A glimpse of gauge-theoretic panoramas

Tommaso Pacini*

April 4, 2022

Abstract

Gauge theory is the mathematical equivalent of a wizard's cauldron. In go lizard tails (Topology: differential and algebraic), bat wings (Analysis: real, complex and PDE), and the pulsing heart of a fiery dragon (Geometry: differential and algebraic). Out comes an amazing, misterious, landscape made of deep, intoxicating, relationships between far-off theories. Brace yourself, breath deeply, plunge in. Enjoy.

Remark. No animals were harmed in the writing of these notes. Don't forget to send me your corrections.

1 Introduction

Analysis on $\mathbb{R}^n$ encounters no difficulties in differentiating vector fields and in higher-order calculus: everything reduces to repeated derivatives of functions. Partial derivatives commute, and the concept of constant vector fields leads to no surprises: constant means constant, no matter what route the vector field follows within its domain.

The concept of a differentiable structure has the stated goal of extending analysis to manifolds, but only does half the job: we can differentiate functions and forms but not vector fields, tensors or sections of vector bundles. In particular, without additional structure we cannot even define second derivatives of a function. In this sense, a differentiable structure only provides the foundation for first-order calculus.

Example 1. Let $E \to M$ denote a $\mathbb{R}$-vector bundle of dimension r. By definition there exists an open covering $\{U_i\}$ of M and local frames of $E_{|U_i}$, thus coordinates. It follows that $E_{|U_i} \simeq U_i \times \mathbb{R}^r$. Let $g_{ij} : U_i \cap U_j \to GL(r, \mathbb{R})$ denote the corresponding transition maps. This means that if a section σ has coordinates $s^i := (s^i_1, \ldots, s^i_r)^T$ over U_i and coordinates $s^j := (s^j_1, \ldots, s^j_r)^T$ over U_j, then $s^i = g_{ij} \cdot s^j$ on $U_i \cap U_j$.

The naive way of differentiating σ in a given direction $X \in T_pM$, which uses only the differentiable structure on M, would be by positing that the vector

*University of Torino, Italy, tommaso.pacini@unito.it

$\nabla_X \sigma(p)$ has coordinates $ds^i(X)$ in any U_i. This definition is however well-posed only if $ds^i(X) = g_{ij} \cdot ds^j(X)$. Substitution on the LHS then leads to $dg_{ij}(X) \cdot s^j + g_{ij} \cdot ds^j(X) = g_{ij} \cdot ds^j(X)$, for all X and p. This implies that g_{ij} must be constant: this generally does not hold, it is a strong assumption on E.

We summarize as follows. (i) There is no difficulty in defining derivatives inside a single chart, eg of sections compactly suppported in the chart. Problems arise with global sections, or if we want to make the definition independent of the chart. (ii) The naive definition works only for a special class of vector bundles: we will meet such bundles later on, in the context of flat connections. (iii) In general, it is instead necessary to neutralize the term dg_{ij} via a more complicated notion of differentiation, encoded in the general notion of "connection".

Example 2. A function on $\mathbb{R}^n$ is convex iff it restricts to a convex function along lines. This seemingly innocuous notion, closely related to the Hessian of f, hides the fact that, even on $\mathbb{R}$, a function can change its nature drastically under reparametrization of the domain or the range: the concave function $\log x$ becomes linear or convex under multiple right-composition with exp, and conversely with left-composition. The point is that the notion itself of line changes. Connections provide a coordinate-free definition of lines (geodesics), eliminating this ambiguity. In turn, this provides the notion of second derivatives of a function.

The full goal of analysis on manifolds is thus achieved only after adding a connection, but connections are not unique so this requires making an extra choice. Starting from a topological manifold, we thus face two choices: a differentiable structure and a connection. Each choice comes with its own moduli spaces, features, etc., and it is natural to try to make a "best choice".

Our viewpoint here will be that we have already made the choice of a differentiable structure. Our goal is thus to focus only on the connections compatible with that choice, describing their properties and providing means to distinguish them. Along the way we shall discuss the new features that arise in calculus on manifolds: non-commutativity of derivatives and holonomy groups. This will help us start to understand which connection might be "best" for our purposes.

Remark. Students generally first encounter connections in the context of Riemannian geometry. This is incongruous given that, as explained above, the concept lies squarely within the wider realm of differential geometry. It is explained by the fact that a Riemannian metric induces a canonical connection on the tangent bundle, uniquely defined by certain properties. This is the Levi-Civita connection.

Extending analysis to manifolds provides the first, basic, reason for studying connections. The next step is based on the observation that connections tend to incorporate important elements of the underlying geometry. We shall discuss several instances of this principle.

The simplest instance is perhaps Chern-Weil theory, which shows how to use connections to construct invariants of vector bundles.

Another is provided by the theory of flat connections and the Narasimhan-Seshadri theorem, which shows how certain types of connections correspond to special structures on a vector bundle: local systems and stable holomorphic structures.

The most sophisticated manifestation of this principle mentioned in these notes concerns Donaldson's work on differentiable structures on 4-manifolds. In this context one considers families (moduli spaces) of connections. It turns out that certain such moduli spaces contain highly non-trivial information on the differentiable structure of the underlying manifold. We refer to [**?**] for a full account of this.

Further motivation for studying gauge theory comes from the wealth of applications and ideas which originated there, then produced results also in other parts of mathematics. The notion of bubbling and the analogies and relationships between gauge theory and calibrated geometry are important examples of this. We shall only touch upon these topics.

Yet another source of motivation comes from particle physics. The theory of connections is the mathematical formulation of the theory of gauge fields, hence its name. We will not investigate this relationship.

2 Foundations

In this section M will refer to a differentiable manifold, generally of dimension n. $\mathbb{K}$ will denote either $\mathbb{R}$ or $\mathbb{C}$.

Vector bundles. Recall that a smooth vector bundle over a field $\mathbb{K}$ is a smooth manifold E endowed with a surjective map $\pi : E \to M$ over a smooth manifold M such that (i) each fibre $E_x := \pi^{-1}(x)$ has the structure of a $\mathbb{K}$-vector space of fixed dimension r, (ii) there exists an open cover $\{U_i\}$ of M and fibre-preserving diffeomorphisms $\phi_i : \pi^{-1}(U_i) \to U_i \times \mathbb{K}^r$ such that $\phi_i \circ \phi_j^{-1} : (U_i \cap U_j) \times \mathbb{K}^r \to (U_i \cap U_j) \times \mathbb{K}^r$ are isomorphisms on each fibre. We thus obtain maps $g_{ij} : U_i \cap U_j \to \mathrm{GL}(r, \mathbb{K})$, $g_{ij}(x) := (\phi_i \circ \phi_j^{-1})_{|x \times \mathbb{K}^r}$. The data $\{U_i, \phi_i\}$ (or, to simplify, just $\{U_i\}$) is called a *local trivializing atlas*, while g_{ij} are called the *transition maps*.

This definition generates two equivalent points of view on the concept of "equivalent" vector bundles. The global point of view is that two vector bundles E, E' over the same M are *equivalent* if there exists a fibre-preserving diffeomorphism $\Lambda : E \to E'$ which covers the identity on M and restricts to an isomorphism on each fibre. The local point of view is that there exist trivializing atlases wrt the same open cover $\{U_i\}$ and maps $\lambda_i : U_i \to \mathrm{GL}(r, \mathbb{K})$ such that $\lambda_i g_{ij} = g'_{ij} \lambda_j$ on $U_i \cap U_j$. The maps λ_i are simply the matrices representing Λ in terms of the given trivializations.

Every vector bundle admits smooth sections. In general, any section will vanish in some points. We are particularly interested in the existence of nowhere-vanishing sections, because any such section determines a trivial line sub-bundle, thus a splitting $E = E' \oplus \mathbb{R}$. This generalizes to looking for sections which are pointwise linearly independent. This is a strictly stronger condition than the global linear independence of sections. Indeed, the space of global sections is infinite-dimensional, but there can exist at most r pointwise linearly independent sections.

G-vector bundles. The vector bundles above are "naked", in the sense that the definition does not include any extra structure. We can incorporate such structure as follows.

We say that E has *structure group* G, or is a G-vector bundle, if there exists a trivializing atlas and an action of G on $\mathbb{K}^r$, ie a homomorphism $\rho : G \to \mathrm{GL}(n, \mathbb{K})$, such that, up to identifications, $g_{ij} \in G$. We will refer to this as a *G-trivializing atlas.* Up to quotienting by the normal subgroup defined by the kernel, one can assume the homomorphism is injective.

The structure group is typically related to the datum of additional algebraic structure on E; alternatively, to special properties of E.

Example. Using the standard actions:

(i) If the bundle is real and has a Euclidean metric g we may use the Gram-Schmidt algorithm to build local orthonormal trivializations. This means that g is locally identified with the standard Euclidean structure on $\mathbb{R}^r$, so the new transition maps are in $\mathrm{O}(r, \mathbb{R})$: we have achieved a reduction to the structure group $G := \mathrm{O}(r, \mathbb{R})$.

(ii) In the complex Hermitian setting with metric h we analogously obtain $G := \mathrm{U}(r)$.

(iii) The condition $G := \mathrm{SL}(n, \mathbb{K})$ implies that the determinant line bundle associated to E, ie the line bundle whose transition maps are $\det(g_{ij}) : U_i \cap U_j \to \mathbb{K}$, is trivial. We can identify this bundle with top wedge product $\Lambda^r(E)$. The dual line bundle is thus also trivial, so it admits a (non-canonical) global section Ω: this is a volume form on E.

(iv) The choice $G := \mathrm{GL}(r, \mathbb{C}) \leq \mathrm{GL}(2r, \mathbb{R})$, obtained via the embedding $A + iB \mapsto \begin{pmatrix} A & -B \\ B & A \end{pmatrix}$, corresponds to a complex vector bundle structure on E.

Notice: at this stage E is only acquiring an algebraic structure. We are not enforcing additional differential properties on this structure such as, in case (iv) above, that E be holomorphic. We will do this later.

We shall be interested in two types of vector bundle morphisms. $\mathrm{Aut}_G(E)$ will denote the automorphisms which preserve each fibre and its structure. $\mathrm{End}_{\mathfrak{g}}(E)$ will denote the endomorphisms which preserve each fibre and act according to the induced action of the Lie algebra $\mathfrak{g}$. For example, if $G = \mathrm{O}(r, \mathbb{R})$ then the automorphisms are the fibrewise isometries and the endomorphisms are the fibrewise anti-symmetric maps.

Example. Set $M := \mathbb{S}^1$. We can build a atlas on M by writing it as the union of three arcs U_1, U_2, U_3 (the intersection of two arcs would have multiple components, thus generating labelling confusion, while using only one open interval would imply a self-intersection: this, properly speaking, does not constitute a chart). We can build a $\mathbb{R}$-line bundle E over $\mathbb{S}^1$ by using the transition map Id on two components, the map $v \mapsto -v$ on the third. The total space of this bundle is the Möbius strip. Notice that $E \otimes E$ is the trivial $\mathbb{R}$-line bundle. These are the only line bundles on $\mathbb{S}^1$.

