormal oriented frame of TM, then

$$\mathcal{D}\, s = \sum_{i=1}^{n} e_i \cdot \nabla_{e_i} s$$

Example 101. The bundle $\Lambda T^*M = \oplus_{k=0}^n \Lambda^k T^*M$ has a natural structure of a Dirac bundle, where the Clifford multiplication is given by (97). The corresponding Dirac operator equals $d+d^*$ [LM89, Thm 5.12], where d^* is the formal adjoint of d, see Section 5.3.2 below for more details.

Exercise 102. Show that the Dirac operator on a closed manifold is formally self-adjoint, i.e., for any $s_1, s_2 \in \Gamma(E)$ we have

$$\int_M \langle \mathcal{D}\, s_1,\, s_2\rangle = \int_M \langle s_1,\, \mathcal{D}\, s_2\rangle.$$

4.3 Spin and Spin^c structures

Let $\mathrm{Fr}_{\mathrm{SO}} \to M$ be the principal bundle of orthonormal oriented frames of M.

Definition 103. M is said to be spinnable, if there is a principal $\mathrm{Spin}(n)$ bundle P equipped with a $\mathrm{Spin}(n)$–equivariant map $\tau\colon P \to \mathrm{Fr}_{\mathrm{SO}}$, which covers the identity map on M and is a fiberwise double covering. Here $\mathrm{Spin}(n)$ acts on $\mathrm{Fr}_{\mathrm{SO}}$ via the homomorphism $\mathrm{Spin}(n) \to \mathrm{SO}(n)$.

A choice of a bundle P as above is called a spin structure. A manifold equipped with a spin structure is called a spin manifold.

For a given M a spin structure may or may not exist. If M is spinnable, there may be many non-equivalent spin structures. The questions on existence and classification of spin structures may be completely answered in terms of the Stiefel–Whitney classes [LM89]. However, I will not go into the details here. For the remaining part of this section I assume throughout that M is spin.

Let $\omega \in \Omega^1\big(\mathrm{Fr}_{\mathrm{SO}};\, \mathfrak{so}(n)\big)$ be the connection 1-form of the Levi–Civita connection. Since the homomorphism $\mathrm{Spin}(n) \to \mathrm{SO}(n)$ is a local diffeomorphism, we have an isomorphism of Lie algebras $\mathfrak{spin}(n) \cong \mathfrak{so}(n)$. Hence, $\tau^*\omega \in \Omega^1\big(P;\, \mathfrak{spin}(n)\big)$ is a connection on P. Slightly abusing terminology, this is still called the Levi–Civita connection.

For any $n \geq 3$ there is a unique complex representation $\rho\colon \mathrm{Spin}(n) \to \mathrm{End}(\$)$ distinguished by the property that it extends to a complex irreducible representation of $\mathrm{Cl}(\mathbb{R}^n)$. Notice that this means neither that $\$$ is a unique $\mathrm{Spin}(n)$–representation, nor that $\$$ is an irreducible $\mathrm{Spin}(n)$–representation. For example, for $n=3$ this representation coincides with the fundamental representation of $\mathrm{Sp}(1) = \mathrm{Spin}(3)$. For $n=4$ we have $\$ = \$^+ \oplus \$^-$, see (99).

If M is spin, we can construct *the spinor bundle*

$$\$:= P \times_{\mathrm{Spin}(n),\,\rho} \$.$$

Here, as it is quite common, we use the same notation both for the representation and the associated vector bundle.

Since ρ extends to a representation of $Cl(\mathbb{R}^n)$, the spinor bundle is in fact a bundle of $Cl(M)$–modules. Hence, the construction of Section 4.2 yields the spin Dirac operator

$$\not{D}\colon \Gamma(\$) \to \Gamma(\$).$$

Remark 104. Recall that in the case $\dim M = 4$, the spinor bundle splits: $\$ = \$^+ \oplus \$^-$. By (99), the Clifford multiplication with 1–forms changes the chirality, i.e., for any $\omega \in \Omega^1(M)$ we have $\omega\cdot\colon \$^\pm \to \$^\mp$. Hence,

$$\not{D} = \begin{pmatrix} 0 & \not{D}^- \\ \not{D}^+ & 0 \end{pmatrix}, \qquad \text{where } \not{D}^\pm\colon \Gamma(\$^\pm) \to \Gamma(\$^\mp).$$

The following two variations of this construction are frequently used. First, let P be a principal G–bundle equipped with a connection A and let $\tau\colon \mathrm{G} \to \mathrm{U}(n)$ be a unitary representation of G so that we have the associated bundle $E = P \times_{\mathrm{G},\tau} \mathbb{C}^n$, which is Hermitian. Then the twisted spinor bundle $\not{S} \otimes E$ is also a Dirac bundle so that we have a twisted Dirac operator

$$\not{D}_A\colon \Gamma(\not{S} \otimes E) \to \Gamma(\not{S} \otimes E).$$

Example 105. Let us assume that $\dim M = 3$ for the sake of definiteness. Choose $E = \not{S}$, which is equipped with the Levi–Civita connection. Then we have

$$\not{S} \otimes \not{S} = \mathrm{Sym}^2(\not{S}) \oplus \Lambda^2\not{S} \cong T^*_{\mathbb{C}}M \oplus \mathbb{C},$$

cf. (98). Hence, the twisted spinors can be identified with the complexification of odd forms. Of course, the Hodge $*$-operator yields and isomorphism between odd and even forms so that we can identify the twisted spinors with even forms too. We already have seen above a Dirac operator acting on forms, namely

$$d + d^*\colon \Omega^{\mathrm{odd}}(M) \to \Omega^{\mathrm{even}}(M),$$

cf. Example 101. One can show that the complexification of $d + d^*$ coincides with the twisted Dirac operator on $\Gamma(\not{S} \otimes \not{S})$.

Exercise 106. The Clifford multiplication combined with the map $\mathrm{ad}\, P \to \mathrm{End}(E)$ yields the 'twisted' Clifford multiplication

$$T^*M \otimes \mathrm{ad}\, P \to \mathrm{End}(\not{S}) \otimes \mathrm{End}(E) \cong \mathrm{End}(\not{S} \otimes E).$$

Show that for $a \in \Omega^1(\mathrm{ad}\, P)$ the following holds:

$$\not{D}_{A+a}\psi = \not{D}_A\psi + a \cdot \psi.$$

Let me explain the second variation, which in essence is not really much different from the first one, however the details are somewhat involved.

Thus, by the construction of $\mathrm{Spin}(n)$ we have the exact sequence

$$\{1\} \to \{\pm 1\} \to \mathrm{Spin}(n) \to \mathrm{SO}(n) \to \{1\}.$$

Moreover, one can show that the kernel $\{\pm 1\}$ lies in the center of $\mathrm{Spin}(n)$. This is clear anyway in our main cases of interest, namely for $n = 3$ and $n = 4$. Denote

$$\mathrm{Spin}^c(n) := \mathrm{Spin}(n) \times \mathrm{U}(1)/\pm 1,$$

where $\{\pm 1\}$ is embedded diagonally. Notice that both $\mathrm{Spin}(n)$ and $\mathrm{U}(1)$ are subgroups of $\mathrm{Spin}^c(n)$ and $\mathrm{U}(1)$ lies in fact in the center of $\mathrm{Spin}^c(n)$.

Furthermore, we have the following exact sequences:

$$\{1\} \to \mathrm{U}(1) \to \mathrm{Spin}^c(n) \xrightarrow{\rho_0} \mathrm{Spin}(n)/\pm 1 = \mathrm{SO}(n) \to \{1\},$$
$$\{1\} \to \mathrm{Spin}(n) \to \mathrm{Spin}^c(n) \xrightarrow{\rho_{\det}} \mathrm{U}(1)/\pm 1 \cong \mathrm{U}(1) \to \{1\}.$$

These in turn give rise to the exact sequence

$$\{1\} \to \{\pm 1\} \to \mathrm{Spin}^c(n) \xrightarrow{(\rho_0,\, \rho_{\det})} \mathrm{SO}(n) \times \mathrm{U}(1) \to \{1\}, \tag{107}$$

which shows that $\mathrm{Spin}^c(n)$ is a double covering of $\mathrm{SO}(n) \times \mathrm{U}(1)$.

Example 108.

(a) For $n = 3$ we have $\mathrm{Spin}^c(3) = \mathrm{SU}(2) \times \mathrm{U}(1)/\pm 1 \cong \mathrm{U}(2)$. In particular, $\rho_{\det}(A) = \det A$ and the sequence (107) has the following form

$$\{1\} \to \big\{\pm\mathbb{1}\big\} \to \mathrm{U}(2) \to \mathrm{SO}(3) \times \mathrm{U}(1) \to \{1\},$$

where the homomorphism $\mathrm{U}(2) \to \mathrm{SO}(3) = \mathrm{PU}(2)$ is the natural projection.

(b) For $n = 4$ we have

$$\begin{aligned}\mathrm{Spin}^c(4) &= \Big(\big(\mathrm{SU}(2) \times \mathrm{SU}(2)\big) \times \mathrm{U}(1)\Big)/\pm 1 \\ &= \big\{(A_+,\, A_-) \in \mathrm{U}(2) \times \mathrm{U}(2) \mid \det A_+ = \det A_-\big\}.\end{aligned}$$

In particular, $\rho_{\det}(A_+,\, A_-) = \det A_+ = \det A_-$.

Notice that in this case we also have the homomorphisms

$$\rho_\pm\colon \mathrm{Spin}^c(4) \to \mathrm{U}(2), \qquad \rho_\pm(A_+,\, A_-) = A_\pm.$$

Definition 109. A spin^c structure on M is a principal $\mathrm{Spin}^c(n)$–bundle $P \to M$ equipped with a $\mathrm{Spin}^c(n)$–equivariant map $P \to \mathrm{Fr}_{\mathrm{SO}}$ which induces an isomorphism $P/\mathrm{U}(1) \cong \mathrm{Fr}_{\mathrm{SO}}$.

Just like in the case of spin structures, spin^c structures may or may not exist. If a spin^c structure exists, it is rarely unique. It is also clear that if M is spin, then there is also a spin^c structure on M.

Our main interest in spin^c structures stems from the following result.

Theorem 110 (Cite:Gompf–Stipsicz)**.** *Any closed oriented four–manifold admits a spin^c structure.* □

Notice however, that there are closed four–manifolds that are not spin. In dimension three the situation is different: Any closed oriented three–manifold is spin , and hence also spin^c. Citation!

With these preliminaries at hand, let M be a manifold equipped with a spin^c structure P. Define *the determinant line bundle*

$$L_{\det} := P \times_{\rho_{\det}} \mathbb{C}.$$

This is clearly a Hermitian line bundle and its $\mathrm{U}(1)$–structure is

$$P_{\det} := P/\mathrm{Spin}(n).$$

By (107), we have a double cover map

$$\tau\colon P \to P/\{\pm 1\} = \mathrm{Fr}_{\mathrm{SO}} \times_M P_{\det}.$$

Hence, if A is a connection on $P_{\det}$, then $\tau^*(\omega + A) \in \Omega^1\big(P;\, \mathfrak{spin}^c(n)\big)$, where we have used the isomorphism $\mathfrak{spin}^c(n) \cong \mathfrak{so}(n) \oplus \mathfrak{u}(1)$. Thus, the choice of a unitary connection on the determinant line bundle together with the Levi–Civita connection determines a connection on the spin^c bundle.

Let $\slashed{S}$ be the distinguished representation of $\mathrm{Spin}(n)$. This can be clearly extended to a $\mathrm{Spin}^c(n)$–representation, which is still denoted by $\slashed{S}$: $[g, z] \cdot s = z\rho(g)\, s$. This in turn yields the spin^c spinor bundle

$$\slashed{S} := P \times_{\mathrm{Spin}^c(n)} \slashed{S},$$

which is a Dirac bundle. Hence, we obtain the spin^c Dirac operator $\slashed{D}_A\colon \Gamma(\slashed{S}) \to \Gamma(\slashed{S})$.

Example 111. Assume M is spin and pick a spin structure P_{Spin}. Pick also a principal $\mathrm{U}(1)$–bundle P_0. Then

$$P := P_{\mathrm{Spin}} \times_M P_0 / \pm 1$$

is a principal $\big(\mathrm{Spin}(n) \times \mathrm{U}(1)\big)/\pm 1 = \mathrm{Spin}^c(n)$ bundle. In this case the spin^c spinor bundle is just the twisted spinor bundle $\not{S} \otimes L_0$, where $\not{S}$ is the pure spinor bundle and $L_0 := P_0 \times_{\mathrm{U}(1)} \mathbb{C}$ is the associated complex line bundle. Likewise, the spin^c Dirac operator is just the twisted Dirac operator.

Remark 112. In some sense, any *spin^c* spinor bundle can be thought of as the twisted spinor bundle just like in the example above. The problem is that in general neither $\not{S}$ nor L_0 is globally well-defined, however their product $\not{S} \otimes L_0$ is well-defined. The notion of a spin^c structure is just a convenient way to make this 'definition' precise.

Exercise 113. For $a \in \Omega^1(M; \mathbb{R}i)$ prove that

$$\not{D}_{A+a}\psi = \not{D}_A \psi + \frac{1}{2}\, a \cdot \psi. \tag{114}$$

The coefficient $\frac{1}{2}$ in (114) can be explained as follows: Think of the spin^c spinor bundle as $\not{S} \otimes L_0$ just like in Remark 112. Then the determinant line bundle is $\Lambda^2(\not{S} \otimes L_0) = L_0^2$ so that $A_0 \in \mathcal{A}(L_0)$ induces some connection A on L_0^2. Then $A_0 + a_0$ corresponds to $A + 2\, a_0$.

4.3.1 On the classification of spin^c structures

Let P be a spin^c structure with the spinor bundle $\not{S}$. If L is any Hermitian line bundle, then $\not{S} \otimes L$ is a spin^c spinor bundle corresponding to a spin^c structure P_L. This defines an action of $H^2(M; \mathbb{Z})$ on the set $\mathcal{S} = \mathcal{S}(M)$ of all spin^c structures on M. This action can be shown to be free and transitive, hence $\mathcal{S}(M)$ can be identified with $H^2(M; \mathbb{Z})$, however, such an identification is not canonical.

For example, let M be spin and let $\not{S}$ be the pure spinor bundle. The spin structure of M can be also viewed as a distinguished spin^c structure so that $\mathcal{S}(M)$ has a preferred point, the origin. This choice fixes an isomorphism $\mathcal{S}(M) \cong H^2(M; \mathbb{Z})$.