Notice that, if $r > 1$, any $\mathbb{R}^r$-bundle E over $\mathbb{S}^1$ admits a non-vanishing section because $\mathbb{R}^r \setminus \{0\}$ is arc-connected. Iterating this procedure, we find that $E = \mathbb{R}^{r-1} \oplus E'$, where E' is possibly non-trivial. The same reasoning shows that any complex vector bundle on $\mathbb{S}^1$ is trivial.

Set $M := \mathbb{CP}^n$. The definition $\mathbb{CP}^n := (\mathbb{C}^{n+1} \setminus \{0\})/\mathbb{C}^*$ generates a tautological line bundle over M. Each fibre is a complex line in $\mathbb{C}^{n+1}$, so it has a canonical Hermitian structure. The corresponding $\mathbb{S}^1$-bundle is known as the Hopf bundle over $\mathbb{CP}^n$. The total space of the Hopf bundle is the unit sphere $\mathbb{S}^{2n+1}$: this corresponds to the fact that $\mathbb{CP}^n$ can also be obtained as $\mathbb{S}^{2n+1}/\mathbb{S}^1$.

Frame bundles. There exists an alternative, very efficient, method of discussing structure groups of vector bundles. It is analogous to defining, for example, a Euclidean structure on a vector space by choosing the set of orthonormal bases. Clearly, such bases must be related by orthogonal changes of coordinates. This point of view can be expressed in terms of frame bundles; more generally, of principal fibre bundles. An important ingredient here is the group action. To explain how it works, we start as follows.

Digression. Let V be a vector space. Given any two bases $\mathcal{C}, \mathcal{C}'$, let $M^{\mathcal{C},\mathcal{C}'}$ denote the matrix in $\mathrm{GL}(n,\mathbb{K})$ whose columns contain the coordinates of the vectors of $\mathcal{C}'$ wrt $\mathcal{C}$. This convention leads to the following fact: $M^{\mathcal{C},\mathcal{C}''} = M^{\mathcal{C},\mathcal{C}'}M^{\mathcal{C}',\mathcal{C}''}$. If we set $\mathcal{C} = \{v_i\}$, $\mathcal{C}' = \{v_i'\}$, the definition corresponds to the rule $(v_1, \ldots, v_n)M = (v_1', \ldots, v_n')$, where multiplication is formally defined in the usual manner.

We shall be using two group actions on the set of bases (ie homomorphisms into the group of permutations of all bases). The first is the action of $\mathrm{GL}(n,\mathbb{K})$ defined as follows: $\mathcal{C} \cdot M := \mathcal{C}'$, where $M = M^{\mathcal{C},\mathcal{C}'}$. The choice of writing the group element on the RHS, rather than on the LHS, is consistent with the above multiplication rule and is a notationally convenient way of emphasing the fact that this action corresponds to a group anti-homomorphism (also known as a right group action). This action is free and transitive, so it defines a 1:1 correspondence between the set of all bases and $\mathrm{GL}(n,\mathbb{K})$.

The second is the action of $\mathrm{Aut}(V)$ defined by $g \cdot \mathcal{B} := \{g(v_1), \ldots, g(v_n)\}$. This corresponds to a group homomorphism (also known as a left group action). It is again free and transitive.

The two actions are related by the following rule: $(g \cdot \mathcal{B}) \cdot M = g \cdot (\mathcal{B} \cdot M)$, ie the two actions commute. It expresses the fact that if two bases are related by M, then the image bases under g are again related by M.

Now assume $E \to M$ has structure group G defined, as usual, via an injective ρ. We first want to identify an appropriate frame bundle. Choose a G-trivializing atlas. We may then consider, for each $x \in E_{|U_i}$, the subset of frames of E_x which are identified with some $g \cdot \mathcal{B}$, where $\mathcal{B}$ is the standard frame of $\mathbb{K}^r$ and $g \in G$. The condition $g_{ij} \in G$ implies that these subsets glue together to form a fibre bundle P over M. The assumption that ρ is injective implies that the fibre is diffeomorphic to G, so locally $P_{|U_i} \simeq U_i \times G$. The above digression shows how to obtain a right group action of $G \leq \mathrm{GL}(r, \mathbb{K})$, so P is a G-principal fibre bundle P over M.

In summary: no restriction on the structure group corresponds to working with the full $\mathrm{GL}(r, \mathbb{K})$-bundle of all frames on E; equivalently, with a generic bundle E. The restriction of the structure group to G corresponds to the choice (when it exists) of a G-subbundle of this $\mathrm{GL}(r, \mathbb{K})$-bundle; equivalently, to the choice of some additional structure on E and of local trivializations which put this structure in standard form.

An interesting aspect of the principle fibre bundle viewpoint is that it provides a uniform method for building many other bundles which are geometrically related to the initial vector bundle: any representation $\rho : G \to \mathrm{Aut}(V)$ defines an equivalence relation on $P \times V$, $(p, v) \sim (pg, \rho(g^{-1})v)$, thus an associated vector bundle

$$P \times_\rho V := \{[p, v]\}.$$

Analogous constructions are possible replacing $\mathrm{Aut}(V)$ with $\mathrm{Aut}(G)$, $\mathrm{Diff}(F)$ (where F is any chosen abstract fibre), etc.

In our case, for example, we can recover E as the vector bundle associated to P via the action $\rho : G \to \mathrm{GL}(r, \mathbb{K}) = \mathrm{Aut}(\mathbb{K}^r)$, ie $E = P \times_\rho \mathbb{K}^r$. In this setting sections of E are equivalence classes.

Appropriate tensor products of $\mathbb{K}^r$ and induced representations yield the corresponding tensor bundles.

Two other canonical bundles are respectively constructed via the *adjoint actions* $Ad : G \to \mathrm{Aut}(G)$, $g \cdot h := ghg^{-1}$ and $ad : G \to \mathrm{Aut}(\mathfrak{g})$, defined by differentiating Ad. When dealing with matrix groups the latter coincides with $g \cdot X := gXg^{-1}$. Let $Ad(P) := P \times_{Ad} G$, $ad(P) := P \times_{ad} \mathfrak{g}$ be the corresponding *adjoint bundles*, whose general fibres are isomorphic to G, respectively $\mathfrak{g}$. In our case it turns out that $Ad(P)$ can be identified with $\mathrm{Aut}_G(E)$ and that $ad(P)$ can be identified with the bundle $\mathrm{End}_{\mathfrak{g}}(E)$. The latter identification, for example, is via the map

$$ad(P) \to \mathrm{End}_{\mathfrak{g}}(E), \quad [p, X]\sigma := [p, \rho(X)s],$$

where $\sigma \in E_x$, elements in $ad(P)$ are represented by classes of the form $[p, X]$ with $p \in P_x$ and $X \in \mathfrak{g}$, and σ is represented by the element $[p, s] \in P \times_\rho \mathbb{K}^r$.

Remark. A second interesting aspect of the principle fibre bundle viewpoint is that it extends beyond frame bundles, allowing us for example to define Spin-structures on a manifold.

Remark. It is interesting to compare the transition maps used to generate a principal fibre bundle with those used for $\mathrm{Aut}(P)$. In the first case, left multiplication is an element of $\mathrm{Diff}(G)$ so the fibres are diffeomorphic to G. They do not have a Lie group structure but they do admit a right G-action. In the second case the maps are in $\mathrm{Aut}(G)$ so each fibre has a group structure isomorphic to G, but does not necessarily admit a G-group action.

Notice that only in first case do we obtain a construction of associated vector bundles. Furthermore, any principal fibre bundle admitting a global section is trivial. In the second case the identity element in each fibre provides a global section.

Consider for example the Klein bottle, viewed as a $\mathbb{S}^1$-bundle over $\mathbb{S}^1$. The transition map is of the form $e^{i\theta} \mapsto e^{-i\theta}$. This is a group automorphism so each fibre has the structure of a Lie group but is not a principal fibre bundle. This corresponds to the fact that the bundle is not trivial and that the Klein bottle is not the frame bundle of a Hermitian complex line bundle over $\mathbb{S}^1$.

Connections. The most practical definition is that a *connection* on E is a linear operator $\nabla : \Lambda^0(E) \to \Lambda^1(E)$ satisfying the Leibniz rule $\nabla(f\sigma) = df \otimes \sigma + f\nabla\sigma$. This immediately provides us with a way of differentiating sections of E. In particular we obtain the notion of "constant" sections, better known as *parallel*, satisfying the condition $\nabla\sigma = 0$.

Remark. The fact $\nabla_{fX}\sigma = f\nabla_X\sigma$ holds by definition of $\Lambda^1(E)$. This is an important feature of any reasonable notion of derivatives, and explains, for example, why Lie derivatives are not to be considered "true derivatives" of tensors.

Any E admits at least one connection, obtained by gluing together locally defined trivial connections. Specifically, choose a trivializing atlas $\{U_i\}$ and a partition of unity ρ_i subordinate to this atlas. Recall that, by definition, for each $x \in M$ there is only a finite number of indices for which $\rho_i(x) \neq 0$. Let ∇^i denote the trivial connection, defined on sections over each U_i via the standard operator d acting on coordinates. Define $\nabla\sigma := \sum \nabla^i(\rho_i\sigma)$. It is not immediately clear that, at each point, this is a finite sum. This however follows from

$$\sum \nabla^i(\rho_i\sigma) = \sum (d\rho_i)\sigma + \sum \rho_i\nabla^i\sigma = d(\sum \rho_i)\sigma + \sum \rho_i\nabla^i\sigma = \sum \rho_i\nabla^i\sigma.$$

It is simple to check that this connection has the required properties.

Remark. The reason this construction works is that the partition functions allow us to localize any section, thus avoiding the use of transition maps to glue together locally defined global sections. Of course, this connection depends on the specific partition of unity.

It is also simple to check that the difference between two connections is a globally defined tensorial quantity: specifically, it is an element in $\Lambda^1(\mathrm{End}(E))$. Conversely, any connection can be perturbed by adding an element of this type. It follows that the space of all connections is an infinite-dimensional affine space parametrized by $\Lambda^1(\mathrm{End}(E))$.

The above clarifies that the definition allows for a great deal of flexibility because it allows the appearance of 0-order terms. This is also apparent in terms of a local trivialization of E, ie a local moving frame $\mathcal{B}$ defined on an open subset $U \subseteq M$. Assume σ has local coordinates $s = (s_1, \dots, s_r)^T$, $s_i \in \mathbb{K}$. The previous reasoning shows that any connection is locally of the form $d + A$ for some $A \in \Lambda^1(\mathrm{End}(\mathbb{K}^r))$.

As usual, for this to make sense it must however satisfy the correct transformation rules. Let us choose a second local moving frame $\mathcal{B}'$ on U and change trivialization via matrices $Q = Q(x) := M^{\mathcal{B},\mathcal{B}'}(x)$, so that $s = Qs'$. In this coordinate system, $\nabla\sigma$ will have coordinates of the form $ds' + A's'$, and the requirement is that $ds + As = Q(ds' + A's')$. On the LHS, substitution shows $d(Qs') + A(Qs') = Q(ds' + (Q^{-1}dQ + Q^{-1}AQ)s')$, so the transformation rule is $A' = Q^{-1}dQ + Q^{-1}AQ$.

In particular, this applies to the case $U := U_i \cap U_j$ and $Q := g_{ij}$, where $\{U_i\}$ is a trivializing atlas for E. We will then write that ∇ is of the form $d + A_i$. It is customary to simplify the notation using A rather than A_i. A is also sometimes used to denote the connection ∇ related to it.

Remark. It may be useful to emphasize that a connection is not a tensorial object. Specifically, it is not C^∞-linear. That said, some types of problems such as how its expression changes when we change trivialization are common to both. In particular, it can be important to find trivializations with respect to which the local expression has better properties. In the context of Riemannian metrics, for examples, one might be interested in local ON frames. In the context of connections it can be useful to impose certain differential conditions on A, such as the "Coulomb gauge" condition used by Uhlenbeck.

Our next job is to understand how to adapt this theory to the context of vector bundles endowed with a G-structure: we want the connection to be somehow compatible with the G-invariant subset P of frames on E.

Locally, ie in terms of a G-trivializing atlas, this compatibility condition is expressed by the requirement $A_i \in \Lambda^1(\mathrm{End}_{\mathfrak{g}}(\mathbb{K}^r))$. Notice that this condition is preserved when we change trivialization using $Q(x) = g(x) \in G$. Indeed, given any $X \in T_xM$ thus $dQ(X) \in T_gG$, we may choose g_t such that $g_0 = g$, $\frac{d}{dt}g_{t|t=0} = dQ(X)$. We then find $Q^{-1}dQ(X) = g^{-1}(\frac{d}{dt}g_{t|t=0}) = \frac{d}{dt}(g^{-1}g_t)_{|t=0} \in \mathfrak{g}$ and $Q^{-1}A_i(X)Q \in \mathfrak{g}$.

The difference between two compatible connections is an element in the restricted space $\Lambda^1(\mathrm{End}_{\mathfrak{g}}(E))$.

Given any G-structure on E, one can prove that a compatible connection always exists. When the G-structure corresponds to an algebraic structure on E, the compatibility condition implies special facts about that structure. We illustrate both facts via the following example.

Example. Let (E, g) be a vector bundle endowed with a Euclidean metric. Assume ∇ is compatible with g, in the above sense: this means that, using local orthonormal frames, $\nabla = d + A$ where A take values in the space $\mathfrak{o}(r)$ of anti-symmetric matrices on $\mathbb{R}^r$. Recall that ∇ induces a connection on all tensor bundles of the form $E \otimes \cdots \otimes E \otimes E^* \otimes \cdots \otimes E^*$. In particular,

$$(\nabla g)(\sigma, \tau) = d(g(\sigma, \tau)) - g(\nabla\sigma, \tau) - g(\sigma, \nabla\tau).$$

With respect to our choice of coordinates, g coincides with the standard structure on $\mathbb{R}^r$. Using these coordinates to calculate the RHS, we find

$$d(s \cdot t) - (ds + As) \cdot t - s \cdot (dt + At) = 0,$$

proving that $\nabla g = 0$. In other words, compatibility is equivalent to the condition that g is parallel.

In the Hermitian setting we analogously find $\nabla h = 0$, while in the case of $\mathrm{SL}(r, \mathbb{K})$-structures we find $\nabla\Omega = 0$, where Ω is any of the global sections of the determinant line bundle defined above.

Let us now prove that any (E, g) admits a compatible connection. Our starting point is an orthonormal trivializing atlas. We then consider the previous construction: $\nabla\sigma = \sum \nabla^i(\rho_i\sigma)$, where $\nabla^i(\rho_i\sigma) \simeq d(\rho_i s)$. We want to prove that

$$d(g(\sigma, \tau)) = g(\nabla\sigma, \tau) + g(\sigma, \nabla\tau).$$

The first term on the RHS can be written, using local coordinates,

$$g(\sum \nabla^i(\rho_i\sigma), \tau) = \sum g((d\rho_i)\sigma, \tau) + \sum \rho_i g(\nabla^i\sigma, \tau) = \sum \rho_i(ds \cdot t)$$

Likewise, the second term can be written $\sum \rho_i(s \cdot dt)$. Adding them produces $\sum \rho_i d(s \cdot t) = \sum \rho_i d(g(\sigma, \tau))$, as desired.

Remark. Above we have privileged the local point of view based on trivializations of E. This has the annoying consequence that a connection is obtained by a collection of local data on the manifold, subject to complicated local transformation rules. The theory of principal fibre bundles would allow us to alternatively present connections in terms of global $\mathfrak{g}$-valued 1-forms, ie sections of $\Lambda^1(P) \otimes \mathfrak{g}$. The price we pay is that these 1-forms live on the principal fibre bundle P rather than on the manifold, and satisfy certain global transformation rules. In any case this emphasizes the fact that principal fibre bundles, by encoding all frames simultaneously, serve to avoid any specific choice of frame as happens when we work with local trivializations.

Parallel transport. Choose a point $x \in M$, a curve $t \mapsto \gamma(t)$ in M starting at x, and any vector $\sigma(0) \in E_x$. The equation $\nabla_{\dot\gamma}\sigma = 0$ defines a linear first order system wrt the unknown σ, admitting a unique solution $t \mapsto \sigma(t) \in E_{\gamma(t)}$: this is the operation known as *parallel transport* along γ. When γ is a loop through x, this construction produces, in a linear fashion, a new vector $\sigma(1) \in E_x$, ie a map $\sigma(0) \mapsto \sigma(1)$. The datum of a connection and a loop through x thus define an isomorphism of E_x.

It is interesting to understand in detail how this isomorphism arises by looking inside each chart. The simplest case is when the transition maps g_{ij} are constant so that we can choose the trivial connection, defined locally by $\nabla \simeq d$. In this case the equation is $ds(\dot\gamma) = 0$ so, as long as γ stays within a same chart, moving along γ does not change the coordinates of σ. When it moves into the transition area, instead, the coordinates get hit by g_{ij}, thus change. At the end of the loop σ re-enters the initial chart but its coordinates have changed according to the composition of these transformations. This means that, within that chart, σ itself has changed.

Remark. The above holds true only when γ moves between different transition regions. As long as γ stays within one chart, or within the intersection of several transition regions, the co-cycle conditions on E imply that σ does not change.

In the general case the equation is $ds(\dot\gamma) + A(\dot\gamma)s = 0$, so the coordinates change not only because of the g_{ij} but also because of the equation itself.

The compatibility condition implies that this isomorphism corresponds to an action of G. Indeed, locally, the parallel condition $ds(\dot\gamma) + A(\dot\gamma)s = 0$ means that the variation of the coordinates s is contained in $\mathfrak{g}$.

The subgroup of G defined by all such isomorphisms, for all loops γ through x, is known as the *holonomy group* of the connection. Up to conjugation it is independent of the point x. The smaller G is, the stronger control we have over these isomorphisms. The bottom line is that holonomy is an inevitable consequence of doing Analysis on (non-trivial) vector bundles. The (usually necessary) presence of 0-order terms in the connection makes it all the more unescapable.

Example. It is a standard exercise to check how parallel transport acts on $\mathbb{S}^2$ along a geodesic triangle, when using the Levi-Civita connection wrt the standard metric.

One reason for interest in parallel transport is that it provides convenient frames for calculations. Indeed, by applying it to a basis $\{\sigma_1(x), \dots, \sigma_r(x)\}$ we obtain a parallel basis $\{\sigma_1, \dots, \sigma_r\}$ above γ, thus a local identification of the fibres $E_{\gamma(t)}$+connection with the standard model $\mathbb{K}^r$+trivial connection.

Achtung! In general an analogous construction of parallel sections or parallel frames over higher-dimensional submanifolds in M is not possible: this requires additional conditions on the curvature, as discussed below.

Curvature. The simplest point of view on curvature is that it measures the lack of commutativity for second derivatives. We will discuss an alternative point of view on curvature, in terms of an integrability condition, later on.

With this in mind, given ∇, it is tempting to try to define the curvature operator $F_\nabla \in \Lambda^2(\mathrm{End}(E))$ in the same way as for the Levi-Civita connection, ie setting $F_\nabla(X,Y)\sigma := (\nabla\nabla\sigma)(X,Y) - (\nabla\nabla\sigma)(Y,X) = (\nabla_X\nabla\sigma)(Y) - (\nabla_Y\nabla\sigma)(X)$. Unfortunately, this definition does not make sense because ∇ does not automatically induce a connection on $\Lambda^1(E)$: this would require a connection also on TM. We can however define F_∇ via the formula that would be a consequence of that definition, via the Leibniz rule:

$$F_\nabla(X,Y)\sigma := \nabla_X\nabla_Y\sigma - \nabla_Y\nabla_X\sigma - \nabla_{[X,Y]}\sigma.$$

One can check that this produces an element $F_\nabla \in \Lambda^2(\mathrm{End}(E))$, as desired. Specifically, it is a tensor. Clearly, the operation $\nabla \mapsto F_\nabla$ is non-linear on the space of connections, and of first order wrt $A \simeq \nabla$.

As in the case of parallel transport, local charts help us understand why, given a connection ∇, the corresponding second derivatives generally do not commute. A first obstruction might seem to arise if the vector fields do not have constant coordinates. This however is dealt with via the correction term $\nabla_{[X,Y]}$. The real reason is apparent from the following calculation.

Once again, the bottom line will be that non-commutativity is an inevitable consequence of the (usually necessary) presence of 0-order terms in the connection.

In terms of a local trivialization, $\nabla_Y\sigma$ corresponds to $Ys + (Y\lrcorner A)s$, so

$$\begin{aligned}\nabla_X\nabla_Y\sigma &\simeq XYs + X((Y\lrcorner A)s) + (X\lrcorner A)(Ys) + (X\lrcorner A)(Y\lrcorner A)s\\ &= XYs + (X(Y\lrcorner A))s + (Y\lrcorner A)(Xs) + (X\lrcorner A)(Ys) + (X\lrcorner A)(Y\lrcorner A)s\end{aligned}$$

Terms 3+4 are symmetric, so they drop under alternation. Thus

$$\nabla_X\nabla_Y\sigma - \nabla_Y\nabla_X\sigma \simeq [X,Y]s + (X(Y\lrcorner A))s - (Y(X\lrcorner A))s + [A,A](X,Y)s,$$

where

$$[A,A](X,Y) := [X\lrcorner A, Y\lrcorner A] = (X\lrcorner A)\circ(Y\lrcorner A) - (Y\lrcorner A)\circ(X\lrcorner A).$$

We thus see that the non-commutativity of second derivatives is due to $dA + [A,A]$. In other words, this calculation provides the local formula

$$F_\nabla = dA + [A,A].$$

When using local coordinates, we also use the notation F_A.