4.4 The Weitzenböck formula

For a function $u\colon \mathbb{R}^4 \to \mathbb{H}$ the operators

$$\begin{aligned}
\not{D}^+(u) &= \quad \frac{\partial u}{\partial x_0} + i\frac{\partial u}{\partial x_1} + j\frac{\partial u}{\partial x_2} + k\frac{\partial u}{\partial x_3}, \\
\not{D}^-(u) &= -\frac{\partial u}{\partial x_0} + i\frac{\partial u}{\partial x_1} + j\frac{\partial u}{\partial x_2} + k\frac{\partial u}{\partial x_3}
\end{aligned}$$

can be thought of as four-dimensional analogues of the $\bar\partial = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$ and $\partial = \frac{1}{2}(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y})$ operators for complex–valued functions of one complex variable $z = x + yi$ respectively. In fact, tracing through the construction, it is easy to see that for the flat four–manifold $\mathbb{R}^4$ the Dirac operator can be written as follows

$$\not{D} = \begin{pmatrix} 0 & \not{D}^- \\ \not{D}^+ & 0 \end{pmatrix},$$

cf. (99). Here I use the fact, that the spinor bundle of $\mathbb{R}^4$ is (canonically) the product bundle. A straightforward computation yields

$$\not{D}^2 = \begin{pmatrix} \not{D}^- \not{D}^+ & 0 \\ 0 & \not{D}^+ \not{D}^- \end{pmatrix} = \Delta, \tag{115}$$

where $\Delta = -\sum_{i=0}^{3} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian. This is usually phrased as "the Dirac operator is the square root of the Laplacian".

Remark 116. The Dirac operator on $\mathbb{R}^2$ has the form

$$\begin{pmatrix} 0 & 2\frac{\partial}{\partial z} \\ -2\frac{\partial}{\partial \bar{z}} & 0 \end{pmatrix} : C^\infty(\mathbb{R}^2;\, \mathbb{C}^2) \to C^\infty(\mathbb{R}^2;\, \mathbb{C}^2)$$

and squares to the Laplacian just like in dimension four. Notice, however, that the case of dimension two requires a special treatment due to the fact that $\pi_1(\mathrm{SO}(2)) \cong \mathbb{Z}$, which was one of the reasons I assumed $n \geq 3$ in this section.

On $\mathbb{R}^3$ the spinor bundle can be identified with the product bundle $\underline{\mathbb{H}}$ so that the corresponding Dirac operator is

$$\not{D}(u) = i\frac{\partial}{\partial x_1} + j\frac{\partial}{\partial x_2} + k\frac{\partial}{\partial x_3},$$

which also squares to the Laplacian.

On general Riemannian manifolds the relation $\not{D}^2 = \Delta$ still holds up to a zero order operator. This is known as the Weitzenböck formula, which I describe next.

Thus, given an Euclidean vector bundle E with a connection ∇ *the connection Laplacian* is defined by

$$\nabla^*\nabla\colon \Gamma(E) \xrightarrow{\nabla} \Omega^1(E) \xrightarrow{\nabla^* = -*d_\nabla *} \Gamma(E). \tag{117}$$

Exercise 118.

(a) Let $(e_1, \ldots, e_n)$ be a local orthonormal frame of TM over an open subset of M. Show that the connection Laplacian can be expressed as follows

$$\nabla^*\nabla s = -\sum_{i=1}^{n} \big(\nabla_{e_i} \nabla_{e_i} s - \nabla_{\nabla_{e_i} e_i}\, s \big),$$

where ∇e_i means the Levi–Civita connection applied to e_i.

(b) Prove that the connection Laplacian is formally self-adjoint.

(c) In the case M is closed, prove the identity $\langle \nabla^*\nabla s,\, s\rangle_{L^2} = \|\nabla s\|^2_{L^2}$.

Assume that E is a Dirac bundle and let $R \in \Omega^2(\mathrm{End}(E))$ be the curvature 2-form of the corresponding connection ∇. Using the Clifford multiplication, we obtain a homomorphism $\Lambda^2 T^*M \otimes \mathrm{End}(E) \to \mathrm{End}(E)$, which maps R to an endomorphism $\mathfrak{R}$. In a local frame (e_i) of TM this can be expressed as follows

$$\mathfrak{R}(s) := \frac{1}{2} \sum_{i,j=1}^{n} e_i \cdot e_j \cdot R_{e_i, e_j}(s).$$

Theorem 119 (Weitzenböck formula)**.** *Let $\not{D}$ be the Dirac operator for the Dirac bundle E. Then the following holds:*

$$\not{D}^2 = \nabla^*\nabla + \mathfrak{R}.$$

Proof. Pick a point $m \in M$ and a local frame (e_i) such that $\nabla_{e_i} e_j$ vanishes at m. Then at m we have the following:

$$\begin{aligned}
\not{D}^2 s &= \sum_{i=1}^n e_i \cdot \nabla_{e_i}\Big(\sum_{j=1}^n e_j \cdot \nabla_{e_j} s\Big) = \sum_{i,j=1}^n e_i \cdot e_j \cdot \nabla_{e_i}\nabla_{e_j} s \\
&= \sum_{i=1}^n e_i \cdot e_i \cdot \nabla_{e_i}\nabla_{e_i} s + \frac{1}{2}\sum_{i\neq j} e_i \cdot e_j \cdot \big(\nabla_{e_i}\nabla_{e_j} - \nabla_{e_j}\nabla_{e_i}\big) s \\
&= \nabla^*\nabla s + \mathfrak{R} s.
\end{aligned}$$

□

In the case of the spin^c spinor bundle, a straightforward computation yields the following corollary, whose proof is left as an exercise.

Corollary 120. *Let A be a Hermitian connection on the determinant line bundle with the curvature form F_A. Then the spin^c Dirac operator satisfies:*

$$\not{D}^2 \psi = \nabla^*\nabla\psi + \frac{1}{4} s_g\,\psi + \frac{1}{2} F_A \cdot \psi,$$

where s_g is the scalar curvature of the background metric g. □

In particular, in the case M is spin and $\not{S}$ is the pure spinor bundle, we have

$$\not{D}^2 \psi = \nabla^*\nabla\psi + \frac{1}{4} s_g\,\psi.$$

This implies, for example, that for a metric with positive scalar curvature the space of harmonic spinors $\ker \not{D}$ is trivial.

5 Linear elliptic operators

5.1 Sobolev spaces

Consider the following classical boundary value problem in the theory of PDEs: Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with a smooth boundary $\partial\Omega$. Does there exist a function $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ such that

$$\Delta u = 0 \quad \text{in } \Omega \qquad \text{and} \qquad u|_{\partial\Omega} = \varphi, \tag{121}$$

where φ is a given function on $\partial\Omega$?

Consider *the energy functional*

$$E(u) := \frac{1}{2}\int_\Omega |\nabla u(x)|^2\, dx,$$

where dx denotes the standard volume form on $\mathbb{R}^n$. Assume there exists an absolute minimum of E, i.e., a function $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ such that

$$E(u) = \inf\big\{ E(v) \mid v \in C^2(\Omega) \cap C^0(\bar{\Omega}),\ v|_{\partial\Omega} = \varphi \big\}.$$

A straightforward computation using integration by parts yields

$$0 = \frac{d}{dt}\Big|_{t=0} E(u + tw) = \int_\Omega w\, \Delta u\, dx$$

for all w such that $w|_{\partial\Omega} = 0$. This implies that u is harmonic in Ω. Thus, we can find a solution of (121) if we can prove that the energy functional attains its minimum.

Hence, a strategy for proving the existence of solutions of (121) may be the following: Pick a sequence u_k such that $E(u_k)$ converges to the infimum of the energy functional and prove that u_k converges to a limit u possibly after extracting a subsequence. This strategy does work indeed, however, it requires certain technology known as the theory of Sobolev spaces, which turned out to be very useful in gauge theory as well. What follows below is a crash course in Sobolev spaces. An interested reader should consult more specialized literature for details.

First recall that the space $L^2(\mathbb{R}^n)$ of square integrable functions equipped with a scalar product

$$\langle u, v\rangle_{L^2} := \int_{\mathbb{R}^n} uv\, dx$$

is a Hilbert space, i.e., $L^2(\mathbb{R}^n)$ is complete with respect to the norm $\|u\|_{L^2} := \sqrt{\langle u, u\rangle}$. This space can be viewed as a completion of the space $C_0^\infty(\mathbb{R}^n)$ of all smooth functions with compact support with respect to the norm $\|\cdot\|_{L^2}$.

This definition admits a number of variations, which will be of use below. First, we can pick any $p > 1$ and put

$$\|u\|_{L^p} := \Big(|u(x)|^p\, dx \Big)^{\frac{1}{p}},$$

where $u \in C_0^\infty(\mathbb{R}^n)$. The completion of $C_0^\infty(\mathbb{R}^n)$ with respect to $\|\cdot\|_{L^p}$ is then the Banach space $L^p(\mathbb{R}^n)$.

Secondly, we can also consider

$$\|u\|_{W^{k,p}} := \Big(\sum_{i=0}^{k} \|\nabla^i u\|_{L^p} \Big)^{\frac{1}{p}}.$$

The completion of $C_0^\infty(\mathbb{R}^n)$ with respect to this norm is denoted by $W^{k,p}(\mathbb{R}^n)$. These spaces are called *Sobolev spaces*.

Remark 122. Many different and inconsistent notations are in use for Sobolev spaces. For example, sometimes L^s_r may mean $W^{r,s}$ and sometimes $W^{s,r}$. Of all notations being used for Sobolev spaces, the symbol $W^{k,p}$ seems to be most consistently used, hence I opted for this one.

Third, we can replace the domain $\mathbb{R}^n$ by any open subset of $\mathbb{R}^n$ or, even more generally, by a Riemannian manifold M. The corresponding spaces will be denoted by $W^{k,p}(M)$. Here ∇u should be understood as the differential of u, that is $\nabla u \in \Gamma(T^*M)$. fixing a connection on T^*M, for example the Levi–Civita one, ∇u can be differentiated again. This clarifies the meaning of $\nabla^2 u \in \Gamma(T^*M \otimes T^*M)$ and so on up to the dependence of the space $W^{k,p}(M)$ on the choice of connection.

This naturally leads us to one more variation of the definitions above. Pick a vector bundle $E \to M$ and connections $\nabla^E \in \mathcal{A}(E), \nabla^M \in \mathcal{A}(T^*M)$. This yields a connection on $T^*M \otimes E$ so that the higher derivatives $\nabla^i(\nabla s)$ are well defined for any $s \in \Gamma(E)$. This yields the Sobolev spaces $W^{k,p}(M;\, E)$ in the same manner as above. While the norm $\|\cdot\|_{W^{k,p}}$ does depend on ∇^E and ∇^M, different choices yield equivalent norms so that the resulting topology is independent of the choices made.

With this understood, for any $p > 1$ we have the sequence of inclusions

$$L^p(M;\, E) = W^{0,p}(M;\, E) \supset W^{1,p}(M;\, E) \supset W^{2,p}(M;\, E) \supset \dots$$

Relations between all these spaces is given by the following theorem, which is of fundamental importance in the theory of PDEs.

Theorem 123. *Let M be a compact manifold.*

(i) If $s \in W^{k,p}(M;\, E)$, then $s \in W^{m,q}(M;\, E)$ provided

$$k - \frac{n}{p} \geq m - \frac{n}{q} \quad \text{and} \quad k \geq m,$$

where $n = \dim M$, and there is a constant C independent of s such that $\|s\|_{W^{m,s}} \leq C\|s\|_{W^{k,p}}$. In other words, the natural embedding

$$j\colon W^{k,p}(M;\, E) \subset W^{m,q}(M;\, E)$$

is continuous.

(ii) j is a compact operator provided

$$k - \frac{n}{p} > m - \frac{n}{q} \quad \text{and} \quad k > m. \tag{124}$$

This means that any sequence bounded in $W^{k,p}$ has a subsequence, which converges in $W^{m,q}$ provided (124) holds.

(iii) We have a natural continuous embedding

$$W^{k,p}(M;\, E) \subset C^r(M;\, E)$$

provided $k - \frac{n}{p} > r$. In particular, if $s \in W^{k,p}(M;\, E)$ for some fixed p and for all $k \geq 0$, then $s \in C^\infty(M;\, E)$.

(iv) (a) In the case $kp > n$ the space $W^{k,p}(M;\, \mathbb{R})$ is an algebra.

(b) In the case $kp < n$, we have a bounded map

$$W^{k_1,p_1} \otimes W^{k_2,p_2} \to W^{k,p}, \qquad \text{provided} \quad k_1 - \frac{n}{p_1} + k_2 - \frac{n}{p_2} \geq k - \frac{n}{p}. \qquad \square$$

Remark 125. Although the proof of this theorem goes beyond the goals of these notes, it may be instructive to see 'the spirit of the proof' in one particular case. Thus, let $M = S^1$ and $u \in C^\infty(S^1; \mathbb{R})$. Denote $\bar u := \frac{1}{2\pi}\int u(\theta)\,d\theta$ and $u_0(\theta) := u(\theta) - \bar u$. By the mean value theorem, there is $\theta_0 \in S^1$ such that $u_0(\theta_0) = 0$. Hence, for any $\theta \in S^1$ we have

$$|u_0(\theta)| = \Big|\int_{\theta_0}^{\theta} u_0'(\varphi)\,d\varphi\Big| \le \sqrt{\int_{\theta_0}^{\theta} |u_0'(\varphi)|^2\,d\varphi}\,\sqrt{\int_{\theta_0}^{\theta} 1^2\,d\varphi} \le \sqrt{2\pi}\,\|u_0\|_{W^{1,2}}, \tag{126}$$

where the first inequality follows by the the Cauchy–Schwarz inequality. This yields the estimate $\|u\|_{C^0} \le C\|u\|_{W^{1,2}}$, which in turn shows that there is a continuous embedding $W^{1,2}(S^1) \subset C^0(S^1)$.

Furthermore, by tracing through (126) it is easy to see that

$$|u(\theta_1) - u(\theta_2)| \le \sqrt{2\pi}\,\|u\|_{W^{1,2}}\ \mathrm{dist}(\theta_1, \theta_2)^{\frac12}.$$

Hence, if u_k is any sequence bounded in $W^{1,2}(S^1)$, then this sequence consists of uniformly bounded and equicontinuous functions on S^1. By the Arzela–Ascoli theorem, this sequence has a convergent subsequence in $C^0(S^1)$, thus proving the compactness of the embedding $W^{1,2}(S^1) \subset L^p(S^1)$ for any p.