We may extend the above operation to obtain an analogue of the wedge product for endomorphism-valued 1-forms:

$$\begin{gathered}\Lambda^1(\mathrm{End}_{\mathfrak{g}}(\mathbb{K}^r)) \times \Lambda^1(\mathrm{End}_{\mathfrak{g}}(\mathbb{K}^r)) \to \Lambda^2(\mathrm{End}_{\mathfrak{g}}(\mathbb{K}^r)),\\ [A,B](X,Y) := \frac{1}{2}\left([X\lrcorner A, Y\lrcorner B] - [Y\lrcorner A, X\lrcorner B]\right).\end{gathered}$$

Changing local trivialization leads to $F_{A'} = Q^{-1}F_AQ$, which corresponds to the fact that F_∇ is globally well-defined. This calculation uses the fact that $d(Q \circ Q^{-1}) = 0$, so

$$d(Q^{-1}) = -Q^{-1}(dQ)Q^{-1}.$$

These formulae show that compatibility between the connection and the G-structure implies $F_\nabla \in \Lambda^2(\mathrm{End}_{\mathfrak{g}}(E))$.

Example. Assume E has constant transition functions g_{ij}. Consider the connection locally defined by $\nabla = d$. In this case $A_i = 0$ so the local formula shows that $F_\nabla = 0$.

If $r = 1$ then A is a locally defined $\mathbb{K}$-valued 1-form. The formula $A' = Q^{-1}dQ + Q^{-1}AQ = d\log Q + A$ shows that is not globally defined. Notice that $[A, A] = 0$, so $F_A = dA$. Also, $F_{A'} = dA' = dA$ so our formula for F_A is actually independent of the local chart used. This confirms that F_∇ is a global 2-form on M. It is clearly closed and, of course, locally exact.

Let us repeat this in the case where E is a U(1)-bundle, ie a complex line bundle endowed with a Hermitian metric h. In this case $\mathfrak{g} = i\mathbb{R}$ so $A \in \Lambda^1(i\mathbb{R})$ and $F_A = dA$, a locally exact i-valued 2-form on M. We change trivializations via 1×1 matrices $Q = e^{i\theta}$. Thus $A' = A + id\theta$. The term $d\theta$ is closed but generally not exact because θ can be multi-valued.

Remark. In the language of principal fibre bundles, curvature belongs to the space $\Lambda^2(P) \otimes \mathfrak{g}$. Notice the pattern: connections and curvatures, related to forms on M in the spaces $\Lambda^i(\mathrm{End}_{\mathfrak{g}}(E))$, lift to forms on P in the simpler spaces $\Lambda^i \otimes \mathfrak{g}$.

Our definition of curvature is very much ad hoc, but it actually fits into a more general construction. In order to introduce this let us start with a digression on the exterior differential operator d and on tensor products.

Digression. The vector bundles $\Lambda^k(M)$ have the very special property that they admit an intrinsic linear differential operator $d : \Lambda^k(M) \to \Lambda^{k+1}(M)$. This is slightly different from the notion of a connection, which would only require $d : \Lambda^k(M) \to \Lambda^1 \otimes \Lambda^k(M)$, but it still satisfies a Leibniz rule wrt the operation $\wedge$. It can be defined locally; it is then well defined globally thanks to the property $d(\phi^*\alpha) = \phi^*(d\alpha)$.

On another note, let us recall that $\mathrm{End}(V)$ contains distinguished elements (called simple, elementary or decomposable) of the form $v \otimes \alpha$, where $v \in V$, $\alpha \in V^*$. It is clear that not every element in $\mathrm{End}(V)$ is of this form, because such decomposable elements have kernel of codimension 1. It is also clear however, via a choice of basis, that such elements generate all others. In order to define a linear map on $\mathrm{End}(V)$, it thus suffices to define it (in a linear fashion) on decomposable elements.

The same holds for tensor products of the form $V \otimes \cdots \otimes V \otimes V^* \otimes \cdots \otimes V^*$, for tensor products of several vector spaces and for tensor products of vector bundles over a manifold. In particular, the choice of a local basis for each vector bundle over a chart U_i shows that any section of the tensor product can be locally written as a finite C^∞-linear combination of decomposable sections. Moreover, a partition of unity argument shows that it can also be written as a locally finite C^∞-linear combination of global, compactly supported, decomposable sections. The bottom line is that we can define a linear operator on tensors simply by defining how it acts on global decomposable sections.

We shall use the above to define a differential operator on $\Lambda^i(E) = \Lambda^i(M) \otimes E$. Choose any $\alpha \in \Lambda^i(M)$ and $\sigma \in \Lambda^0(E)$ and set

$$d_\nabla(\alpha \otimes \sigma) := d\alpha \otimes \sigma + (-1)^i \alpha \wedge \nabla\sigma.$$

This defines d_∇ on decomposable sections. By the above, it induces a linear operator $d_\nabla : \Lambda^i(E) \to \Lambda^{i+1}(E)$.

Notice that $d_\nabla = \nabla$ on $\Lambda^0(E)$. Furthermore $F_\nabla = d_\nabla \circ \nabla$, ie $F_\nabla \sigma = d_\nabla(\nabla\sigma)$. To check this, it suffices to restrict to the decomposable case $\nabla\sigma = \alpha \otimes \tau$ and show

$$\nabla_X(\alpha(Y)\tau) - \nabla_Y(\alpha(X)\tau) - \alpha([X,Y])\tau = d_\nabla(\alpha \otimes \tau)(X,Y),$$

which is a simple calculation. This legitimates writing $F_\nabla = d_\nabla \circ d_\nabla$. This shows that $d_\nabla \circ d_\nabla$ is generally not zero. However, the *Bianchi identity* shows that $d_\nabla F_\nabla = 0$, for any connection.

Remark. Our presentation mentions two unexpected consequences of doing analysis on manifolds via a connection: holonomy and non-commutativity of derivatives. It turns out that these are two sides of a same coin: the holonomy group is a Lie group and its Lie algebra is generated by the endomorphisms defined by curvature, ie of the form $F_A(X,Y) \in \mathrm{End}_{\mathfrak{g}}(E)$.

3 Chern-Weil theory

Chern-Weil theory is perhaps not part of gauge theory, properly understood: it concerns only connections, the gauge group and the Yang-Mills functional play no role. However, it is an eye-opener. It provides perhaps the most elementary indication that connections incorporate fundamental information about the underlying geometry. In this sense it provides a toy model for gauge theory itself.

Characteristic classes. Given M, the basic problem concerning vector bundles is to classify the possible vector bundles over M, up to equivalence. This requires good ways of detecting non-equivalent vector bundles, thus justifying the search for invariants. The basic idea is to extract from the data which defines a vector bundle some simpler quantities which can be used to distinguish non-equivalent bundles. In general, one attempts to package this simplified data into some recognizable algebraic or geometric form. Some such invariants are trivial, such as the rank r of the vector bundle: this is a numerical invariant. Notice that its pointwise analogue, dimension, provides a complete invariant for vector spaces. Other invariants are more sophisticated. We are interested in a class of invariants known as *characteristic classes*, which take the shape of cohomology classes on M.

The basic intuition underlying characteristic classes is that they are a measure of the "twisting" of the vector bundle, as determined by the transition maps g_{ij}. Alternatively, they encode an obstruction to the existence of pointwise linearly independent sections, as above. In particular, characteristic classes vanish for trivial bundles.

Characteristic classes vary depending on the choice of coefficients $\mathbb{K}$. In any case, they are required to obey a certain set of axioms which, one proves, suffice to uniquely characterize them. For each $\mathbb{K}$ there exist several equivalent definitions, eg in terms of algebraic topology, universal classifying spaces, Chern-Weil theory, etc. This explains one reason why characteristic classes are so important: each definition leads to specific relationships and constraints between topology and the corresponding branch of geometry. The fact that the various definitions lead to the same objects is due to the fact that they satisfy the required axioms.

Ad hoc constructions. In order to explain the geometric intuition more precisely, let us consider the Euler characteristic class of orientable real vector bundles. We will actually restrict our attention to the simplest case $r = 2$ following Bott-Tu, page 70. As we will see, this will suffice for several further purposes and has the advantage of being fairly transparent in regards to the idea of simplifying the data of the vector bundle so as to extract information on its twisting.

Let $E \to M$ be an orientable rank 2 real vector bundle. Choose a partition of unity ρ_i on M subordinate to the trivializing atlas $\{U_i\}$, and a Riemannian metric on E. For each $x \in U_i$ we thus obtain coordinates r_i, θ_i on $\{x\}\times(\mathbb{R}^2\setminus\{0\})$. Let us assume that the trivializing atlas is defined by local ON sections so that $g_{ij} = g_{ij}(x) \in \mathrm{SO}(2)$, and is thus equivalent to angles $\theta_{ij} = \theta_{ij}(x)$. By definition, on $U_i \cap U_j$ we obtain the relationship $\theta_i = \theta_{ij} + \theta_j$. This is the "twisting" we have been referring to.

Our next task is to package this data into something recognizable. To do so, notice that another change of atlas leads to the relationship $\theta_{ij} + \theta_{jk} = \theta_{ik}$ on $U_i \cap U_j \cap U_k$. At this point, however, it is important to pause and remember that anything we do wrt angles is intrinsically ill-defined: angles are only well-defined up to multiples of 2π. We can address this problem by taking derivatives, ie working in terms of 1-forms, again defined on $U_i \cap U_j \cap U_k$: we then obtain the "cocycle condition"

$$d\theta_{ij} + d\theta_{jk} = d\theta_{ik}.$$

Now notice that, for any l, the 1-form $\rho_l d\theta_{il}$, although a priori defined only on $U_i \cap U_l$, extends smoothly to 0 when we move off the support U_l of ρ_l. In particular, it extends to U_i. This implies that $\xi_i := -\sum \rho_l d\theta_{il}$ is a well-defined 1-form on U_i. A simple calculation using the cocycle condition and the fact $\theta_{ij} = -\theta_{ji}$ then shows that $d\theta_{ij} = \xi_j - \xi_i$ on $U_i \cap U_j$. This implies that $d\xi_i = d\xi_j$ on $U_i \cap U_j$, so it defines a global 2-form on M. It is clear that this form is closed, so by general theory (but made clear also from the above) it is locally exact; however, it is not necessarily globally exact. Its de Rham cohomology class (usually normalized to $\frac{1}{2\pi} d\xi_i$, leading to integer coefficients) is the *Euler class* $e(E) \in H^2(M)$.

In summary, we have simplified and manipulated the transition data which defined the vector bundle to obtain a completely different object: a cohomology class. If two vector bundles are isomorphic, their Euler classses must coincide.

More generally, the Euler class of an orientable real vector bundle E of rank r is an element $e(E) \in H^r(M; \mathbb{Z})$.

Remark. It is interesting to examine the above construction from the point of view of Cech cohomology theory. Abstractly, the data $\{d\theta_{ij}\}$ is an example of a "Cech 1-cocycle" in the sheaf of 1-forms on M, relative to the given atlas. The ξ_i are an example of a "Cech 0-cocyle" for that sheaf. The relation $d\theta_{ij} = \xi_j - \xi_i$ shows that the Cech cohmology class of $\{d\theta_{ij}\}$ vanishes, so it is not a useful object. This justifies why we shift our attention to $d\xi_i$ and to a different cohomology theory.

The construction of ξ_i is itself interesting. The sheaf of 1-forms is an example of a "fine sheaf". In particular it is acyclic, ie the higher cohomology groups vanish. The construction above is the standard trick to prove this, by explicitly showing that any cohomology class of degree at least 1 is trivial.

The above construction thus incorporates ideas and mechanisms of sheaf theory. In the case of complex line bundles this is emphasized by an alternative, more common, construction of $c_1(E)$ in terms of a long exact sequence between certain sheaf cohomology groups.

The alternative intuition, of characteristic classes as obstructions to the existence of pointwise linearly independent sections, is perhaps more transparent from the dual point of view of homology classes. Consider the case $E = TM$, where M is a smooth oriented manifold. A generic section of E, ie a tangent vector field, vanishes in isolated points. Using appropriate signs, these point create a 0-dimensional homology class on M which turns out to be independent of the vector field. Its dual in $H^n(M; \mathbb{Z})$ is the Euler class of TM. In particular, this shows that the existence of a nowhere-vanishing vector field implies that the Euler class vanishes. In order to generalize this in a uniform way to other homology dimensions it is necessary to use $\mathbb{Z}_2$ coefficents. The locus of points in M where k generic vector fields fail to be linearly independent then defines a homology class which, dualized, produces the "Stiefel-Whitney classes" in $H^*(M; \mathbb{Z}_2)$.

Now consider complex vector bundles. A complex line bundle E provides a special case of an orientable rank 2 vector bundle. In this case, the Euler class is also known as the *first Chern class* $c_1(E) \in H^2(M; \mathbb{Z})$. We can define the first Chern class of any complex vector bundle to be the first Chern class of the corresponding complex determinant line bundle.

More generally, given any complex vector bundle $E \to M$, there also exist higher Chern classes $c_i(E) \in H^{2i}(M;\mathbb{Z})$. Chern classes vanish for $i > r$, and the $i = r$ case coincides with the Euler class of E.

Once again, in speciific contexts one can sometime describe these classes very geometrically. Consider the case of $c_2(E)$, where E is a complex rank 2 SU(2)-bundle over $\mathbb{S}^4$. In terms of the trivializing atlas given by two hemispheres intersecting along an open neighbourhood of $\mathbb{S}^3$, the transition map can be reduced to $g_{12} : \mathbb{S}^3 \to \mathrm{SU}(2) \simeq \mathbb{S}^3$, and $c_2(E) \in H^4(M;\mathbb{Z})$ corresponds precisely to the degree of this map, see [**?**] p.40. As already mentioned, this coincides with $e(E)$.

The Chern-Weil construction. It is clear from the examples above that geometrically intuitive definitions of a characteristic class rely on very ad hoc constructions. Chern-Weil theory provides a uniform construction for certain classes of characteristic classes. In this case, information about the bundle is extracted through an auxiliary choice of a connection on the bundle. More specifically, via the curvature of this connection.

We will focus on the Chern-Weil construction of Chern classes for a complex vector bundle $E \to M$. Our main goal will be to define the *Chern-Weil homomorphism* between algebras

$$c : \mathcal{I}(\mathfrak{g}) \to H^*(M).$$

This requires the following linear-algebraic digression.

Digression. Given a vector space V, recall the standard correspondence $Q \leftrightarrow \varphi$ between quadratic forms Q and bilinear symmetric forms φ:

$$Q(v) = \varphi(v, v), \quad \varphi(v, w) = 1/2(Q(v + w) - Q(v) - Q(w)).$$

Some examples:

1. On $\mathbb{R} = \{x\}$:

$$Q(x) = x^2 \leftrightarrow \varphi(x_1, x_2) = x_1x_2.$$

2. On $\mathbb{R}^2 = \{(x, y)\}$:

$$Q((x, y)) = x^2 \leftrightarrow \varphi((x_1, y_1), (x_2, y_2)) = x_1x_2,$$
$$Q((x, y)) = xy \leftrightarrow \varphi((x_1, y_1), (x_2, y_2)) = 1/2(x_1y_2 + x_2y_1).$$

More generally, on $\mathbb{R}^n$ (using the variable x) there exists a correspondence between k-homogeneous polynomials $p(x)$ and k-multilinear symmetric functions $\tilde{p}(x_1, \dots, x_k)$. It can be extended to all vector spaces as follows.

A *polynomial* on V is a function $V \to \mathbb{K}$ which, in terms of some basis, can be expressed as a polynomial of the corresponding coordinates. We say that it is k-homogeneous if this is true in terms of the basis. These concepts are independent of the specific basis.

There exists a 1:1 correspondence between such polynomials and the symmetric k-multilinear maps on V. Specifically, the correspondence is given by the *polarization* formula

$$\tilde{p}(v_1, \dots, v_k) := \frac{(-1)^k}{k!} \sum_{j=1}^{k} \sum_{i_1 < \cdots < i_j} (-1)^j p(v_{i_1} + \cdots + v_{i_j}).$$

We can recover p by restricting to the diagonal: $p(v) = \tilde{p}(v, \dots, v)$.

Now let G be a Lie group. Let $\mathfrak{g}$ denote its Lie algebra. In this case it is interesting to restrict our attention to ad-invariant functions. We will denote the algebra generated by homogeneous ad-invariant polynomials by $\mathcal{I}(\mathfrak{g})$. The corresponding multilinear maps are also ad-invariant.

We are mostly interested in the case where $\mathfrak{g}$ is a subalgebra of $\mathrm{gl}(r, \mathbb{K})$. Examples of homogeneous ad-invariant functions are then $\mathrm{tr}(M)$, $\det(M)$. More generally, the functions $p_k = p_k(M)$ determined by the equality $\det(I+M) = 1 + p_1(M) + \cdots + p_r(M)$ are clearly ad-invariant. They are homogeneous polynomials of degree k and they generate the algebra $\mathcal{I}(\mathfrak{g})$.

One can check that $p_k(M) = \sum_I \det M_{I,I}$, where I is a multi-index of order k and $M_{I,I}$ is extracted from M by keeping only the rows and columns with position I. This is clear in the special case where M is diagonal. In particular, $p_1(M) = \mathrm{tr}(M)$ and $p_r(M) = \det(M)$.

Example. Consider $p_2 = \det$ on $\mathrm{gl}(2, \mathbb{C})$. The polarization formula yields

$$\widetilde{\det}(M_1, M_2) = \frac{1}{2}(-\det(M_1) - \det(M_2) + \det(M_1 + M_2)).$$

Conversely,

$$\begin{aligned} \det(M_1 + M_2) &= \widetilde{\det}(M_1 + M_2, M_1 + M_2) \\ &= \widetilde{\det}(M_1, M_1) + \widetilde{\det}(M_2, M_2) + 2\widetilde{\det}(M_1, M_2). \end{aligned}$$

An analogous formula holds for $\det(M_1 + \cdots + M_m)$.

The reason we are interested in this correspondence is that multilinear maps are easier to define and compute than polynomials: multilinearity allows us to focus on how the function acts on elements of a basis. We can apply this as follows.

Let V be a vector space. Choose a polynomial p of degree k on $\mathfrak{g}$. Given any $\mathfrak{g}$-valued s-form on V, we want to apply the polynomial to obtain a $\mathbb{K}$-valued form on V. We shall do this in two steps.

1. Assume the form is decomposable, ie we can write it as $\alpha \otimes M \in \Lambda^s(V) \otimes \mathfrak{g}$. Given that p acts on $\mathfrak{g}$ in a homogeneous fashion, in order to obtain a well-defined operation on the tensor product $\alpha \otimes M$ we need to perform some operation on

α which has the same homogeneity. The natural choice is to use the wedge product. We thus obtain the object $c_p(\alpha\otimes M) := p(M)\alpha^k$: it is a ks-form on V.

2. In general the form will be a linear combination of decomposable forms $\alpha_1\otimes M_1,\dots,\alpha_m\otimes M_m$. In this case we can rely on the multilinearity of $\tilde{p}$, as follows. Set

$$c_p(\sum_{i=1}^m \alpha_i\otimes M_i) := \sum_I \tilde{p}(M_{i_1},\dots,M_{i_k})(\alpha_{i_1}\wedge\cdots\wedge\alpha_{i_k}),$$

where on the RHS we sum over all multi-indices $I=(i_1,\dots,i_k)$. This can be computed in terms of p via polarization. We have thus achieved our goal.

To conclude this digression, notice that more generally we can set

$$c_p(\sum_{i=1}^m \alpha_i^1\otimes M_i^1,\dots,\sum_{i=1}^m \alpha_i^k\otimes M_i^k)) := \sum_I \tilde{p}(M_{i_1}^1,\dots,M_{i_k}^k)(\alpha_{i_1}^1\wedge\cdots\wedge\alpha_{i_k}^k),$$

to obtain a map

$$c_p:\Lambda^s(V)\otimes\mathfrak{g}\times\cdots\times\Lambda^s(V)\otimes\mathfrak{g}\to\Lambda^{ks}(V).$$

Coming back to Chern-Weil theory, let E be a G-vector bundle over M. Choose a *ad*-invariant homogeneous polynomial p on $\mathfrak{g}$. Any local G-basis for E allows us to identify E_x with $\mathbb{K}^r$, thus $\mathrm{End}_{\mathfrak{g}}(E_x)$ with $\mathfrak{g}$. We can then apply p to any $\phi\in\mathrm{End}_{\mathfrak{g}}(E)$: the *ad*-invariance ensures that the result does not depend on the choice of G-basis. We will write the result as $p(\phi)$: it defines a function on M.

More generally, the above digression shows how to evaluate p on any element in $\Lambda^s(M)\otimes\mathrm{End}_{\mathfrak{g}}(E)$, obtaining a ks-form on M. In particular, choose a G-compatible connection on E and let $F_\nabla\in\Lambda^2(\mathrm{End}_{\mathfrak{g}}(E))$ denote its curvature. The above process produces, for any p of degree k, a form $c_p(F_\nabla)\in\Lambda^{2k}(M;\mathbb{C})$.

The Bianchi identity shows that it is closed, and one can check that changing the connection implies that it changes by an exact form. The result is that the corresponding cohomology class $[c_p(F_\nabla)]\in H^{2k}(M;\mathbb{C})$ is independent of all choices. Wrt the variable p we thus obtain the Chern-Weil homomorphism associated to the bundle E.

Example. We obtain better control over the coeffiicients when $G=\mathrm{U}(r)$. In this case we can normalize the classes c_p so as to obtain $\mathbb{R}$-valued forms.