5.2 Elliptic operators

A map $L\colon C^\infty(\Omega; \mathbb{R}^r) \to C^\infty(\Omega; \mathbb{R}^s)$, where $\Omega \subset \mathbb{R}^n$ is an open subset, is said to be a linear differential operator of order ℓ if L can be expressed in the form

$$Lf = \sum_{|\alpha|\le \ell} A_\alpha(x)\frac{\partial^{|\alpha|}}{\partial x^\alpha} f, \tag{127}$$

where $\alpha = (\alpha_1, \dots, \alpha_n),\ \alpha_i \in \mathbb{Z}_{\ge 0}$, is a multi-index, $|\alpha| := \sum \alpha_i$, and $A_\alpha \in C^\infty\big(\Omega; \mathrm{Hom}(\mathbb{R}^r; \mathbb{R}^s)\big)$. For example, in the case of operator of order $\ell = 2$ acting on functions of two variables, we have

$$Lf = A_{20}\frac{\partial^2}{\partial x_1^2} + A_{11}\frac{\partial^2}{\partial x_1 \partial x_2} + A_{02}\frac{\partial^2}{\partial x_2^2} + A_{10}\frac{\partial}{\partial x_1} + A_{01}\frac{\partial}{\partial x_2} + A_{00},$$

where A_{ij} are smooth functions on Ω.

It is intuitively clear that the highest order terms determine some essential properties of L. Thus, we say that

$$\sigma_L(\xi) := \sum_{|\alpha|=\ell} A_\alpha(x)\xi^\alpha = \sum_{|\alpha|=\ell} A_\alpha(x)\xi_1^{\alpha_1}\cdot\ldots\cdot\xi_n^{\alpha_n} \qquad \text{where } \xi \in (\mathbb{R}^n)^* \cong \mathbb{R}^n$$

is *the principal symbol* of L. The symbol can be conveniently viewed as a map $\Omega \times \mathbb{R}^n \to \mathrm{Hom}(\mathbb{R}^r; \mathbb{R}^s)$, which is polynomial in the ξ-variable.

Definition 128. A linear differential operator L is called *elliptic*, if for all $x \in \Omega$ and all $\xi \in \mathbb{R}^n,\ \xi \ne 0$, we have: $\sigma_L(\xi)$ is an invertible homomorphism.

Notice in particular that we must require $r = s$ to have an elliptic operator.

Example 129. For the (non-negative) Laplacian on $\mathbb{R}^n$ acting on functions we have

$$\sigma_\Delta(\xi) = -\sum_{i=1}^n \xi_i^2 = -|\xi|^2.$$

Hence, Δ is an elliptic operator. This is in fact a prototypical example of an elliptic operator.

The concepts above make sense for a more general setting of vector bundles. To spell some details, let E and F be vector bundles over a manifold M of rank r and s respectively. We say that a map $L\colon \Gamma(E) \to \Gamma(F)$ is a linear differential operator of order ℓ, if for any choice of local coordinates on $\Omega \subset M$ and any trivializations of $E|_\Omega$ and $F|_\Omega$ the map L can be represented as in (127), where the coefficients A_α are allowed to depend on the choices made.

Let $\pi\colon T^*M \to M$ be the natural projection. A straightforward computation shows that the symbol makes sense as a section of $\mathrm{Hom}(\pi^*E;\, \pi^*F)$. Then L is said to be elliptic if the symbol is a pointwise invertible homomorphism away from the zero section of T^*M.

Example 130. Let M be an oriented Riemannian manifold. Consider the Laplace–Beltrami operator acting on the space of functions on M:

$$\Delta f = - * d * df,$$

where $*\colon \Lambda^k T^*M \to \Lambda^{n-k}T^*M$ is the Hodge operator.

Choose local coordinates $(x_1, \dots, x_n)$ and write $g = g_{ij}\, dx_i \otimes dx_j$, $|g| = \det(g_{ij})$. Denoting by (g^{ij}) the inverse matrix, a straightforward computation yields the local form of the Laplace–Beltrami operator:

$$\Delta f = -|g|^{-\frac{1}{2}} \sum_{i,j=1}^n \frac{\partial}{dx_i}\Big(|g|^{\frac{1}{2}} g^{ij} \frac{\partial f}{\partial x_j}\Big)$$

From this it is easy to compute the symbol. Indeed, if $\xi = \sum \xi_i\, dx_i$, then

$$\sigma_\Delta(\xi) = -|g|^{-\frac{1}{2}} \sum_{i,j=1}^n \xi_i |g|^{\frac{1}{2}} g^{ij} \xi_j = -|\xi|^2,$$

where $|\cdot|$ denotes the norm on T^*X. In particular, the Laplace–Beltrami operator is elliptic.

Example 131. For any Dirac bundle E the corresponding Dirac operator $\mathcal{D}\colon \Gamma(E) \to \Gamma(E)$ is elliptic. Indeed, it is easy to see that the principal symbol $\sigma_\xi(D)$ is just the Clifford multiplication with ξ. This is invertible with the inverse given by the Clifford multiplication with $|\xi|^{-2}\xi$.

In dimension four (and in fact in any even dimension) the chiral Dirac operators $\not{D}^{\pm}$ are also elliptic.

Any linear elliptic operator $L\colon C^\infty(M;\, E) \to C^\infty(M;\, F)$ of order ℓ extends as a bounded linear map

$$L\colon W^{k+\ell,p}(M;\, E) \to W^{k,p}(M;\, F) \tag{132}$$

for any $k \geq 0$ and any $p > 1$.

Theorem 133 (Elliptic estimate). *Let M be a compact manifold. For any linear elliptic operator L there is a constant $C > 0$ with the following property: If $Ls \in W^{k,p}(M;\, F)$, then $s \in W^{k+\ell,p}(M;\, E)$ and*

$$\|s\|_{W^{k+\ell,p}} \leq C\big(\|Ls\|_{W^{k,p}} + \|s\|_{L^p}\big).$$

Here C depends on k and p but not on s. □

Assume that both E and F are equipped with an euclidean structure. An operator $L^* \colon C^\infty(M;\, F) \to C^\infty(M;\, E)$ is said to be formal adjoint of L if

$$\langle Ls, t\rangle_{L^2} = \langle s, L^* t\rangle_{L^2} \tag{134}$$

holds for any $s \in C^\infty(M;\, E)$ and any $t \in C^\infty(M;\, F)$. One can show that L^* exists and is a linear differential operator of order ℓ. Moreover, L is elliptic if and only if L^* is elliptic.

One of the most important results in the theory of elliptic differential operators is the following.

Theorem 135 (Fredholm alternative). *Let L be elliptic, M compact, and $t \in C^\infty(M;\, F)$. The equation*

$$Ls = t$$

has a smooth solution if and only if $t \in \ker L^*$. □

Remark 136. A corollary of Theorem 135 can be formulated as follows. Under the hypotheses of Theorem 135 one and only one of the following statements hold:

(i) The homogeneous equation $L^* s = 0$ has a non-trivial solution.

(ii) The inhomogeneous equation $Ls = t$ has a unique solution for any smooth t.

This form of Theorem 135 is widely used in the theory of PDEs.

Definition 137. A bounded linear map $B \colon X \to Y$ between two Banach spaces is called Fredholm, if the following conditions hold:

(a) $\dim \ker B < \infty$;

(b) $\operatorname{Im} B$ is a closed subspace of Y;

(c) $\operatorname{coker} B := Y / \operatorname{Im} B < \infty$.

If B is Fredholm, the integer

$$\operatorname{index} B := \dim \ker B - \dim \operatorname{coker} B$$

is called the index of B.

Remark 138. One can show that (b) follows the other two conditions. Nevertheless, it will be useful to know that $\operatorname{Im} B$ is closed, even if one does not necessarily need to check this.

For example, any linear map between finite dimensional spaces is Fredholm and its index equals $\dim Y - \dim X$. In fact, Fredholm operators resemble linear maps between finite dimensional vector spaces well known from the basic course of linear algebra and this largely explains the importance of Fredholm operators for us.

Exercise 139. Let B_0 be a Fredholm operator. Show that there is an $\varepsilon > 0$ such that any bounded operator B with

$$\|B - B_0\| < \varepsilon$$

is also Fredholm. Here $\|\cdot\|$ means the operator norm in the space of bounded linear maps $X \to Y$. Prove also that $\operatorname{index} B = \operatorname{index} B_0$.

Theorem 140. *For any elliptic operator L of order $\ell > 0$ on a compact manifold M, (132) is a Fredholm operator. Moreover, $\ker L$ consists of smooth sections only.*

Sketch of proof. If $s \in W^{\ell,p}$ is in the kernel of L, then $s \in W^{k,p}$ for any $k \geq 0$ by Theorem 133. Hence, s is smooth by Theorem 123.

Let $s_j \in W^{\ell,p}$ be any sequence such that $Ls_j = 0$ and $\|s_j\|_{W^{\ell,p}} \leq 1$. Then the sequence $\|s_j\|_{W^{\ell+1,p}}$ is bounded by Theorem 133. Hence, by Theorem 123 after passing to a subsequence if necessary, s_j converges to some limit s_∞ in $W^{\ell,p}$. Since $L\colon W^{\ell,p} \to L^p$ is bounded, $s_\infty \in \ker L$. In other words, the unit ball in $\ker L$ is compact with respect to the $W^{\ell,p}$ norm. This implies that $\ker L$ is finite dimensional.

Next, let us assume that $p = 2$, which simplifies the discussion somewhat. Denote $V := \big(\operatorname{Im} L\colon W^{k+\ell,2} \to W^{k,2}\big)^{\perp}$, where the orthogonal complement is understood in the sense of L^2–scalar product. One can show that all sections lying in V are in fact in $W^{\ell,p}$. Then (134) implies that $V = \ker L^*$. This in turn yields that V is finite dimensional (and consists of smooth sections only). □

5.3 Elliptic complexes

Fix a manifold M and a sequence of vector bundles $E_1, \dots, E_k$, which I assume to be finite for the simplicity of exposition. Let

$$0 \to \Gamma(E_1) \xrightarrow{L_1} \Gamma(E_2) \xrightarrow{L_2} \dots \xrightarrow{L_{k-1}} \Gamma(E_k) \to 0 \tag{141}$$

be a sequence of differential operators such that $L_j \circ L_{j-1} = 0$ for all integer $j \in [1, k-1]$, i.e., (141) is a complex. In particular, we can define the corresponding cohomology groups

$$H^j(E) := \ker L_j / \operatorname{im} L_{j-1},$$

where $j \in \{1, 2, \dots, k\}$.

Associated to (141) is the sequence of principal symbols:

$$0 \to \pi^*E_1 \xrightarrow{\sigma_{L_1}} \pi^*E_2 \xrightarrow{\sigma_{L_2}} \pi^*E_3 \xrightarrow{\sigma_{L_3}} \dots \xrightarrow{\sigma_{L_{k-1}}} \pi^*E_k \to 0.$$

If this is exact on the complement of the zero section, (141) is called an elliptic complex. For example, a very short complex $0 \to \Gamma(E) \xrightarrow{L} \Gamma(F) \to 0$ is elliptic if and only if L is elliptic.

With these preliminaries at hand, we can construct the associated Laplacians:

$$\Delta = \Delta_j\colon \Gamma(E_j) \to \Gamma(E_j), \qquad \Delta_j = L_j^* L_j + L_{j-1} L_{j-1}^*.$$

Exercise 142. Prove that each Δ_j is an elliptic operator provided the initial complex is elliptic.

In the sequel, I will drop the index of the differentials L_i to simplify the notations so that all maps in (141) are denoted by the same symbol L. I assume also that all differentials L are of order 1, since only this case will show up in these notes, but operators of other orders could be considered as well. Also, I assume that all bundles E_i are equipped with an Euclidean structure.

Denote

$$\mathcal{H}_j(E) := \big\{ s \in \Gamma(E_j) \mid \Delta s = 0 \big\}.$$

Elements of $\mathcal{H}_j(E)$ are called harmonic sections.

Theorem 143. *For an elliptic complex on a compact manifold the following holds:*

(i) *Each $\mathcal{H}_j(E)$ is a finite dimensional vector space.*

(ii) *$s \in \mathcal{H}_j(E)$ if and only if $Ls = 0$ and $L^* s = 0$.*

(iii) *The natural homomorphism*

$$\mathcal{H}_j(E) \to H^j(E), \qquad s \mapsto [s]$$

is an isomorphism.

Proof. The first statement follows from the ellipticity of Δ. The second statement follows easily from the equality

$$\langle \Delta s, s \rangle_{L^2} = \big\langle (LL^* + L^*L)s, s \big\rangle_{L^2} = \|L^* s\|^2_{L^2} + \|Ls\|^2_{L^2}.$$

It remains to prove the last claim. Thus pick any $s_0 \in \Gamma(E_j)$ such that $Ls_0 = 0$. We wish to show that the equation

$$(L + L^*)\big(s_0 + Lt\big) = 0$$

has a solution $t \in \Gamma(E_{j-1})$. Notice that this equation is equivalent to

$$L^* L t = -L^* s_0. \tag{144}$$

Consider instead the equation $\Delta t = -L^* s_0$, whose right hand side is clearly L^2-orthogonal to $\ker \Delta^* = \ker \Delta$. Hence, by Theorem 135 there is a unique solution of $\Delta t = -L^* s_0$, which can be rewritten as

$$L(L^* t) + L^*(s_0 + Lt) = 0.$$

A moment's thought shows that $\operatorname{Im} L$ is L^2-orthogonal to $\operatorname{Im} L^*$. Hence, t is a solution of (144) in fact. This finishes the proof of the existence part.

The uniqueness is easy to show: If $s_1, s_2 \in \mathcal{H}_j(E)$ are such that $s_1 - s_2 = Lt$ for some $t \in \Gamma(E_{j-1})$, then

$$\|Lt\|^2_{L^2} = \langle L^* L t, t \rangle_{L^2} = \langle L^*(s_1 - s_2), t \rangle_{L^2} = 0.$$

□

A refinement of the argument in the proof of the above theorem shows that in fact we have the following decomposition

$$\Gamma(E_j) = \operatorname{Im} L \oplus \mathcal{H}_j(E) \oplus \operatorname{Im} L^*, \tag{145}$$

which is L^2–orthogonal. Details can be found for example in [Wel80].

Exercise 146. Show that the short complex

$$0 \to \Gamma(E_1) \xrightarrow{L_1} \Gamma(E_2) \xrightarrow{L_2} \Gamma(E_3) \to 0$$

is elliptic if and only if the operator

$$(L_2, L_1^*)\colon \Gamma(E_2) \to \Gamma(E_3 \oplus E_1)$$

is elliptic.