Specifically, set $c_k := [\frac{i}{2\pi}c_{p_k}(F_\nabla)]$, where p_k are the canonical invariant polynomials defined above. These *Chern classes* are $\mathbb{R}$-valued. Since p_k generate $\mathcal{I}(\mathfrak{g})$, the Chern-Weil homomorphism becomes a map $c:\mathcal{I}(\mathfrak{u}(r))\to H^*(M;\mathbb{R})$. Notice that any complex bundle admits a Hermitian metric, ie a $\mathrm{U}(r)$-structure, so $\mathbb{R}$-valued Chern classes are always possible. The choice of metric will however influence the specific connection used in the construction.

It is sometimes useful to consider the characteristic classes associated to other $p \in \mathcal{I}(\mathfrak{g})$. Consider for example the case $p(M) := \mathrm{tr}(M^2)$, leading to $\mathrm{tr}(F^2_\nabla) \in \Lambda^4(M)$. It has the property that

$$\frac{1}{8\pi^2}[\mathrm{tr}(F^2_\nabla)] = c_2(E) - \frac{1}{2}c_1(E)^2 \in H^4(X;\mathbb{R}).$$

This is very relevant to the Yang-Mills functional, defined below, whose integrand is related to the LHS.

Achtung! There is an important, though subtle, difference between characteristic classes provided by Chern-Weil theory and those built using other definitions. Topological definitions lead to coefficients in $\mathbb{Z}$, while Chern-Weil theory generally leads to coefficients in $\mathbb{R}$. The relationship between these classes is provided by the map $H^*(M;\mathbb{Z}) \to H^*(M;\mathbb{R})$, which however is sometimes not injective, so there is some loss in information. The problem lies in the fact that torsion classes in $H^*(M;\mathbb{Z})$ are killed in $\mathbb{R}$. If the topology of M does not allow for torsion classes in cohomology, eg Riemann surfaces, this is no problem. In general, however, the fact $c_1(E) = 0$ proved via Chern-Weil theory might not imply that the integral class $c_1(E)$ vanishes. In particular, the flatness of a complex line bundle does not necessarily imply that it is trivial.

Applications. Characteristic classes are the key-stone for achieving a deeper understanding of various topological constraints on geometry, such as the classical Poincaré-Hopf and Gauss-Bonnet theorems. These results also serve to emphasize the power of having multiple viewpoints on characteristic classes. Let us review them.

Poincaré-Hopf. Recall that, depending on the choice of homology theory, one can define the Euler characteristic $\chi(M)$ of a compact manifold M as the alternating sums of the ranks of either the singular or the simplicial homology groups (the latter defined via a triangulation). This number depends only on the topology of the manifold, not on any additional smooth structure.

The Poincaré-Hopf theorem for a smooth compact oriented manifold M states that $\chi(M)$ coincides with the sum of indices of any vector field on M with isolated, transverse, zeroes.

The idea of the proof is to first show that the sum of indices is independent of the particular vector field. This is true because each index (defined as the degree of an induced map between spheres) coincides with the intersection index of X with the zero section Z, viewed as submanifolds in TM, so the sum of indices is the intersection number of X and Z in TM. (Alternatively, integrating the vector field produces a diffeomorphism, and the sum of indices is the intersection number of the submanifolds in $M \times M$ defined by the graphs of the identity map and of this diffeomorphism). We can then tie this number to triangulations by showing that any triangulation on M induces a vector field with one zero of index ± 1 on each facet of the triangulation. For example, in dimension 2 we can find a vector field with a sink for each vertex, a source for each face, and a saddle for each side. The sum of indices of this specific vector field is thus the Euler characteristic. (Alternatively, we can use the gradient vector field of any Morse function and the Morse equality).

We can summarize this using characteristic classes as follows. Recall that the Euler class $e(TM)$ coincides with the top Stiefel-Whitney class, which is the dual of the 0-cycle defined by the zeroes of a generic vector field. The sum of these zeroes is thus $\int_M e(TM)$, and the Poincaré-Hopf theorem states that $\int_M e(TM) = \chi(M)$.

The interest in this theorem lies in the fact that, by definition, $\chi(M)$ depends only the on the topology of M, not on any smooth structure. Vector fields and the Euler class depend instead on the tangent bundle, thus on the smooth structure. We have thus showed that a certain aspect of any smooth structure on M is constrained by the topology: $\mathbb{S}^2$ cannot have a nowhere-vanishing vector field, $\mathbb{T}^2$ can.

Remark. Intersections can be further formalized as follows. The *Thom class* of a rank r oriented vector bundle E is a real cohomology class $\Theta(E) \in H^r(E)$, defined by choosing a generator for the compactly supported top-dimensional cohomology of each fibre, ie a compactly supported volume form on each E_x with integral 1. An interesting case is when E is the normal bundle of an embedded oriented submanifold $M \subset N$. In this case one can check that, for any form ω on N, $\int_M \omega = \int_N \omega \wedge \Theta(E)$, which is the formula defining the Poincaré dual $\eta(M)$: in other words, the Thom class coincides with the Poincaré dual of the homology cycle defined by the submanifold. It follows that the Poincaré dual of the intersection of two submanifolds in N is the wedge product of the Thom classes of their normal bundles.

For any bundle, it turns out that $e(E) := Z^*(\Theta(E))$, where Z is the zero section. Thus $\int_M e(TM) = \int_M Z^*(\Theta(TM)) = \int_Z \Theta(TM) = \int_{TM} \Theta(TM) \wedge PD(Z) = \int_{TM} PD(Z \cdot Z)$, where we use the fact that the normal and tangent bundles of Z in TM are isomorphic. This confirms that $\int_M e(TM)$ coincides with the self-intersection number of Z in TM (more generally, with the intersection number of any two vector fields, because they are all homologous).

Gauss-Bonnet. Chern-Weil theory provides a direct path towards an intricate network of relationships between topology and curvature. The classical Gauss-Bonnet theorem is one of its simplest manifestations.

Given a compact oriented smooth surface M, the theorem states that, for any metric, the integral of the Gaussian curvature K coincides with $2\pi\,\chi(M)$.

This result is very natural from the Chern-Weil theory viewpoint: one shows that the class $e(TM)$ is represented by the 2-form $\frac{1}{2\pi} K \,\mathrm{vol}_g$ defined using the given metric on the bundle TM, so $\frac{1}{2\pi}\int_M K\,\mathrm{vol}_g = \int_M e(TM) = \chi(M)$.

4 Gauge theory: basic ingredients

Up to now we have discussed connections from the point of view of their use in Analysis, ie in providing a higher-order calculus on manifolds, and in topology. In particular, we have highlighted the existence in each context of infinite possible connections. As already mentioned, in each specific geometric context it is natural to look for a “best choice”. This is sometimes done algebraically, as in the Levi-Civita case, by defining a unique connection in terms of extra properties.

Gauge theory proper begins with a different approach to this problem, based on the Yang-Mills functional.

Gauge transformations. It is a fact of life that vector bundles have a huge set of automorphisms. Specifically, the *gauge group* of E is the infinite-dimensional Lie group of sections of the bundle $\mathrm{Aut}(E)$. This group has a natural left action on the space of connections on E:

$$(g \cdot \nabla)\sigma := g(\nabla(g^{-1}\sigma)).$$

Locally, this action follows rules similar to the changes of trivialization seen above (via the identification $g \simeq Q^{-1}$):

$$g \cdot A = gAg^{-1} - (dg)g^{-1}, \quad F_{g\cdot A} = g(F_A)g^{-1}.$$

We can take compatibility conditions into account by restricting to sections of $\mathrm{Aut}_G(E)$.

Connections related this way are considered equivalent: anything we do with them leads to the same result, up to the group action. In other words, the gauge group generates a huge amount of redundancy within the space of connections.

Achtung! The previous paragraph makes sense only after we have specified which gauge group we are interested in. Below, we will meet the problem of fixing a complex gauge group orbit of (partial) connections and a metric, then looking for a (partial) connection (whose Chern connection is) most compatible with the metric. In this situation we shall consider equivalent only those connections which are compatible with a same metric, not those which belong to the same complex orbit.

Geometrically interesting conditions on a connection are also generally preserved by the gauge group action, so the natural moduli spaces in this context are aways infinite-dimensional. We can remedy this fact by quotienting by this action, ie working on the space of orbits. Given a connection ∇, this is locally equivalent to restricting one's attention to a “slice”, ie a submanifold in the space of connections which is transverse to the group orbit. If we parametrize the set of all connections via $A \mapsto \nabla + A$, a slice can be obtained by restricting to those A which satisfy a condition similar to the “Coulomb gauge” condition introduced by Uhlenbeck.

Remark. Group actions, orbits and slices are omni-present in geometry and beyond.

In linear algebra, an analogous situation is given by the set of symmetric bilinear forms on a vector space V, under the action of the group $\mathrm{Aut}(V)$. Sylvester's theorem shows that the space of orbits is discrete, classified by the signature. In this case any transversal slice has dimension 0.

A more geometric example is $\mathbb{S}^2$, acted upon by the group of rotations around the z-axis. In this case the space of orbits is 1-dimensional, and a slice at a point $x \in \mathbb{S}^2$ is given by any curve in $\mathbb{S}^2$, through x, transverse to the horizontal plane.

An analytic example is provided by Hodge theory on a compact Riemannian manifold. Here, we consider the space of closed k-forms, acted upon via translation by the Abelian group of exact forms. De Rham cohomology is the space of orbits. A slice is provided by the harmonic k-forms, which allow us to change cohomology class. In this example there also appears an energy functional, similar to the Yang-Mills functional below. We find the harmonic k-forms by minimizing this functional within a given cohomology class.

The Yang-Mills functional. In Riemannian geometry one chooses a homology class, considers the infinite-dimensional space of all submanifolds representing that class and the corresponding volume functional. The critical points are the minimal submanifolds. One may be particularly interested in the stable critical points, or in volume-minimizing submanifolds.

We now want to set up an analogous framework for studying the infinite-dimensional space of connections on a vector bundle. We start with the following digression.

Digression. We shall need a metric on $\Lambda^2(\mathrm{End}_{\mathfrak{g}}(E))$. We can obtain it as the tensor product of metrics on $\Lambda^2(M)$ and on $\mathrm{End}_{\mathfrak{g}}(E)$, defined as follows.

1. Assume (V, g) is a Euclidean vector space. We obtain a metric on $V \otimes \cdots \otimes V$ by setting $g(v_1 \otimes \cdots \otimes v_k, w_1 \otimes \cdots \otimes w_k) := g(v_1, w_1) \ldots g(v_k, w_k)$ on decomposable elements, then extending by linearity. We obtain a metric on $\Lambda^k(V)$ by setting $g(v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_k) := \det(g(v_i, w_j))$ on decomposable elements, then extending by linearity. Dually, we endow V^* with the metric such that $V \to V^*$, $v \mapsto g(v, \cdot)$ is an isomorphism. We like-wise obtain a metric on the dual tensor products and k-forms by dualizing the metrics above.