5.3.1 A gauge-theoretic interpretation

The cohomology group $H^j(E)$ is an example of a linear gauge–theoretic moduli space, cf. Section 1. Let me spell some details. The manifold[4] $\mathcal{B} = \Gamma(E_j)$ carries an action of the additive group $\mathcal{G} := \Gamma(E_{j-1})$ by translations: $(s,t) \mapsto s + Lt$. If $\Gamma(E_{j+1})$ is viewed as a trivial $\mathcal{G}$–representation, $L\colon \Gamma(E_j) \to \Gamma(E_{j+1})$ is a $\mathcal{G}$-equivariant map, which in this particular case just means that L is $\mathcal{G}$-invariant. The action of the gauge group on $L^{-1}(0)$ is not free in general, however $\ker L$ acts trivially so that the corresponding 'moduli space' $L^{-1}(0)/\mathcal{G} = H^j(E)$ is a finite dimensional manifold (a vector space, in fact) provided the complex is elliptic and the base manifold is compact. This explains our interest in the theory of elliptic operators.

Notice, however, that because of the linear setting we can not expect that the 'moduli space' $H^j(E)$ will be compact. The reason is that $H^j(E)$ inherits the action of $\mathbb{R}_{>0}$ by dilations and if we take the quotient of $H^j \setminus \{0\}$ by this action the resulting space is compact. Thus, in this setting $\dim H^j(E) < \infty$ is a suitable replacement for the compactness of the moduli space.

One more important feature we have seen is the so called gauge fixing. Namely, we have shown that for each point $s \in L^{-1}(0)$ there is a unique representative $h(s)$ in the 'gauge-equivalence class of s' such that $h(s)$ is harmonic. Moreover, the map

$$L^{-1}(0) \to \mathcal{H}_j(E), \qquad s \mapsto h(s)$$

induces a diffeomorphism $L^{-1}(0)/\mathcal{G} \to \mathcal{H}^j(E)$.

Furthermore, an isomorphism class of a finite dimensional vector space is determined by a unique non-positive integer, namely its dimension. Hence, $b_j(E) := \dim H^j(E)$ is an 'invariant' of E. In many cases, these invariants capture a subtle information about the underlying manifold M.

The example of the de Rham complex in the following subsection will make these constructions more concrete.

5.3.2 The de Rham complex

Recall that for any manifold M of dimension n we have the de Rham complex

$$0 \to \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \longrightarrow \dots \longrightarrow \Omega^n(M) \to 0.$$

[4] One can consider $\Gamma(E_j)$ as a Banach manifold by taking a Sobolev completion. Since this is not really important at this point, I describe the setting in the smooth category.

Exercise 147. Show that the de Rham complex is elliptic.

If M is oriented and Riemannian, we have the Hodge operator $*\colon \Lambda^j T^*M \to \Lambda^{n-j}T^*M$. Using this, the formal adjoint of the exterior derivative can be expressed as

$$d^* = (-1)^{n(j-1)+1} * d *\colon \Omega^j(M) \to \Omega^{j-1}(M)$$

so that the Hodge–de Rham Laplacian is $\Delta = dd^* + d^*d$. If M is in addition compact, the space $\mathcal{H}_j$ of harmonic forms of degree j is naturally isomorphic to the jth de Rham cohomology group. Then $b_j(M) := \dim \mathcal{H}_j$ is the jth Betti number of M. This is a topological invariant of M even though in this approach a smooth structure has been used.

6 Fredholm maps

6.1 The Kuranishi model and the Sard–Smale theorem

Let X and Y be Banach manifolds.

Definition 148. A map $F \in C^\infty(X; Y)$ is said to be Fredholm if the differential dF is a Fredholm linear map at each point.

Hence, for each $x \in X$ the index of d_xF is well defined. If X is connected, then $\operatorname{index} d_xF$ does not depend on x and this common value is denoted by $\operatorname{index} F$.

Fredholm maps have a lot in common with smooth maps between finite dimensional manifolds. A manifestation of this is the following result.

Theorem 149 (Kuranishi model)**.** *Let X and Y be Banach spaces and $F\colon X \to Y$ a Fredholm map. Pick a point $p \in F^{-1}(0)$ and denote $X_0 = \ker d_pF$, $\quad Y_0 := \operatorname{im} d_pF$. Furthermore, choose subspaces $X_1 \subset X$ and $Y_1 \subset Y$ such that*

$$X = X_0 \oplus X_1 \qquad \textit{and} \qquad Y = Y_0 \oplus Y_1.$$

Then there is a diffeomorphism φ of a neighborhood of the origin in X onto a neighbourhood of p such that $\varphi(0) = p$, a linear isomorphism $T\colon X_1 \to Y_0$, and a smooth map $f\colon X \to Y_1$ such that

$$F \circ \varphi(x_0, x_1) = Tx_1 + f(x_0, x_1)$$

for all $(x_0, x_1) \in X_0 \oplus X_1$ in the neighborhood of the origin.

In particular, if $f_0\colon X_0 \to Y_1$ denotes the restriction of f to X_0, then a neighborhood of p in $F^{-1}(0)$ is homeomorphic to a neighborhood of the origin in $f_0^{-1}(0)$.

Proof.

To be added.

□

Corollary 150. *Assume the hypotheses of Theorem 149. Suppose also that 0 is a regular value of F, i.e., the d_pF is surjective for all $p \in F^{-1}(0)$. Then $F^{-1}(0)$ is a smooth manifold of dimension* $\operatorname{index} F$.

Proof. Since $\operatorname{Im} d_pF = Y$, we necessarily have $Y_1 = \{0\}$ so that f_0 is a constant map. Hence, $F^{-1}(0)$ is diffeomorphic to a neighborhood of the origin in X_0, i.e., a manifold of dimension $\dim X_0 = \operatorname{index} F$. □

For a smooth map between finite dimensional manifolds, almost any value is regular by Sard's theorem [BT03, Thm. 9.5.4]. There is a generalization of this statement for an infinite dimensional setting due to Smale. This is commonly known as the Sard–Smale theorem.

Recall that a subset A of a topological space is said to be of *second category*, if A can be represented as a countable intersection of open dense subsets. If the underlying topological space is a Banach manifold, then a subset of second category is dense.

Theorem 151. *Let F be a smooth Fredholm map between paracompact Banach manifolds. Then the set of regular values of F is of second category, in particular dense.* □

Let $Z \subset Y$ be a smoothly embedded finite dimensional submanifold. A map F is said to be *transverse* to Z, if for any $z \in Z$ and any $x \in F^{-1}(Z)$ the following holds:

$$\operatorname{Im} d_xF + T_zZ = T_zY.$$

In particular, F is transverse to $Z = \{z\}$ is and only if z is a regular value of F. It is well-known that the notion of transversality is a useful generalization of the notion of regular value, see for instance [GP10].

Just as in the finite dimensional case we have the following result.

Theorem 152. *Let $Z \subset Y$ be a smoothly embedded finite dimensional submanifold. If F is transverse to Z, then $F^{-1}(Z)$ is a smooth submanifold of X and*

$$\dim F^{-1}(Z) = \operatorname{index} F + \dim W. \qquad \square$$

6.2 The $\mathbb{Z}/2\mathbb{Z}$ degree

Recall that a map $F\colon X \to Y$ is called *proper* if preimages of compact subsets are compact.

Let $F\colon X \to Y$ be a proper Fredholm map between (paracompact) Banach manifolds of index zero, where Y is connected. Then for any regular value $y \in Y$ the preimage $F^{-1}(y)$ is a compact manifold of dimension zero, hence a finite number of points. The number

$$\deg_2 F := \# f^{-1}(y) \mod 2$$

is called the $\mathbb{Z}/2\mathbb{Z}$ degree of F.

Theorem 153.

(i) $\deg_2 F$ *does not depend on the choice of the regular value* y*;*

(ii) If F_0 and F_1 are homotopic within the class of proper Fredholm maps of vanishing index, then $\deg_2 F_0 = \deg_2 F_1$.

Proof. The proof requires a number of steps.

Step 1. *If F is a proper Fredholm map, then the set of regular values of F is open and dense.*

First notice that the set of critical points $\mathrm{Crit}(F)$ is closed. Since any proper map is closed, the set of critical values $F(\mathrm{Crit}(F))$ is closed.

Step 2. *The function*

$$y \mapsto \#F^{-1}(y) \mod 2$$

is locally constant on the set of regular values.

Let y be a regular point of F, $F^{-1}(y) = \{x_1, \dots, x_k\}$. By the inverse function theorem, there is a neighborhood U of y and a neighborhood V_j of x_j such that $F\colon V_j \to U$ is a diffeomorphism. Then, for any $y' \in U$ we have $\#f^{-1}(y') = k$, hence the claim.

Step 3. *Let F_0 and F_1 be homotopic so that the homotopy is within the class of proper Fredholm maps. Then*

$$\#F_0^{-1}(y) = \#F_1^{-1}(y) \mod 2 \tag{154}$$

for any y, which is a regular value for both F_0 and F_1.

Let F_t, $t \in [0,1]$, be a homotopy. Let me assume first that, y is a regular value for $F\colon X \times [0,1] \to Y$. Then $F^{-1}(y)$ is a 1-dimensional manifold with boundary such that $\partial F^{-1}(y) = F_0^{-1}(y) \cup F_1^{-1}(y)$. Hence, (154) holds.

If y is not a regular value of F, then we can choose y_1 arbitrarily close to y such that y_1 is a regular value for any of F_0, F_1, F. The conclusion of this step follows by Step 2.

Step 4. *Let x be an arbitrary point in the unit ball of a Banach space B. Then there is a diffeomorphism φ of B such that the following holds: $\varphi(0) = x$, φ is the identity map on the complement of the Ball of radius 2, and φ is homotopic to the identity map relative to the complement of the ball of radius 2.*

This is really a finite dimensional statement. Indeed, choose decomposition $B = V \oplus V'$, where V is finite dimensional and $x \in V$. By [Mil97, P. 22] there is a 1-parameter family $\psi_t\colon V \to V$ of diffeomorphisms such that $\psi_1(0) = x$, $\psi_0 = \mathrm{id}_V$, and ψ_t is identity outside the unit ball in V.

Choose a smooth function $\chi\colon \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ such that $\chi(0) = 1$ and $\chi(t) = 0$ for $t \ge 1$. Then

$$\varphi_t(v, v') := \psi_{t\chi(|v'|)}(v) + v'$$

is a family of diffeomorphisms such that $\varphi_0 = \mathrm{id}$ and $\varphi_1(0,0) = \psi_1(0) = x$. Clearly, for any t we have $\|\varphi_t(v, v')\| \le 2$ by the triangle inequality.

Let us check that φ_t is identity on the complement of the ball of radius 2. Thus, pick any (v, v') such that $\|v\| + \|v'\| \ge 2$. Then either $\|v\| \ge 1$ or $\|v'\| \ge 1$. The first case is clear and the second one we have $\varphi_t(v, v') = \psi_0(v) + v' = v + v'$ since $\chi(\|v'\|) = 0$ by the construction.

Step 5. *If Y is a connected Banach manifold, then for any $y_1, y_2 \in Y$ there is a diffeomorphism $\varphi \in \mathrm{Diff}_0(Y)$ such that $\varphi(y_1) = y_2$, where $\mathrm{Diff}_0(Y)$ denotes the subgroup of all diffeomorphisms homotopic to the identity.*

Fix $y_2 \in Y$ and denote

$$C(y_2) := \big\{ y_1 \in Y \mid \exists \varphi \in \mathrm{Diff}_0(Y) \text{ such that } \varphi(y_1) = y_2 \big\}$$

By the previous step, $C(y_2)$ is open and non-empty, hence $C(y_2) = Y$.

Step 6. *We prove the claim of this theorem.*

Let F be as in the statement of this theorem. Pick any regular values y_1 and y_2 and $\varphi \in \mathrm{Diff}_0(Y)$ such that $\varphi(y_1) = y_2$. Then y_2 is a common regular value for F and $\varphi \circ F$. Since these maps are homotopic, by Step 3 we have

$$F^{-1}(y_2) = (\varphi \circ F)^{-1}(y_2) = F^{-1}(y_1) \mod 2.$$

This proves *(i)*.

Furthermore, if $F\colon X \times [0,1] \to Y$ is a homotopy between F_0 and F_1, we can choose $y \in Y$ which is a regular value for any of the following maps: F_0, F_1, and F. This implies the second assertion. □

Corollary 155. *Let $F\colon X \to Y$ be a proper Fredholm map between (paracompact) Banach manifolds of index zero, where Y is connected. If $\deg_2 F \neq 0$, then F is surjective.* □

6.3 The parametric transversality

Given a 'random' point y in the target, there is no reason to expect that the preimage $F^{-1}(y)$ will be a submanifold. Of course, if F is Fredholm the theorem of Sard–Smale implies that we can choose y' arbitrarily close to y so that $F^{-1}(y')$ is a manifold indeed. In practice, however, it is not always the case that one has a freedom to choose a point in the target. This is typically the case, which will be considered in some detail below, if X and Y are equipped with an action of a Lie group $\mathcal{G}$ and F is $\mathcal{G}$–equivariant. In this case it is natural to choose y as a fixed point of the $\mathcal{G}$–action on Y so that $F^{-1}(y)$ inherits a $\mathcal{G}$–action. This, however, restricts severely possible choices of y so that the Sard–Smale theorem is not applicable in a straightforward manner.

One way to deal with this problem is as follows. Assume there is a connected Banach manifold W and a smooth Fredholm map $\mathcal{F}\colon X \times W \to Y$ with the following properties:

(A) For each $w \in W$ the map $\mathcal{F}_w\colon X \times \{w\} \to Y$ is Fredholm;

(B) There is a point $w_0 \in W$ such that $\mathcal{F}_{w_0} = F$;

(C) y is a regular value of $\mathcal{F}$;

Typically, it is not too hard to construct a map $\mathcal{F}$ satisfying these properties.

Since y is a regular value, $\mathcal{F}^{-1}(y) \subset X \times W$ is a smooth Banach submanifold. We have a natural projection $\pi\colon \mathcal{F}^{-1}(y) \to W$, which is just the restriction of the projection $X \times W \to W$.

Lemma 156.

(i) *π is a Fredholm map and* $\operatorname{index} \pi = \operatorname{index} F$;

(ii) *There is a subset $W_0 \subset W$ of second category such that y is a regular value for $\mathcal{F}_w$ for all $w \in W_0$.*

Proof.

To be inserted.

□

Hence, by the Sard–Smale theorem, there is w arbitrarily close to w_0 such that

$$\pi^{-1}(w) = \mathcal{F}_w^{-1}(y)$$

is a smooth submanifold of dimension $\operatorname{index} F$.

Suppose that in addition to hypotheses (A)–(B), the following holds:

(D) Each map $\mathcal{F}_w$ is proper.

In what follows I will also assume that $\operatorname{index} F = 0$, since this is the setting where the degree of a Fredholm map was defined. However, this is by no means essential.

Thus, if (A)–(D) holds, $\mathcal{F}_w^{-1}(y) = \pi^{-1}(w)$ consists of finitely many points. Since W is connected by assumption, there is a path connecting w_0 and w so that $\mathcal{F}_w$ is homotopic to $\mathcal{F}_{w_0} = F$. Hence,

$$\deg_2 F = \#\mathcal{F}_w^{-1}(y) \mod 2.$$

In particular, $\#\mathcal{F}_w^{-1}(y) \mod 2$ does not depend on w.

Of course, this conclusion is pretty much straightforward thanks to the facts we have established in the preceding subsection. The point is that the number $\#\mathcal{F}_w^{-1}(y) \mod 2$ could be taken as a definition of the degree of F thus omitting the construction of Section 6.2. For instance, this is commonly used in the equivariant setting as discussed at the beginning of this section.

Let me describe this alternative construction of the degree of a proper Fredholm map in some detail. Thus, choose any $w_1, w_2 \in W_0$ so that y is a regular value for $\mathcal{F}_j := \mathcal{F}_{w_j}$. For a fixed $k \geq 0$ denote

$$\Gamma(w_1, w_2) := \Big\{\gamma \in C^k\big([1,2]; W\big) \mid \gamma(1) = w_1 \text{ and } \gamma(2) = w_2\Big\},$$

which is a Banach manifold. Consider the map

$$\mathbf{F}\colon X \times \Gamma(w_1, w_2) \times [1,2] \to Y, \qquad \mathbf{F}(x, \gamma, t) := \mathcal{F}(x, \gamma(t)).$$

Clearly, y is a regular value for $\mathbf{F}$ too so that $\mathbf{F}^{-1}(y)$ is a submanifold of $X \times \Gamma(w_1, w_2) \times [1,2]$. Applying the Sard–Smale theorem to the restriction of the projection $X \times \Gamma(w_1, w_2) \times [1,2] \to \Gamma(w_1, w_2)$ we obtain that for a generic $\gamma \in \Gamma(w_1, w_2)$ the subset

$$\mathcal{M}_\gamma := \big\{(x,t) \in X \times [1,2] \mid \mathbf{F}(x, \gamma, t) = y\big\}$$

is a smooth submanifold of dimension 1 and

$$\partial \mathcal{M}_\gamma = \mathcal{F}_1^{-1}(y) \cup \mathcal{F}_2^{-1}(y).$$

Hence,

$$\#\mathcal{F}_1^{-1}(y) = \#\mathcal{F}_2^{-1}(y) \mod 2$$

so that the number $\#\mathcal{F}_w^{-1}(y) \mod 2$ does not depend on $w \in W_0$. This common value, as we already know, is just $\deg_2(F)$.

Figure ?

6.4 The determinant line bundle

If X and Y are closed *oriented* manifolds of the same finite dimension, for any map $F\colon X \to Y$ the degree can be extended to take values in $\mathbb{Z}$ rather than $\mathbb{Z}/2\mathbb{Z}$ [GP10]. This generalization requires orientations of the background manifolds and this tool is not readily available in infinite dimensions. In this section I will describe how to deal with this problem.

Let $\mathcal{P}$ be a topological space and $\{T_p \mid p \in \mathcal{P}\}$ be a continuous family of linear Fredholm maps. This means that the map

$$\mathcal{P} \to \mathrm{Fred}(X;Y), \qquad p \mapsto T_p$$

is continuous with respect to the operator norm, where X, Y are Banach spaces and $\mathrm{Fred}(X;Y)$ denotes the subspace of linear Fredholm maps.

Pick a point $p \in \mathcal{P}$. Since both the kernel and cokernel of T_p are finite dimensional, we can construct the (real) line

$$\det T_p := \Lambda^{\mathrm{top}} \ker T_p \otimes \Lambda^{\mathrm{top}} (\operatorname{coker} T_p)^*.$$

As p varies, we obtain a family of vector spaces of dimension one. It turns out that this family is actually a locally trivial vector bundle, which is somewhat non-obvious given that the dimensions of the kernel as well as cokernel may 'jump' as p varies.

To understand why this is the case, pick a point p_0 and a finite dimensional subspace $V \subset Y$ transverse to $\operatorname{Im} T_{p_0}$. The linear map

$$T_{p,V}\colon X \oplus V \to Y, \qquad T_{p,V}(x,v) = T_p x + v$$

is surjective for $p = p_0$ and, hence, also for all p sufficiently close to p_0. Therefore, we have the exact sequence

$$0 \to \ker T_p \to \ker T_{p,V} \to V \to V/\operatorname{Im} T_p \cap V \to 0 \tag{157}$$

Notice that $\operatorname{coker} T_p = (\operatorname{Im} T_p + V)/\operatorname{Im} T_p = V/\operatorname{Im} T_p \cap V$.

Exercise 158. Let $0 \to U_0 \to U_1 \to U_2 \to 0$ be a short exact sequence of real vector spaces. Show that the 'inner product map'

$$\Lambda^{\mathrm{top}} U_0 \otimes \Lambda^{\mathrm{top}} U_1^* \to \Lambda^{\mathrm{top}} (U_1/U_0)^* = \Lambda^{\mathrm{top}} U_2^*$$

induces an isomorphism $\Lambda^{\mathrm{top}} U_1 \cong \Lambda^{\mathrm{top}} U_0 \otimes \Lambda^{\mathrm{top}} U_2$.

More generally, show that for any exact sequence of real vector spaces $0 \to U_0 \to U_1 \to U_2 \to \cdots \to U_k \to 0$ there is a canonical isomorphism

$$\bigotimes \Lambda^{\mathrm{top}} U_{\mathrm{even}} \cong \bigotimes \Lambda^{\mathrm{top}} U_{\mathrm{odd}}.$$

Hence, by (157) we have a canonical isomorphism

$$\det T_p \cong \Lambda^{\mathrm{top}} \ker T_{p,V} \otimes \Lambda^{\mathrm{top}} V^*.$$

It can be shown not only that $\dim \ker T_{p,V}$ is constant in p near p_0 but also that $\ker T_{p,V}$ is trivial in a neighborhood W of p_0. Details can be found for instance in [MS12, Thm. A.2.2]. Thus, the family of real lines $\{\det T_p \mid p \in W\}$ admits a trivialization, i.e., $\det T$ is a vector bundle.

Definition 159. Two continuous families of Fredholm operators $\{T_{p,i} \mid p \in \mathcal{P}\},\ i \in \{0,1\}$ are said to be *homotopic*, if there is a continuous family of Fredholm operators $\{T_{p,t} \mid (p,t) \in \mathcal{P} \times [0,1]\}$, whose restriction to $\mathcal{P} \times \{0\}$ and $\mathcal{P} \times \{1\}$ yields the initial families.

Assume that $\{T_{p,0}\}$ is homotopic to $\{T_{p,1}\}$ and the determinant line bundle $\det T_{p,1}$ is trivial. Since $\mathcal{P} \times [0,1]$ is homotopy equivalent to $\mathcal{P}$, the bundle $\det T_{p,t}$ is also trivial. In particular, $\det T_{p,0}$ is trivial. More precisely, we have the following statement.

Proposition 160. *Let $\{T_{p,0}\}$ and $\{T_{p,1}\}$ be homotopic families of Fredholm operators such that $\det T_{p,1}$ is trivial. Then $\det T_{p,0}$ is also trivial. Moreover, a choice of a trivialization of $\det T_{p,1}$ and a homotopy $\{T_{p,t} \mid t \in [0,1]\}$ yields a trivialization of $\det T_{p,0}$. This is unique up to a multiplication with a positive function.* □

6.5 Orientations and the $\mathbb{Z}$–valued degree

Let $X, Y,$ and W be Banach manifolds. Let $F \colon X \times W \to Y$ be a smooth map such that the following holds:

(a) $F_w = F|_{X \times \{w\}} \colon X \to Y$ is a Fredholm map of index d for each $w \in W$;

(b) $y \in Y$ is a regular value of F;

(c) $F_w^{-1}(y)$ is compact for any $w \in W$;

(d) $\det d_x F_w$ is trivial over $X \times W$.

Furthermore, I assume that a trivialization of the determinant line bundle has been fixed.

It should suffice that $\det d_x F_w$ is trivial over $F^{-1}(y)$.

Pick any $w \in W$ such that y is a regular value of F_w. Then for any $x \in F_w^{-1}(y)$ we have

$$\det d_x F_w = \Lambda^{\mathrm{top}} \ker d_x F_w = \Lambda^{\mathrm{top}} T_x F_w^{-1}(y)$$

so that the hypotheses above imply that $\mathcal{M}_w := F_w^{-1}(y)$ is an oriented d–manifold.

Furthermore, let $\gamma \colon [0,1] \to W$ be a path connecting w_0 and w_1 akin to the situation considered in Section 6.3. For generic γ the space

$$\mathcal{M}_\gamma := \big\{(x,t) \in X \times [0,1] \mid F_{\gamma(t)}(x) = y\big\}$$

is an oriented manifold of dimension $d+1$ such that

$$\partial \mathcal{M}_\gamma = \mathcal{M}_{w_1} \sqcup \overline{\mathcal{M}}_{w_0}, \tag{161}$$

where $\overline{\mathcal{M}}_{w_0}$ means that the orientation of $\mathcal{M}_{w_0}$ is reversed.

In the particular case $d = 0$, $\mathcal{M}_w$ is just a finite collection of points $\{m_1, \dots, m_k\}$ equipped with signs $\{\varepsilon_1, \dots, \varepsilon_k\}$ so that we can define

$$\deg F_w := \sum_{i=1}^{k} \varepsilon_k \in \mathbb{Z}. \tag{162}$$

In fact, for any two choices w_0 and w_1 as above, (161) implies that $\deg F_{w_0} = \deg F_{w_1}$ provided w_0 and w_1 are in the same connected component. Thus, $\deg F_w$ does not depend on w as long as W is connected.

The following result summarizes the considerations above.

Theorem 163. *Assume $F\colon X \times W \to Y$ satisfies Hypotheses (a)–(d) above. Then $\deg F_w$ does not depend on w.* □

Clearly, the case $d = 0$ is not really special. Indeed, as we already know for any $d \geq 0$, $\mathcal{M}_w = F_w^{-1}(y)$ is a smooth oriented manifold for all w in a dense subset in W. Then (161) shows that for any two choices w_0 and w_1 in this subset, $\mathcal{M}_{w_0}$ and $\mathcal{M}_{w_1}$ are cobordant, i.e., the oriented cobordism class of $[\mathcal{M}_w]$ is well–defined and does not depend on w. One can take this oriented cobordism class as an invariant, however, in practice it may be hard to deal with. One way to extract a number out of this cobordism class is as follows.

Let $P \to X$ be a (principal) bundle. Assume there are characteristic classes $\alpha_1, \dots \alpha_k$ of P such that $\alpha := \alpha_1 \cup \dots \cup \alpha_k \in H^d(X;\mathbb{Z}) \subset H^d(X;\mathbb{R})$. Since the restriction of α to $\mathcal{M}_w$ can be represented by a closed d–form, say ω, Stokes' theorem implies that

$$\langle \alpha, [\mathcal{M}_w] \rangle = \int_{\mathcal{M}_w} \omega$$

does not depend on w. In fact, this is an integer, since $[\omega]$ represents an integral cohomology class.

6.6 An equivariant setup

Let X be a Banach manifold equipped with an action of a Banach Lie group $\mathcal{G}$. For any $x \in X$ the infinitesimal action of $\mathcal{G}$ at x is given by the linear map

$$R_x\colon \operatorname{Lie}(\mathcal{G}) \to T_x X, \tag{164}$$

whose image is the tangent space to the orbit through x.

For the sake of simplicity of exposition let me assume that $\mathcal{G}$ acts freely on X. It will be also convenient to assume that X and $\mathcal{G}$ are Hilbert manifolds.

Definition 165. A Hilbert submanifold $S \subset X$ containing x is said to be a local slice of the $\mathcal{G}$-action at x, if the set $\mathcal{G}S := \{g \cdot s \mid g \in \mathcal{G},\ s \in S\}$ is open in X and the natural map

$$\mathcal{G} \times S \to \mathcal{G}S, \qquad (g, s) \mapsto g \cdot s$$

is a diffeomorphism.

Proposition 166. *Assume $\mathcal{G}$ acts freely on X. If the $\mathcal{G}$–action admits a slice at any point, then the quotient $X/\mathcal{G}$ is a manifold.*

The proof of this proposition is clear: S can be identified with a neighbourhood of the orbit $\mathcal{G} \cdot x$ in the quotient space $X/\mathcal{G}$.

Example 167. The group $\mathrm{U}(1)$ acts on S^2 by rotations around the z-axis. If we remove the north and the south poles, this action is free. The submanifold

$$S := \{(x, 0, z) \mid x^2 + z^2 = 1, x > 0\} \cong (-1, 1)$$

is a global slice for the $\mathrm{U}(1)$–action. In particular, the quotient is a manifold, which in this case is naturally diffeomorphic to an interval.

Let Y be a smooth manifold equipped with a $\mathcal{G}$–action and $F\colon X \times W \to Y$ be a smooth map such that each $F_w\colon X \to Y$ is $\mathcal{G}$–equivariant. Let $y \in Y$ be a fixed point of the $\mathcal{G}$–action, i.e., $g \cdot y = y$ for all $g \in \mathcal{G}$. For any $x \in F_w^{-1}(y)$ we can construct the sequence

$$0 \to \mathrm{Lie}(\mathcal{G}) \xrightarrow{R_x} T_xX \xrightarrow{d_xF_w} T_yY \to 0, \tag{168}$$

which is in fact a complex. This follows immediately from the equivariancy of F_w (and the assumption that y is a fixed point). This is called *the deformation complex* at x.

The zero's cohomology group of the deformation complex is just the Lie algebra of the stabilizer of x. Our assumption implies that this is trivial.

The second cohomology group is just the cokernel of d_xF_w. This is trivial if and only if y is a regular value for F_w.

It is also easy to understand the meaning of the first cohomology group. Indeed, $\ker d_xF_w$ is the Zariski tangent space to $F_w^{-1}(y)$. Since y is fixed, $\mathcal{G}$ acts on $F_w^{-1}(y)$ so that the first cohomology group can be thought of as the tangent space to the 'moduli space'

$$\mathcal{M}_w := F_w^{-1}(y)/\mathcal{G}$$

at $\mathcal{G} \cdot x$.

Since T_xX is by assumption a Hilbert space, we have a linear map

$$D_x := (R_x^*,\, d_xF_w)\colon T_xX \to \mathrm{Lie}(\mathcal{G}) \oplus T_yY \tag{169}$$

whose kernel can be identified with the first cohomology group of the deformation complex.