Notice that these definitions clash with certain other conventions. For example, if e_i is orthonormal, the above implies that $e_i \wedge e_j$ has length 1 (corresponding to the natural area of the corresponding parallelogram). Like-wise, it implies $\mathrm{vol}_g := e_1^* \wedge \cdots \wedge e_n^*$ has length 1. With this choice, standard identifications such as $v \wedge w = \frac{1}{2}(v \otimes w - w \otimes v)$ or $v \wedge w = v \otimes w - w \otimes v$ are not isometries.

Applying this construction pointwise we obtain a metric on $\Lambda^2(M)$, where M is a Riemannian manifold.

2. We now want a metric on $\mathrm{End}_{\mathfrak{g}}(E)$. Again we shall define it pointwise, using a different process which provides good invariance properties.

Let V be a vector space. There exists a canonical symmetric bilinear form on $\mathrm{End}(V)$ defined by $\varphi(\phi,\psi) := \mathrm{tr}(\phi\circ\psi)$.

In order to understand its signature it is useful to choose a metric on V. This yields a notion of adjoint operators, thus a metric $(\phi,\psi)\mapsto \mathrm{tr}(\phi^T\circ\psi)$. The two forms coincide on the subspace of selfadjoint endomorphisms and have opposite sign on the subspace of anti-selfadjoint operators. This shows that our form φ is non-degenerate, of mixed signature.

Any injective representation $G\to\mathrm{Aut}(V)$ leads to a representation $\mathfrak{g}\to\mathrm{End}(V)$, identifying $\mathfrak{g}$ with the subspace we have denoted $\mathrm{End}_{\mathfrak{g}}(V)$.

We are interested in the signature of the restriction of φ to $\mathrm{End}_{\mathfrak{g}}(V)$. Once again, we can study this via a metric on V. If G is compact, by averaging wrt the G-action we may assume this metric is G-invariant, ie G acts on V by isometries. This implies that the derivatives of this action, ie the endomorphisms in $\mathrm{End}_{\mathfrak{g}}(V)$, are anti-selfadjoint. As above, it follows that the restriction of φ is negative definite.

In summary: when G is compact, $-\varphi$ provides a canonical invariant metric on $\mathrm{End}_{\mathfrak{g}}(V)$.

Remark. Within the framework of principle fibre bundles we identify $\mathrm{End}_{\mathfrak{g}}(E)$ with the bundle $ad(P) = P\times_\rho\mathfrak{g}$. In this language our metric can be obtained starting from the adjoint representation $\mathfrak{g}\to\mathrm{End}(\mathfrak{g})$, $X\mapsto[X,\cdot]$. The Jacobi identity is equivalent to saying that this map is a Lie algebra homomorphism. In this case the restriction of our form φ coincides, up to identifications, with the *Killing form* $K(X,Y) := \mathrm{tr}([X,\cdot])\circ[Y,\cdot])$ on $\mathfrak{g}$.

In general the restriction of a non-degenerate form may become degenerate but one can prove that, if G is semi-simple, K (thus the restriction of φ) is non-degenerate. An argument similar to the above proves that if G is compact then the Killing form is negative-definite.

Let (M,g) be a compact oriented Riemannian manifold. Let E be a G-vector bundle over M, where G is compact. Consider the space of all connections on E compatible with the G-structure. The *Yang-Mills functional* is defined by

$$\nabla\mapsto YM(\nabla) := \int_M \|F_\nabla\|^2\,\mathrm{vol}_g,$$

where $\|F_\nabla\|$ is obtained using the metric on the tensor bundle $\Lambda^2(\mathrm{End}_{\mathfrak{g}}(E))$ defined as in the digression, above. These choices imply that the Yang-Mills functional is gauge-invariant wrt to sections of $\mathrm{Aut}_G(E)$.

Remark. The functional is also invariant wrt sections of $\mathrm{Aut}(E)$. This action changes the specific G-structure in question, relating the corresponding spaces of G-connections. In this sense the functional notices only that the connections are compatible wrt some G-structure, it doesn't care which.

The role of the gauge group is one of the big differences between the theory of submanifolds, mentioned above, and the theory of connections.

Now recall that on any n-dimensional oriented Euclidean vector space V there exists a canonical operator $\star : \Lambda^k(V) \to \Lambda^{n-k}(V)$, known as the *Hodge star*. It is defined by imposing the condition $\alpha \wedge \star\beta = g(\alpha, \beta)\,\mathrm{vol}_g$. Using the Hodge star operator we can equivalently write $YM(\nabla) = \int_M F_\nabla \wedge \star F_\nabla$, where we hide the role of the norm on $\mathrm{End}_{\mathfrak{g}}(E)$.

We can linearize the functional at ∇ by choosing $A \in \Lambda^1(\mathrm{End}_{\mathfrak{g}}(E))$, thus a new connection $\nabla + A$, and writing

$$F_{\nabla+A}(X,Y)\sigma = (d_{\nabla+A})_X(d_{\nabla+A})_Y\sigma - (d_{\nabla+A})_Y(d_{\nabla+A})_X\sigma - (d_{\nabla+A})_{[X,Y]}\sigma.$$

We then find

$$F_{\nabla+A} = F_\nabla + d_\nabla A + [A, A],$$

where, as usual, $(d_\nabla A)(X,Y) = \nabla_X(Y \lrcorner A) - \nabla_Y(X \lrcorner A) - [X,Y] \lrcorner A$. Thus

$$\frac{d}{dt}YM(\nabla + tA)_{|t=0} = 2\int_M (F_\nabla, \frac{d}{dt}F_{\nabla+tA|t=0})\,\mathrm{vol}_g = 2\int_M (F_\nabla, d_\nabla A)\,\mathrm{vol}_g,$$

so the Euler-Lagrange equation is the second order system $d_\nabla^* F_\nabla = 0$, where d_∇^* denotes the adjoint operator on 2-forms. This equation should be coupled with the Bianchi identity $d_\nabla F_\nabla = 0$ to obtain ellipticity (modulo the gauge group action). We will refer to solutions of these equations as *Yang-Mills connections.*

As usual, one can check that $d_\nabla^* = (-1)^{n-1} \star d_\nabla \star$. With this formulation the notion of Yang-Mills connections extends to cases where the functional is not defined, eg non-compact groups G.

Example. Assume E is a U(1)-line bundle over M. In this case F_∇ is an i-valued 2-form and the relevant operator d_∇ coincides with the standard operator d. The Yang-Mills functional is the standard energy functional and ∇ is a Yang-Mills connection iff its curvature is harmonic in the usual sense.

The gauge group is Abelian. An element $g = e^{i\theta}$, where $\theta = \theta(x)$, acts on connections as follows: $g \cdot A = g^{-1}dg + g^{-1}Ag = A + id\theta$. This action is not trivial on connections, but it induces a trivial action on curvature.

Remark. The above example shows that Yang-Mills theory can be seen as an extension to higher-rank vector bundles of standard Hodge theory in degree 2. Like-wise, when E is a trivial line bundle then any connection is of the form $d + A$, where A is a global 1-form. In this sense the theory of connections is an extension to higher-rank, non-trivial, vector bundles of the standard theory of 1-forms.

Remark. The formula above shows what happens to curvature under the perturbation $\nabla \mapsto \nabla + A$ of a given connection ∇. Is the new connection gauge-equivalent to the initial one? Our formulae provide an important tool: equivalence implies that, pointwise, their curvature tensors are conjugation-equivalent.

This shows the relevance of the linear-algebraic problem of understanding the conjugation classes of endomorphisms on a vector space. In the Abelian case, eg line bundles, the two curvature tensors must actually coincide.

5 Flat connections

A connection is *flat* if $F_A = 0$. Notice that this condition is gauge-invariant. Our interest in such connections stems from the fact that, in the appropriate context, they are the simplest example of a Yang-Mills connection. Furthermore, they are clearly absolute minimizers of the functional. The discussion below holds however for any G and M.

We have already met examples of flat connections on vector bundles with constant transition maps g_{ij}, defined by the choice $A_i = 0$ (in each chart). It turns out that, in a certain sense, all examples are of this type. Before explaining this, a digression.

Digression. Assume given a topological manifold. Recall that one usually associates to it its maximal atlas. Charts are related by transition maps which are homeomorphisms. A differentiable structure is then a subatlas whose transition maps are diffeomorphisms. It sometimes turns out that one can find several distinct such subatlases, leading to the notion of exotic structures on the same topological manifold.

This construction reappears in several other contexts. A similar picture arises for example in complex geometry: a holomorphic structure is a subatlas whose transition maps are biholomorphisms.

Now consider a given vector bundle with a maximal trivializing atlas with smooth transition maps. One can try to extract a trivializing atlas whose transition maps are constant. This is known as a *local system*. This is not always possible: we will see obstructions, below. When it is, it sometimes turns out that one can do it in several distinct ways. In other words, the same vector bundle may support several different local systems.

Sheaf theory provides perhaps the best language for this notion, but at the cost of increased technicalities. We shall avoid it.

Given a G-vector bundle E, flatness has an interesting characterization in terms of the frame bundle P. As already mentioned, the datum of a connection translates, in terms of P, into a $\mathfrak{g}$-valued 1-form on P with certain properties. The kernel of this 1-form defines a distribution of n-planes in P, horizontal with respect to the projection $P \to M$. Flatness corresponds to the Frobenius integrability condition for this distribution: this is related to the fact that curvature is obtained by differentiating the connection 1-form. Each integral leaf in P, as x varies in M, defines local parallel frames $\{\sigma_1, \dots, \sigma_r\}$, ie local sections of P satisfying $\nabla\sigma_i = 0$. This construction has the following features.

1. With respect to such frames, the induced trivialization of E provides an identification $E_{|U_i}$+connection ∇ with the standard model $U_i \times \mathbb{K}^r$+standard

connection d. This means each $A_i = 0$.

2. If we change trivialization passing from one parallel basis to another, the parallel condition implies that the transition matrix Q is constant. Analogously, if we use these local trivializations to build a trivializing atlas the transition maps g_{ij} are constant. The connection thus defines a local system.

To summarize: on any given G-vector bundle E there exists a 1:1 correspondence between flat connections and local systems (compatible with G).

Example. Let $E \to M$ be the trivial line bundle. Choose a global non-vanishing section σ, thus an isomorphism $E \simeq M \times \mathbb{K}$. This induces a connection on E, by identification with the trivial connection d on $M \times \mathbb{K}$; equivalently, this connection can be defined by positing that the parallel sections are those which are constant multiples of σ. Since d is flat, the connection on E is also flat. The corresponding local system is given by a trivializing atlas whose transition maps are $g_{ij} = Id$.

In this context the gauge group is given by non-zero functions, acting by multiplication. Any two global non-vanishing sections are related in this way, so the corresponding connections are equivalent. We will refer to connections built this way as *canonical*.

Remark. Any connection ∇ on any vector bundle E over $\mathbb{S}^1$ is flat for dimensional reasons: $F_\nabla \in \Lambda^2(E) = 0$.