Theorem 170. *Let y is a fixed point of the $\mathcal{G}$–action on Y. Assume that the following holds:*

(i) $\mathcal{G}$ acts freely on X;

(ii) y is a regular value for F_w;

(iii) There is a local slice at each point $x \in F_w^{-1}(y)$;

(iv) D_x is a Fredholm linear map of index d.

Then $\mathcal{M}_w$ is a smooth manifold of dimension d.

Proof. The statement is local, so we can restrict our attention to a neighborhood of a point $x \in F_w^{-1}(y)$. Let S be a slice at x so that T_xS and $\mathrm{Im}\, R_w$ are complementary subspaces in T_xX. Then y is still a regular value for $F_w|_S$ and

$$\ker d_xF_w|_S = \ker(R_x^*,\ d_xF_w)$$

so that $F_w^{-1} \cap S$ is a manifold of dimension $\dim\ker D_x = d$. $\square$

By tracing through the discussion of Section 6.5 it is easy to see that the following theorem holds.

Theorem 171. *Assume that in addition to Hypotheses (i)–(iv) of Theorem 170 the determinant line bundle* $\det D_x$ *is trivialized and this trivialization is preserved by the action of* $\mathcal{G}$. *Then* $\mathcal{M}_w$ *is oriented. If in addition* $\mathcal{M}_w$ *is compact for any* w, *then the oriented bordism class of* $\mathcal{M}_w$ *does not depend on* w. □

7 The Seiberg–Witten gauge theory

7.1 The Seiberg–Witten equations

Recall that in dimension 4 we have an isomorphism of $\mathrm{Spin}(4)$ representations

$$\mathbb{R}^4 \otimes \mathbb{C} \cong \mathrm{Hom}(\$^+;\, \$^-), \tag{172}$$

where $\mathrm{Spin}(4)$ acts on $\mathbb{R}^4$ via the homomorphism $\mathrm{Spin}(4) \to \mathrm{SO}(4)$. (172) is still valid as an isomorphisms of $\mathrm{Spin}^c(4)$–representations. Somewhat more explicitly, the standard inclusion $\mathbb{R}^4 \to Cl(\mathbb{R}^4)$ yields a monomorphism

$$\mathbb{R}^4 \to \mathrm{Hom}(\$^+;\, \$^-), \qquad v \mapsto (\psi \mapsto v \cdot \psi).$$

Of course, we have also an inclusion $\mathbb{R}^4 \to \mathrm{Hom}(\$^-;\, \$^+)$. Hence, the Clifford multiplication with a 2–form yields a map $\Lambda^2\mathbb{R}^4 \to \mathrm{End}(\$^\pm)$, whose kernel is $\Lambda^2_\mp\mathbb{R}^4$ as a straightforward computation shows. Moreover, the image of this map consists of skew-Hermitian endomorphisms so that we obtain an isomorphism

$$\Lambda^2_+\mathbb{R}^4 \to \mathfrak{su}(\$^+),$$

whose complexification yields $\Lambda^2_+\mathbb{R}^4 \cong \mathrm{End}_0(\$)$, cf. (98).

Notice also that we have a quadratic map

$$\mu\colon \$^+ \to i\,\mathfrak{su}(\$^+) \subset \mathrm{End}_0(\$^+), \qquad \mu(\psi) = \psi\psi^* - \frac{1}{2}|\psi|^2,$$

where the expression on the right hand side means the following: $\mu(\psi)(\varphi) = \langle\varphi,\psi\rangle\psi + \frac{1}{2}|\psi|^2\,\varphi$. Somewhat more concretely, μ is just the map

$$\mathbb{C}^2 \to i\,\mathfrak{su}(2), \qquad \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \mapsto \frac{1}{2}\begin{pmatrix}|\psi_1|^2 - |\psi_2|^2 & 2\,\psi_1\bar{\psi}_2\\ 2\,\bar{\psi}_1\psi_2 & |\psi_2|^2 - |\psi_1|^2\end{pmatrix}.$$

Hence, we can think of $\mu(\psi)$ as a purely imaginary self–dual 2-form.

A more global version of these identifications is as follows. Pick an oriented Riemannian four–manifold X equipped with a spin^c structure. Denote by $\$^\pm$ the corresponding spinor bundles, see Section 4.3 for details. Then we have the isomorphisms of vector bundles

$$T^*_{\mathbb{C}}X \cong \mathrm{Hom}(\$^+;\, \$^-), \tag{173}$$

$$i\,\Lambda^2_+T^*X \cong i\,\mathfrak{su}(\$^+) \tag{174}$$

and a fiberwise quadratic map

$$\mu\colon \slashed{S}^+ \to i\,\Lambda^2_+ T^*X.$$

With this understood, *the Seiberg–Witten equations* read

$$\slashed{D}^+_A\psi = 0 \quad \text{and} \quad F_A^+ = \mu(\psi). \tag{175}$$

These are the equations for the pair $(\psi, A) \in \Gamma(\slashed{S}^+) \times \mathcal{A}(P_{\det})$. The space of solutions is clearly the zero level set of the Seiberg–Witten map

$$SW\colon \Gamma(\slashed{S}^+) \times \mathcal{A}(P_{\det}) \to \Gamma(\slashed{S}^-) \times \Omega^2_+(X;\, \mathbb{R}i), \qquad SW(\psi, A) = \Big(\slashed{D}^+\psi, F_A^+ - \mu(\psi)\Big).$$

7.1.1 The gauge group action

Since the structure group of $P_{\det}$ is the abelian group $\mathrm{U}(1)$, we have an identification $\mathcal{G} := \mathcal{G}(P_{\det}) \cong C^\infty(X;\, \mathrm{U}(1))$. This acts on $\mathcal{A}(P_{\det})$ on the right by gauge transformations:

$$A \cdot g = A + 2\, g^{-1} dg. \tag{176}$$

We can extend this to the right action of $\mathcal{G}$ on *the configuration space* $\Gamma(\slashed{S}^+) \times \mathcal{A}(P_{\det})$ as follows:

$$(\psi, A) \cdot g = (\bar g \psi,\ A \cdot g).$$

We let also $\mathcal{G}$ act on $\Gamma(\slashed{S}^-)$ on the left in the obvious manner. Extending this action by the trivial one on $\Omega^2_+(X;\, \mathbb{R}i)$, we obtain a left action of $\mathcal{G}$ on $\Gamma(\slashed{S}^-) \times \Omega^2_+(X;\, \mathbb{R}i)$.

Lemma 177. *The Seiberg–Witten map is $\mathcal{G}$-equivariant, i.e.,*

$$SW\big((\psi, A) \cdot g\big) = \bar g \cdot SW(\psi, A).$$

From this we obtain that $\mathcal{G}$ acts on the space of solutions of (175). The quotient

$$\mathcal{M}_{SW} = \big\{ (\psi, A) \text{ is a solution of (175) } \big\} / \mathcal{G}$$

is called *the Seiberg–Witten moduli space.*

Notice that the action of $\mathcal{G}$ on the configuration space is *not* free. Indeed, if $g \in \mathcal{G}$ is the stabilizer of a point (ψ, A) in the configuration space, then (176) implies that g is constant. Hence,

$$\mathrm{Stab}(\psi, A) \neq \{1\} \quad \Longleftrightarrow \quad \psi \equiv 0.$$

The point (ψ, A) with a non-vanishing spinor ψ are called *irreducible*, while points of the form $(0, A)$ are called *reducible*.

Denote also

$$\mathcal{M}^*_{SW} := \big\{ (\psi, A) \text{ is an irreducible solution of (175) } \big\} / \mathcal{G}.$$

7.1.2 The deformation complex

As we know from Section 6.6, for any solution (ψ, A) of the Seiberg–Witten equations, we can associate the deformation complex[5]:

$$0 \to \Omega^0(X;\, \mathbb{R}i) \xrightarrow{R_{(\psi,A)}} \Gamma(\not{S}^+) \oplus \Omega^1(X;\, \mathbb{R}i) \xrightarrow{d_{(\psi,A)}SW} \Gamma(\not{S}^-) \oplus \Omega^2_+(X;\, \mathbb{R}i) \to 0. \tag{178}$$

To explain, since $\mathcal{A}(P_{\det})$ is an affine space modelled on $\Omega^1(X;\, \mathbb{R}i)$, the tangent space to the configuration space at any point can be naturally identified with the middle space of the complex. It is easy to compute the infinitesimal action of the gauge group:

$$R_{(\psi,A)}\,\xi = \big(-\xi\psi,\ 2\, d\xi\,\big), \qquad \xi \in \Omega^0(X;\, \mathbb{R}).$$

Lemma 179. *We have*

$$d_{(\psi,A)}SW(\dot\psi,\, \dot a) = \Big(\not{D}^+_A\dot\psi + \frac{1}{2}\,\dot a\cdot\psi,\ d^+\dot a - 2\mu(\psi, \dot\psi)\Big),$$

where the second summand of the first component means the Clifford multiplication of $\dot a$ and ψ, $d^+\dot a$ is the projection of $d\dot a$ onto the space of self–dual 2-forms, and $\mu(\cdot,\cdot)$ is the polarization of μ.

Proposition 180. *For any solution (ψ, A) of the Seiberg–Witten equations (178) is an elliptic complex.*

The proof of this proposition hinges on the following result, which is of independent interest.

Proposition 181. *For any Riemannian oriented four–manifold X, the Atiyah complex*

$$0 \to \Omega^0(X) \xrightarrow{d} \Omega^1(X) \xrightarrow{d^+} \Omega^2_+(X) \to 0$$

is elliptic.

One can prove this proposition either by computing the principal symbols in local coordinates or by noticing that $d^+ + d^*$ is a twisted Dirac operator. I leave the details to the reader.

Proof of Proposition 180. Modulo zero order terms, which are clearly immaterial for the statement of this proposition, (178) can be written as the direct sum of the Atiyah complex and

$$0 \to 0 \to \Gamma(\not{S}^+) \xrightarrow{\not{D}^+} \Gamma(\not{S}^-) \to 0.$$

The claim follows from the ellipticity of both complexes. □

[5] Strictly speaking, at this point the constructions of Section 6.6 are not applicable, but we will see below how to fix this.

7.1.3 Sobolev completions

The smooth category, which was used to define the Seiberg–Witten map, does not allow us to use the technique developed in Section 6. Hence, we will work with Sobolev completions of the spaces under considerations. This is described next.

Pick any $A_0 \in \mathcal{A}(P_{\det})$ so that we can identify $\mathcal{A}(P_{\det})$ with $\Gamma(T^*X \otimes \mathbb{R}i)$ even if not canonically so. For any fixed (k, p) the space

$$\mathcal{A}^{k,p}(P_{\det}) := A_0 + W^{k,p}(T^*X \otimes \mathbb{R}i).$$

is an affine Banach space, hence a Banach manifold. It is easy to see that the resulting structure is independent of the choice of A_0.

Just like the configuration space, it is also convenient to complete the gauge group. Namely, a map $X \to S^1$ is said to be of class $W^{k,p}$ if the composition $X \to S^1 \subset \mathbb{R}^2$ is in $W^{k,p}(X; \mathbb{R}^2)$. The subset $\mathcal{G}^{k,p}$ of all such maps is not a vector space, however this is a Banach manifold. Moreover, if $kp > 4$ the Sobolev multiplication theorem shows that $\mathcal{G}^{k,p}$ is closed under the pointwise multiplication, so that $\mathcal{G}^{k,p}$ is in fact a Banach Lie group. Its Li algebra is given by

$$\mathrm{Lie}(\mathcal{G}^{k,p}) = W^{k,p}(X; \mathbb{R}i).$$

Proposition 182. *For any k and any $p > 1$ such that $kp > 4 = \dim X$ the Seiberg–Witten map extends as a smooth map*

$$SW \colon W^{k+1,p}(\$^+) \times \mathcal{A}^{k+1,p}(P_{\det}) \to W^{k,p}(\$^-) \times W^{k,p}(\Lambda^2_+ T^*X \otimes \mathbb{R}i).$$

The action of the gauge group extends to a smooth action of $\mathcal{G}^{k+2,p}$ and SW is equivariant with respect to this action.

Proof. Pick a smooth connection A_0 as a reference point. Then recalling (114) we obtain by Theorem 123 (iv)

$$\begin{aligned} \not{D}^+_{A_0+a}\psi &= \not{D}^+_{A_0}\psi + \frac{1}{2}\, a \cdot \psi \in W^{k,p}(\$^-), \\ F^+_{A_0+a} - \mu(\psi) &= F^+_{A_0} + d^+a - \mu(\psi) \in W^{k,p}(\Lambda^2_+ T^*X \otimes \mathbb{R}i), \end{aligned}$$

where $a \in W^{k+1,p}(T^*X \otimes \mathbb{R}i)$.

The fact that the action of the gauge group extends follows from the Sobolev multiplication theorem and (176), which explains the choice $k+2$ when completing the gauge group. □

In what follows, any (k, p) with k sufficiently large would work. For the sake of definiteness, I will stick to $(k, p) = (4, 2)$, which suffices for the arguments invoked below.

7.1.4 Compactness of the Seiberg–Witten moduli space

The most important property of the Seiberg–Witten moduli space is its compactness. In this section I will provide some details why this is the case.

Before going into details, let me briefly explain the approach. Given any sequence (ψ_n, A_n) we need to show that there is a convergent subsequence. This would follow from the compactness of

Sobolev embedding, if we could establish that the sequence $\|(\psi_n, A_n)\|_{W^{5,2}}$ is bounded from above by a constant independent of n. Here we have the freedom to change solutions by the gauge group action, since in fact we want to extract a subsequence, which converges in the quotient space $\mathcal{C}/\mathcal{G}$. In fact, we will see below that (ψ_n, A_n) is bounded in $W^{k,2}$ for any $k \geq 0$ possibly after applying gauge transformations.

The formal proof requires a number of technical lemmas. The key property is as follows.

Lemma 183. *There is a non-negative constant κ, which depends on the background Riemannian metric g only, such that for any solution*

$$(\psi, A) \in \mathcal{C} := W^{4,2}(\$^+) \times \mathcal{A}^{4,2}(P_{\det})$$

of the Seiberg–Witten equations the following estimate holds:

$$\|\psi\|_{L^4} \leq \kappa.$$

Proof. By the Weitzenböck formula, we have

$$\begin{aligned} 0 = \big\langle D\!\!\!/_A^- D\!\!\!/_A^+ \psi, \psi \big\rangle_{L^2} &= \big\langle \nabla_A^* \nabla_A \psi, \psi \big\rangle_{L^2} + \frac{1}{4}\langle s_g \psi, \psi\rangle_{L^2} + \frac{1}{2}\langle F_A^+ \cdot \psi, \psi\rangle_{L^2} \\ &= \|\nabla_A \psi\|_{L^2}^2 + \frac{1}{4}\langle s_g\psi, \psi\rangle_{L^2} + \frac{1}{2}\langle \mu(\psi)\psi, \psi\rangle_{L^2}. \end{aligned}$$

Notice that we have the following pointwise equality:

$$\langle \mu(\psi)\psi, \psi\rangle = \Big\langle |\psi|^2\psi - \frac{1}{2}|\psi|^2\psi,\ \psi\Big\rangle = \frac{1}{2}\,|\psi|^4.$$

Using this we obtain

$$\|\psi\|_{L^4}^4 \leq -\langle s_g\psi, \psi\rangle_{L^2} \leq s_g^-\,\|\psi\|_{L^2}^2,$$

where $s_g^- = \max_{x\in X}\max\{-s_g(x), 0\}$. Using $\|\psi\|_{L^2} \leq C\|\psi\|_{L^4}$, where C is a positive constant, we arrive at $\|\psi\|_{L^4}^4 \leq C s_g^-\,\|\psi\|_{L^4}^2$, which yields the required estimate. □

Remark 184. Notice that the proof of this lemma does not go through for the equations $D\!\!\!/_A^+\psi = 0, F_A^+ = -\mu(\psi)$, which differ from (175) just by a sign.

Corollary 185. *For any solution (ψ, A) of the Seiberg–Witten equations we have the estimates*

$$\|F_A^+\|_{L^2} \leq C \qquad \textit{and} \qquad \|F_A^-\|_{L^2} \leq C - 4\pi^2\, c_1(L_{\det})^2, \tag{186}$$

where the constants depend on the background metric only.

Proof. The second of the Seiberg–Witten equations yields

$$|F_A^+|^2 = |\mu(\psi)|^2 = \frac{1}{2}\,|\psi|^4$$

so that $\|F_A^+\|_{L^2}^2 \leq \kappa^4/2$ by Lemma 183.

To prove the second bound, recall that the first Chern class of $L_{\det}$ is represented by $\frac{i}{2\pi}F_A$ so that we have

$$c_1(L_{\det})^2 = \frac{i^2}{(2\pi)^2}\int_M F_A \wedge F_A = -\frac{1}{4\pi^2}\int_M (F_A^+ \wedge F_A^+ + F_A^- \wedge F_A^-)$$
$$= \frac{1}{4\pi^2}\bigl(\|F_A^+\|_{L^2}^2 - \|F_A^-\|_{L^2}^2\bigr),$$

which yields the required bound. □

Remark 187. Of course, the right hand side of the second inequality of (186) is just a constant independent of the solution. However, this explicit form will be useful below.

Lemma 188. *There is a constant $C > 0$ such that for any solution of the Seiberg–Witten equations we have*

$$\|\psi\|_{C^0} \le C.$$

The proof of the proposition below hinges on the following technical lemma, which follows essentially from the elliptic estimate.

Lemma 189 ([Mor96, Lemma 5.3.1])**.** *Let L be any Hermitian line bundle over M. Fix a smooth reference connection A_0. For sny $k \ge 0$ there are positive constants C_1 and C_2 with the following property: For any $W^{k,2}$–connection A on L there is a gauge transformation $g \in \mathcal{G}^{k+1,2}$ such that $A \cdot g = A_0 + \alpha$, where $\alpha \in W^{k,2}(T^*M \otimes \mathbb{R}i)$ satisfies*

$$d^*\alpha = 0 \qquad \textit{and} \qquad \|\alpha\|_{W^{k,2}} \le C_1\|F_A^+\|_{W^{k-1,2}} + C_2.$$

Moreover, the harmonic component α_h of α can be assumed to be bounded in L^2 by a constant independent of k. □

Proposition 190. *For each $k \ge 0$ there are positive constants $C_k > 0$ with the following property: For any solution (ψ, A) of the Seiberg–Witten equations there is a gauge transformation $g \in \mathcal{G}^{k+1,2}$ such that*

$$\|(\psi, \alpha)\|_{W^{k,2}} \le C_k, \tag{191}$$

where $A \cdot g = A_0 + \alpha$.

Proof. The proof is given via the induction on k. However, a few first values of k require a special treatment.

For $k = 0$ we know already that $\|(\psi, F_A^+)\|_{L^2}$ is bounded by a constant independent of (ψ, A). Lemma 189 yields immediately the required estimate for α.

For $k = 1$, using the Seiberg–Witten equations we have

$$0 = D\!\!\!\!/_A^+\psi = D\!\!\!\!/_{A_0}^+\psi + \frac{1}{2}\alpha \cdot \psi,$$

where the second summand is bounded in L^2 by Lemma 188. Invoking the elliptic estimate for $D\!\!\!\!/_{A_0}^+$, we obtain a bound for ψ in $W^{1,2}$. This proves (191) for $k = 1$.

Let us consider the case $k = 2$. Notice that by the Seiberg–Witten equations we have the pointwise estimate

$$|\nabla^{LC} F_A^+| = |\nabla^{LC}\mu(\psi)| \le C|\nabla^{A_0}\psi||\psi| \le C|\nabla^{A_0}\psi|,$$

where the first inequality follows from the fact that μ is quadratic and the second one follows by Lemma 188. This clearly implies a bound on the $W^{1,2}$-norm of F_A^+. This in turn, yields a bound on the $W^{2,2}$–norm of α.

Furthermore, the bound for $\|\alpha\|_{W^{2,2}}$ together with the Sobolev multiplication theorem yields that $\alpha\cdot\psi$ is bounded in $W^{1,2}$. By the same token as above, this yields the $W^{2,2}$–bound on ψ.

By now, the reader will have no difficulties in proving (191) for $k = 3$. I leave this as an exercise.

We are now prepared to make the induction step. Thus, assume that (191) has been established for some $k \ge 3$. Since $W^{k,2}$ is an algebra for $k \ge 3$, we have a bound on F_A^+ in $W^{k,2}$. Lemma 189 yields a $W^{k+1,2}$–bound on α.

By a similar argument, $\alpha\cdot\psi$ is bounded in $W^{k,2}$ so that the elliptic estimate yields a $W^{k+1,2}$–bound for ψ. This finishes the proof of this proposition. □

Corollary 192. *The Seiberg–Witten moduli space*

$$\mathcal{M} := \big\{(\psi, A) \in \mathcal{C} = \mathcal{C}^{4,2} \mid SW(\psi, A) = 0\big\}/\mathcal{G}^{5,2}$$

is compact.

Proof. Any sequence of solutions (ψ_n, A_n) is gauge equivalent to a sequence $(\psi_n, A_0 + \alpha_n)$ such that (ψ_n, α_n) bounded in $W^{5,2}$. By the Sobolev embedding theorem, a subsequence converges in $\mathcal{C}^{4,2}$ and the limit is a solution of the Seiberg–Witten equations. □

In fact, Corollary 192 can be substantially strengthened, as the following result shows.

Theorem 193. *For each solution $(\psi, A) \in \mathcal{C}^{4,2}$ of the Seiberg–Witten equations there is a gauge transformation $g \in \mathcal{G}^{5,2}$ such that $(\psi, A)\cdot g$ is smooth. Furthermore, $\mathcal{M}$ is homeomorphic to*

$$\mathcal{M}^\infty := \big\{(\psi, A) \in \Gamma(\$^+) \times \mathcal{A}(P_{\det}) \mid SW(\psi, A) = 0\big\}/C^\infty(M;\, \mathrm{U}(1)).$$

This space is compact in the C^∞–topology.

Proof. Let $(\psi, A) \in \mathcal{C}^{4,2}$ be any solution. By Proposition 190, $F_A = F_{A\cdot g} = F_{A_0} + d\alpha \in W^{k-1,2}$ for all k. In particular, F_A is smooth.

Writing $A\cdot g = A_0 + \alpha$ as before, we obtain

$$(d + d^*)\,\alpha = d\alpha = F_A - F_{A_0} \in C^\infty.$$

By the elliptic regularity, α is smooth, i.e., A is smooth. Hence, ψ is also smooth as an element of the kernel of a smooth elliptic differential operator $\not{D}_A^+$.

Furthermore, let (ψ_n, A_n) be any sequence of smooth solutions. By Corollary 192, this contains a subsequence still denoted by (ψ_n, A_n), which converges in the $W^{4,2}$–topology after possibly applying a sequence of $\mathcal{G}^{5,2}$ gauge transformations and the limit lies also in $W^{4,2}$.

Since (ψ_n, A_n) is bounded in $W^{6,2}$, a subsequence converges in $W^{5,2}$ after possibly applying a sequence of $\mathcal{G}^{6,2}$ gauge transformations. Repeating this process for each $k \ge 4$ and choosing the diagonal subsequence, we obtain the claim. □

7.1.5 Slices

In this section we wish to construct local slices for the gauge group action on the subspace of irreducible configurations

$$\mathcal{C}^{4,2}_{\mathrm{irr}} := \big\{ (\psi, A) \in W^{4,2}(\not{S}^+) \times \mathcal{A}^{4,2}(P_{\mathrm{det}}) \mid \psi \not\equiv 0 \big\},$$

where $\mathcal{G}^{5,2}$ acts freely.

The tangent space to the slice at a point (ψ, A) must be transversal to $\operatorname{Im} R_{(\psi,A)}$, hence it is natural to consider the kernel of the formal adjoint operator

$$R^*_{(\psi,A)}(\dot{\psi}, \dot{a}) = 2\, d^*\dot{a} + i\, \operatorname{Re}\langle \psi, i\,\dot{\psi}\rangle.$$

Here $\langle\cdot,\cdot\rangle$ denotes the Hermitian scalar product on the spinor bundle.

Proposition 194. *For any irreducible configuration (ψ, A) the subspace*

$$(\psi, A) + \ker R^*_{(\psi,A)}$$

is a slice for the $\mathcal{G}^{5,2}$–action on $\mathcal{C}^{4,2}_{\mathrm{irr}}$. Here $\ker R^*_{(\psi,A)}$ means the kernel of the map $R^*_{(\psi,A)}\colon W^{4,2} \to W^{3,2}$. □

7.1.6 A perturbation

In general, there is no reason to expect that the origin will be a regular value for the Seiberg–Witten map. However, a family of perturbations just like in Section 6.3 can be constructed by hand.

Proposition 195. *The origin is a regular value of the map*

$$\begin{gathered}\mathcal{SW}\colon \mathcal{C}^{4,2}_{\mathrm{irr}} \times W^{3,2}(\Lambda^2_+T^*M \otimes \mathbb{R}i) \to W^{3,2}(\not{S}^-) \times W^{3,2}(\Lambda^2_+T^*M \otimes \mathbb{R}i),\\ \mathcal{SW}(\psi, A, \eta) = \big(\not{D}_A\psi,\ F_A^+ - \mu(\psi) - \eta\big).\end{gathered}$$

Proof. The perturbation has been chosen so that

$$\mathrm{pr}_2\,\frac{\partial\mathcal{SW}}{\partial\eta}\colon W^{3,2}(\Lambda^2_+T^*M \otimes \mathbb{R}i) \to W^{3,2}(\Lambda^2_+T^*M \otimes \mathbb{R}i)$$

is surjective. Hence, it is enough to show that the map

$$T := \mathrm{pr}_1\, d_{(\psi,A)}\mathcal{SW}_\eta\colon T_{(\psi,A)}\mathcal{C}^{4,2} \to W^{3,2}(\not{S}^-), \qquad (\dot{\psi},\, \dot{A}) \mapsto \not{D}_A\dot{\psi} + \frac{1}{2}\,\dot{A}\cdot\psi$$

is also surjective for any solution (ψ, A) of the perturbed Seiberg–Witten equations.

Assume this is not the case. Then there is a non-zero vector

$$\varphi \in \operatorname{Im} T^\perp := \big\{ \varphi \in W^{3,2} \mid \langle T(\dot{\psi},\, \dot{A}),\, \varphi\rangle_{L^2} = 0 \quad \forall(\dot{\psi},\, \dot{A}) \big\}.$$

In particular, we have $0 = \langle \not{D}^+_A\dot{\psi}, \varphi\rangle_{L^2} = \langle\dot{\psi}, \not{D}^-_A\varphi\rangle_{L^2}$, which implies

$$\not{D}^-_A\varphi = 0. \tag{196}$$

Also,
$$0 = \langle T(0, \dot{A}),\, \varphi \rangle_{L^2} = \langle \dot{A} \cdot \psi,\, \varphi \rangle_{L^2} \tag{197}$$
for all $\dot{A} \in W^{4,2}(T^*M \otimes \mathbb{R}i)$.

Since $W^{4,2} \subset C^0$ in dimension four, ψ is continuous and by assumption does not vanish identically. Hence, there is a point $m \in M$ such that ψ does not vanish on a neighborhood U of m.

Furthermore, notice that the Clifford multiplication with a fixed non-zero vector $\psi_0 \in \$^+$ is surjective, i.e., the map
$$\mathbb{R}^4 \to \$^-, \qquad \mathrm{v} \mapsto \mathrm{v} \cdot \psi_0$$
(here $\$^\pm$ are thought of as $\mathrm{Spin}^c(4)$–representations) is surjective. This implies the following: if φ does not vanish on U, then there is some $\dot{A}$ supported in U such that $\langle \dot{A} \cdot \psi,\, \varphi \rangle_{L^2} > 0$. However, this contradicts (197) so that we conclude that φ must vanish on an open set. Then, by a theorem of Aronszajn [Aro57], (196) implies that φ vanishes identically. This in turn proves the surjectivity of T and finishes the proof of this proposition. ☐

The deformation complex at a solution of the perturbed Seiberg–Witten equations
$$\not{D}_A^+ \psi = 0, \qquad F_A^+ = \mu(\psi) + \eta$$
is given again by (178) since the differentials of $SW = \mathcal{SW}_0$ and $\mathcal{SW}_\eta$ coincide. Hence, this complex is elliptic and therefore the operator
$$\big(d_{(\psi,A)}\mathcal{SW}_\eta,\ R^*_{(\psi,A)}\big) \colon W^{4,2}(\$^+ \oplus T^*X \otimes \mathbb{R}i) \to W^{3,2}(\$^- \oplus \Lambda^2_+ T^*X \otimes \mathbb{R}i) \tag{198}$$
is also elliptic by Exercise 146. Thus, (198) is Fredholm so that by appealing to Theorem 170, we obtain the following result.

Corollary 199. *There is a subset* $\mathrm{H} \subset W^{3,2}(\Lambda^2_+ T^*X \otimes \mathbb{R}i)$ *of the second category such that for any* $\eta \in \mathrm{H}$ *the space*
$$\mathcal{M}_\eta^{irr} := \big\{ (\psi, A) \in \mathcal{C}^{4,2} \mid \mathcal{SW}(\psi, A, \eta) = 0, \quad \psi \not\equiv 0 \big\} / \mathcal{G}^{5,2}$$
is a smooth manifold of dimension
$$d = \frac{1}{4}\Big(c_1(L_{\det})^2 - 2\,\chi(M) - 3\ \mathrm{sign}(M) \Big), \tag{200}$$
where $\chi(M)$ *and* $\mathrm{sign}(M) := b_2^+ - b_2^-$ *are the Euler characteristic and the signature of* M *respectively.*

Proof. We only need to prove the formula for the dimension of the moduli space. To this end the following fact will be useful: The index is a locally constant function on the space of all Fredholm operators. Equivalently, if $\{T_t \mid t \in [0,1]\}$ is a one-parameter family of Fredholm operators, then $\mathrm{index}\, T_0 = \mathrm{index}\, T_1$.

With this understood, consider the operator
$$D_{(\psi,A)} := \big(d_{(\psi,A)} SW_\eta,\ R^*_{(\psi,\eta)}\big) \colon T_{(\psi,A)}\mathcal{C}^{4,2} \to W^{3,2}\big(\$^- \oplus\ \Lambda^2_+ T^*M \otimes \mathbb{R}\, i\ \oplus\ \underline{\mathbb{R}}\, i\big). \tag{201}$$

Writing

$$D_{(\psi,A)} = \begin{pmatrix} \not{D}_A^+ & \\ & d^+ + d^* \end{pmatrix} + B = D_0 + B, \tag{202}$$

where B is a zero order operator, we see that $D_{(\psi,A)}$ is homotopic through Fredholm operators to D_0. Since D_0 decouples, we have

$$\operatorname{index} D_0 = \operatorname{index} \not{D}_A^+ + \operatorname{index}(d^+ + d^*).$$

The latter index is easily computed: $\operatorname{index}(d^+ + d^*) = b_1 - b_0 - b_2^+ = \frac{1}{2}(\chi + \operatorname{sign})$. The index of $\not{D}_A^+$ can be computed by applying the Atiyah–Singer index theorem:

$$\operatorname{index}_{\mathbb{C}} \not{D}_A^+ = \frac{1}{8}\big(c_1(L_{\det})^2 - \operatorname{sign}\big).$$

This yields the result. □

Remark 203. Strictly speaking, at this point we should redo the analysis of the compactness of the Seiberg–Witten moduli space for the perturbed equations. However, this requires cosmetic changes only. The reader should have no difficulties to check that this is indeed the case.

7.1.7 Reducible solutions

While Corollary 199 yields a smooth moduli space, by removing some points we lost an essential property, namely compactness. The following lemma demonstrates how to deal with this problem.

Proposition 204. *Assume $b_2^+ \geq 1$. Then there is an affine subspace $Q \subset W^{3,2}(\Lambda_+^2 T^*M \otimes \mathbb{R}\, i)$ of codimension b_2^+ with the following property: If $\eta \notin Q$, then there are no reducible solutions of the Seiberg–Witten equations.*

Proof. Pick a reference connection A_0 on $L_{\det}$ and denote $Q := F_{A_0}^+ + \operatorname{Im} d^+$, which is a subspace of codimension b_2^+. Clearly, if $\eta \notin Q$, then there are no reducible solutions. □

Corollary 205. *Assume $b_2^+ \geq 1$. Then for a generic η the perturbed Seiberg–Witten moduli space $\mathcal{M}_\eta$ contains no reducible solutions.* □

7.1.8 Orientability of the Seiberg–Witten moduli space

As we have already seen in the proof of Corollary 199, $D_{(\psi,A)}$ defined by (201) is homotopic through elliptic operators to D_0, which is given by (202). This can be also used to orient the Seiberg–Witten moduli space. The key is the following simple observation.

Lemma 206. *If $\{T_{p,0} \mid p \in \mathcal{P}\}$ and $\{T_{p,1} \mid p \in \mathcal{P}\}$ are two homotopic families of linear Fredholm maps, then $\det T_0 \cong \det T_1$. More precisely, this means that $\det T_1$ is trivial if and only if $\det T_0$ is trivial and a trivialization of $\det T_0$ induces a trivialization of $\det T_1$. The latter is well-defined up to a multiplication with an everywhere positive function.*

Proof. Let $\{T_{p,t} \mid (p,t) \in \mathcal{P} \times [0,1]\}$ be a homotopy through linear Fredholm maps. Then $\det T$ is well-defined over $\mathcal{P} \times [0,1]$ and restricts to $\det T_0$ and $\det T_1$ on the corresponding components of the boundary. This implies the statement of this lemma. □

Remark 207. Let $\det T$ be as in the proof of the lemma above. One can construct an isomorphism between $\det T_0$ and $\det T_1$ explicitly by introducing a connection on $\det T$ and taking the parallel transport along the curves $t \mapsto (p, t)$.

With this understood, to orient the Seiberg–Witten moduli space it suffices to check that $\det D_0$ is trivial and pick a trivialization of this bundle. Since D_0 splits, we have

$$\det D_0 \cong \det \not{D}_A^+ \otimes \det(d^+ + d^*).$$

Notice that $\not{D}_A^+$ is a complex linear map. In particular, both $\ker D_A^+$ and $\operatorname{coker} D_A^+ \cong \ker D_A^-$ are complex linear subspaces, hence oriented. This implies that the real determinant bundle $\det \not{D}_A^+$ is trivial.

Furthermore, we have

$$\ker(d^+ + d^*) = H^1_{dR}(M; \mathbb{R}\, i), \qquad \operatorname{coker}(d^+ + d^*) = H^0(M; \mathbb{R}\, i) \oplus H^2_+(M; \mathbb{R}\, i).$$

Hence, by picking an orientation of these cohomology groups, a trivialization of $\det(d^+ + d^*)$ is fixed. This in turn yields a trivialization of $\det D_0$, hence also of $\det D_{(\psi,A)}$. Thus, the Seiberg–Witten moduli space is orientable and a choice of orientations of the cohomology groups $H^0(M; \mathbb{R})$, $H^1(M; \mathbb{R}\, i)$ and $H^2_+(M; \mathbb{R}\, i)$ yields an orientation of the Seiberg–Witten moduli space.

Remark 208. Strictly speaking, the orientation of the Seiberg–Witten moduli space we constructed in this section is *not* canonical. First, there are at least two commonly used non-equivalent conventions how to orient complex linear spaces. Namely, if V is a complex linear space and $(\mathrm{v}_1, \dots, \mathrm{v}_k)$ is a complex basis of V, the underlying real vector space $V_{\mathbb{R}}$ can be oriented by saying that

$$(\mathrm{v}_1, \dots, \mathrm{v}_k, i\mathrm{v}_1, \dots, i\mathrm{v}_k) \qquad \text{or} \qquad (\mathrm{v}_1, i\mathrm{v}_1, \dots, \mathrm{v}_k, i\mathrm{v}_k)$$

determines an orientation of $V_{\mathbb{R}}$. This yields the opposite orientations in the case $k = \dim_{\mathbb{C}} V$ is even.

Secondly, orientations of $H^0(M; \mathbb{R})$, $H^1(M; \mathbb{R}\, i)$ and $H^2_+(M; \mathbb{R}\, i)$ is again a choice. Notice however, that these vector spaces depend on the topological structure of M only.

7.2 The Seiberg–Witten invariant

Combining results obtained in the preceding sections we arrive at our main result.

Theorem 209. *Assume $b_2^+(M) \geq 2$. For any spin^c-structure $\sigma \in \mathcal{S}(M)$ and any generic η the perturbed Seiberg–Witten moduli space $\mathcal{M}_\eta$ is a smooth compact oriented manifold of dimension d, which is given by* (200). *Moreover, if η_0 and η_1 are any two generic perturbations, then $\mathcal{M}_{\eta_0}$ and $\mathcal{M}_{\eta_1}$ are oriented-bordant.* □

The only thing, which perhaps needs an explanation, is the hypothesis $b_2^+(M) \geq 2$. The point is that this ensures that a generic bordism between $\mathcal{M}_{\eta_0}$ and $\mathcal{M}_{\eta_1}$ does not contain reducible solutions and therefore is smooth.

Remark 210. Notice that $\mathcal{M}_\eta$ depends on the choice of σ but this dependence is suppressed in the notations.

With this understood, we can proceed to the definition of the Seiberg–Witten invariant. Pick a point $m_0 \in M$ and consider *the based gauge group* $\mathcal{G}_0 := \{ g \in W^{5,2}(M; \mathrm{U}(1)) \mid g(m_0) = 1 \}$, which fits into the exact sequence

$$\{1\} \to \mathcal{G}_0 \to \mathcal{G} \xrightarrow{\mathrm{ev}_{m_0}} \mathrm{U}(1) \to \{1\},$$

where ev_{m_0} is the evaluation at m_0. Then the quotient

$$\hat{\mathcal{M}}_\eta := \{(\psi, A) \in \mathcal{C}^{4,2} \mid SW_\eta(\psi, A) = 0\}/\mathcal{G}_0$$

is called *the framed moduli space* and is equipped with a free action of $\mathrm{U}(1) = \mathcal{G}/\mathcal{G}_0$. Clearly, the quotient $\hat{\mathcal{M}}_\eta/\mathrm{U}(1)$ is just $\mathcal{M}_\eta$ so that $\hat{\mathcal{M}}_\eta$ is a principal $\mathrm{U}(1)$–bundle over $\mathcal{M}_\eta$. Let μ be the first Chern class of this $\mathrm{U}(1)$–bundle. Notice that this $\mathrm{U}(1)$–bundle is just a restriction to $\mathcal{M}_\eta$ of the following bundle: $\mathcal{C}_{\mathrm{irr}}/\mathcal{G}_0 \to \mathcal{C}_{\mathrm{irr}}/\mathcal{G}$.

Definition 211. The Seiberg–Witten invariant of M is a function $\mathrm{sw}\colon \mathcal{S}(M) \to \mathbb{Z}$ defined by

$$\mathrm{sw}(\sigma) := \begin{cases} \left\langle [\mathcal{M}_\eta], \mu^{d/2} \right\rangle & \text{if } d \text{ is even,} \\ 0 & \text{if } d \text{ is odd.} \end{cases}$$

Theorem 212. *The Seiberg–Witten invariant vanishes for all but finitely many spin^c structures.*

Proof. If $\mathrm{sw}(\sigma) \neq 0$, then $\mathcal{M}_\eta$ is a smooth manifold of dimension $d \geq 0$. Hence,

$$c_1(L_{\det})^2 \geq 2\,\chi(M) + 3\,\mathrm{sign}(M).$$

By (186), we have[6] $\|F_A^-\|_{L^2}^2 \leq C$, where C depends on the background Riemannian metric and the topology of M only. This yields that $\|F_A\|_{L^2}^2$ is also bounded by a constant, which depends on the background Riemannian metric and the topology of M only. This in turn shows that $|c_1(L_{\det})|^2 = |c_1(L_{\det}) \cup \mathrm{PD}(c_1(L_{\det}))| = \frac{1}{4\pi^2}\|F_A\|_{L^2}^2 \leq C$, where PD stays for the Poincaré dual class. This proves the statement of this theorem. $\square$

7.2.1 Sample application of the Seiberg–Witten invariant

The Seiberg–Witten invariant are most well–known for its applications to the topology of smooth four–manifolds. I restrict myself just to a list of some sample applications.

Theorem 213 (Witten)**.** *If M with $b_2^+ \geq 2$ admits a metric of positive scalar curvature, then* $\mathrm{sw}_M \equiv 0$. $\square$

The proof of this theorem follows easily from the Weitzenböck formula.

Theorem 214 (Witten)**.** *Let M_1 and M_2 be closed four-manifolds both with $b_2^+ \geq 1$. Then the Seiberg–Witten invariant of the connected sum $M_1 \# M_2$ vanishes.* $\square$

[6] At this point we should have used an analogous a priory estimate for the perturbed Seiberg–Witten equations, cf. Remark 203. However, this does not change the argument.

These vanishing theorems are complemented by the following non-vanishing result, which can be also viewed as the statement about obstructions for the existence of symplectic strutures on closed smooth four-manifolds.

Theorem 215 (Taubes)**.** *If M is symplectic and $b_2^+(M) \geq 2$, then* $\mathrm{sw}_M \not\equiv 0$.

The results below were originally obtained by other methods, however can be also obtained with the help of the Seiberg–Witten theory.

Theorem 216 (Donaldson)**.** *There are (many) closed topological four-manifolds, which do not admit a smooth structure.*

Theorem 217 (Fintushel–Stern)**.** *There are infinitely many closed four-manifolds, which are all homeomorphic, but pairwise non-difeomorphic.*

Put differently, the last two theorems say the following: On a given closed topological four-manifold there may or may not be a smooth structure and if a smooth structure exists, there may be infinitely many smooth structures.

References

[Aro57] N. Aronszajn, *A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order*, J. Math. Pures Appl. (9) **36** (1957), 235–249. MR0092067 ↑65

[BT03] D. Barden and C. Thomas, *An introduction to differential manifolds*, Imperial College Press, London, 2003. MR1992457 ↑49

[GP10] V. Guillemin and A. Pollack, *Differential topology*, AMS Chelsea Publishing, Providence, RI, 2010. Reprint of the 1974 original. MR2680546 ↑49, 53

[GS99] V. Guillemin and S. Sternberg, *Supersymmetry and equivariant de Rham theory*, Mathematics Past and Present, Springer-Verlag, Berlin, 1999. MR1689252 ↑20

[LM89] B. Lawson and M.-L. Michelsohn, *Spin geometry*, Princeton Mathematical Series, vol. 38, Princeton University Press, Princeton, NJ, 1989. MR1031992 (91g:53001) ↑31, 34

[Mil97] J. Milnor, *Topology from the differentiable viewpoint*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Based on notes by David W. Weaver, Revised reprint of the 1965 original. MR1487640 ↑50

[Mor96] J. Morgan, *The Seiberg–Witten equations and applications to the topology of smooth four-manifolds*, Mathematical Notes, vol. 44, Princeton University Press, Princeton, NJ, 1996. MR1367507 ↑62

[MS12] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, Second, American Mathematical Society Colloquium Publications, vol. 52, American Mathematical Society, Providence, RI, 2012. MR2954391 ↑53

[Wel80] R. O. Wells Jr., *Differential analysis on complex manifolds*, Second, Graduate Texts in Mathematics, vol. 65, Springer-Verlag, New York-Berlin, 1980. MR608414 ↑46