The existence of local parallel frames, thus charts in which ∇ coincides with d, can be alternatively proved via repeated applications of parallel transport, thus avoiding the language of principle fibre bundles. We refer to [**?**] for details.

The existence of a flat connection, thus of a local system, has topological consequences on E: Chern-Weil theory shows that all $\mathbb{K}$-valued characteristic classes vanish.

Example. Any complex vector bundle admitting a flat unitary connection has vanishing real Chern classes of any order. Recall that, in the case of a complex line bundle, the vanishing of the integral first Chern class $c_1(L) = 0 \in H^2(M;\mathbb{Z})$ implies that L is differentiably trivial. The vanishing of the real first Chern class is slightly weaker: it implies that the integral class $c_1(L)$ is a torsion class, ie $kc_1(L) = 0$ for some k. This implies that $c_1(L^k) = 0$, so L^k is trivial. Of course, if $H^2(M;\mathbb{Z})$ has no torsion classes (eg any Riemann surface), then $c_1(L) = 0$ so L is trivial.

Remark. The strength of the flatness condition is particularly clear in the category of complex bundles over complex manifolds: locally constant maps g_{ij} are holomorphic, so a flat connection induces a holomorphic structure on E. It in an interesting question which holomorphic structures arise this way. This is related to stability and the Narasimhan-Seshadri theorem, discussed below.

The above discussion raises the question whether, on a given E (for example E trivial), there exist non-equivalent flat connections/local systems. To answer this we need to find further invariants of connections, beyond curvature. This brings us to the concept of monodromy.

The starting point is provided by the notions of parallel transport and holonomy. In general, parallel transport depends heavily on the specific choice of curve γ. Let us assume, however, that the connection is flat. Any parallel frame along γ can then be seen as a curve inside one of the integral leaves mentioned above. Assume γ is a closed curve through $x \in M$, so that parallel transport generates an isomorphism of E_x. We have already mentioned that this isomorphism is generated by concatenating the maps g_{ij} encountered as γ moves from one trivalizing chart to another. Since these maps are locally constant, the isomorphism is clearly independent of small deformations of γ. One can also show that it is independent of γ within its homotopy class. We thus obtain a homomorphism $\pi_1(M, x) \to G$, known as the *monodromy representation*, whose image is the holonomy group. The element g, ie automorphism of E_x, associated to a specific loop γ (better: to its class) is known as the *monodromy* of that loop (or class).

Example. The canonical flat connections on a trivial line bundle have trivial monodromy because the transition maps are trivial.

It turns out that this process can be reversed. The construction is an extension of the process used in covering space theory by which we write $M \simeq \tilde{M}/\pi_1(M)$, where $\tilde{M}$ is the universal cover endowed with the usual monodromy action of $\pi_1(M)$. Indeed, choose $\rho \in \mathrm{Hom}(\pi_1(M), G)$. We then obtain an action

$$\gamma : \tilde{M} \times \mathbb{K}^r \to \tilde{M} \times \mathbb{K}^r, \quad \gamma \cdot (x, v) := (\gamma(x), \rho(\gamma)(v)),$$

where $\gamma \in \pi_1(M)$. The quotient $E := \tilde{M} \times \mathbb{K}^r/\pi_1(M)$ is a G-vector bundle over M with constant transition maps, thus a flat connection.

The space $\mathrm{Hom}(\pi_1(M), G)$ thus classifies the space of all flat G-vector bundles on M. We will denote by $\mathrm{Hom}_E(\pi_1(M), G)$ the set of representations which yield the G-bundle E. In this sense, monodromy defines a complete invariant for flat connections on E.

We may summarize as follows. Given a G-vector bundle E, there exist 1:1 correspondences between (i) equivalence classes of compatible flat connections, (ii) equivalence classes of compatible local systems, (iii) equivalence classes of homomorphisms in $\mathrm{Hom}_E(\pi_1(M), G)$. In (i) equivalence is defined in terms of sections of $\mathrm{Aut}_G(E)$, in (ii) in terms of G-vector bundle isomorphisms, in (iii) in terms of conjugation by G.

We may think of this as a correspondence between geometric, topological, and algebraic objects.

Remark. Assume two local frames $\mathcal{B}$, $\mathcal{B}'$ are parallel wrt a flat connection ∇, so that $\mathcal{B}' = \mathcal{B} \cdot Q$ for some constant $Q \in G$. Then, given any section g of $\mathrm{Aut}(E)$,

$g\mathcal{B}$, $g\mathcal{B}'$ are related by the same matrix Q and are parallel for the connection $g\nabla g^{-1}$. This shows that the monodromy of $g\nabla g^{-1}$ coincides with that of ∇. In this sense monodromy sees only that ∇ is G-compatible, but not wrt which specific G-structure.

The condition $F_A = 0$ implies that $d_A : \Lambda^i(E) \to \Lambda^{i+1}(E)$ defines a co-chain complex, so one obtains an analogue of de Rham cohomology for E-valued forms on M.

Example. We have mentioned that there are only two line bundles over $\mathbb{S}^1$. The trivial line bundle over $\mathbb{S}^1$ is obviously flat. Now consider the Möbius strip. Its description in terms of transition maps shows that these maps are constant, so it is also flat. We obtain a flat connection by positing that the frames corresponding to the local sections ± 1 over each chart are parallel. In this case the frame bundle is a connected 2:1 covering of $\mathbb{S}^1$. The existence of these two flat line bundles corresponds to the fact that $\pi_1(\mathbb{S}^1) = \mathbb{Z}$ and $\mathrm{Hom}(\mathbb{Z}, \mathbb{Z}_2) = \mathbb{Z}_2$.

Example. Assume E is a complex line bundle over a manifold whose $H^2(M;\mathbb{Z})$ has no torsion. Then any flat bundle has $c_1 = 0$, so it is trivial. Let $G := \mathrm{U}(1)$. Recall that any two Hermitian metrics on a vector space are isomorphic. The analogue holds for $\mathrm{U}(1)$-structures on E. We may thus choose $E = M \times \mathbb{C}$ endowed with the standard Hermitian structure. In this case $\mathrm{Hom}(\pi_1(M), G) = \mathrm{Hom}_E(\pi_1(M), G)$. The existence of non-canonical flat connections depends entirely on this group. Furthermore, since G is Abelian, conjugation acts trivially.

Let us consider three examples of this type.

1. Assume M is a compact surface of genus g. Then, after abelianizing the fundamental group, $\mathrm{Hom}(\pi_1(M), \mathrm{U}(1)) \simeq \mathrm{Hom}(H^1(M;\mathbb{Z}), \mathrm{U}(1)) \simeq \mathrm{U}(1)^{2g}$. This classifies all flat structures on the G-bundle E.

2. Assume $M := \mathbb{R}^n$. Unitary connections on E can be identified with i-valued 1-forms $A \in \Lambda^1(\mathbb{R}^n)$, and flat connections can be identified with closed such 1-forms. We have seen that two such connections A, A' are gauge-equivalent if $A' = A + id\theta$, for some $\theta = \theta(x)$. On the other hand, the fact $H^1(\mathbb{R}^n;\mathbb{R}) = 0$ implies $A' - A = idf$, for some $f : \mathbb{R}^n \to \mathbb{R}$. This shows that they are equivalent with respect to the choice $\theta := \pi \circ f$, where π is the standard projection $\pi : \mathbb{R} \to \mathbb{S}^1$. In other words, up to gauge equivalence there is a unique flat unitary connection on the G-bundle$E$. This corresponds to the fact that $\pi_1(\mathbb{R}^n)$ is trivial, so the only homomorphism $\pi_1(\mathbb{R}^n) \to \mathrm{U}(1)$ is trivial.

3. Assume $M := T^n$ is a n-torus. As before (but dropping the factor i), flat unitary connections are classified by closed 1-forms A. In this case $A = d\theta$ for some θ precisely when $\int_\gamma A \in 2\pi\mathbb{Z}$ for any loop γ: indeed, this assumption allows us to define θ by integration. This includes (but extends) the case $A = df$ which leads to $\theta := \pi \circ f$ as above. It follows that, up to gauge equivalence, flat connections are classified by $H^1(T^n;\mathbb{R})/(2\pi H^1(T^n;\mathbb{Z}))$, which is again a n-torus. This corresponds to the fact that $\pi_1(T^n) = \mathbb{Z}^n$ and homomorphisms $\mathbb{Z}^n \to \mathrm{U}(1) = \mathbb{S}^1$ are completely defined by the values they give to a basis, thus

are classified by $(\mathbb{S}^1)^n$. In Harmonic Analysis the group $\mathrm{Hom}(\mathbb{Z}^n, \mathbb{S}^1)$ is known as the Pontrjagin dual of $\mathbb{Z}^n$. Summarizing: the moduli space of flat unitary connections on E is again a n-torus.

Remark. We may rephrase the above by observing that any element in $H^1(T^n; \mathbb{R})$ can be represented by a constant 1-form A on $\mathbb{R}^n$. Indeed, these forms are closed and, although $A = df$ for some linear function on $\mathbb{R}^n$, this f is not well-defined on T^n so A is not exact on the torus. It follows that, if we write $T^n = V/\Lambda$ for some vector space V and lattice Λ, we can choose constant forms in V^* as representatives and identify the moduli space of flat connections with the "dual torus" $V^*/(2\pi\Lambda^*)$. Here, Λ^* is defined to be the elements in V^* which take integral values on Λ. Our correspondence thus gives an identification between the dual torus and the Pontrjagin dual of the fundamental group of T^n. Notice that the latter is very different from the Pontrjagin dual of T^n. Indeed, $\mathrm{Hom}(\mathbb{S}^1, \mathbb{S}^1) = \mathbb{Z}$ so $\mathrm{Hom}(T^n, \mathbb{S}^1) = \mathbb{Z}^n$.

A note regarding terminology: just like the canonical identification $V = V^{**}$ justifies the term duality between V and V^* because it implies that each is dual to the other, in the same way iterating the above construction provides an identification $V/\Lambda = V^{**}/\Lambda^{**}$.

Example. Let $M := \mathbb{S}^1$. As mentioned, for dimensional reasons any connection on any vector bundle must be flat. This corresponds to the statement that any vector bundle with structure group G over $\mathbb{S}^1$ can be trivialized in such a way as to make the transition maps become constant. Classifying flat connections thus corresponds to classifying all connections. Furthermore, any complex vector bundle is trivial. Let us thus assume E is the trivial complex line bundle with the standard U(1)-structure. As above, we may then conclude that the moduli space of all unitary connections on the U(1)-vector bundle E is in 1:1 correspondence with homomorphisms $\pi_1(\mathbb{S}^1) \simeq \mathbb{Z} \to \mathrm{U}(1)$, so it is isomorphic to U(1).

In this case it is also interesting to study the parallel U(1)-frames in $P = \mathbb{S}^1 \times \mathrm{U}(1)$. Any such frame can be written as $\sigma := e^{if}$, for some $f = f(\theta)$, not necessarily periodic. Given a U(1)-connection $A = ia$, where $a \in \Lambda^1(\mathbb{S}^1)$, the frame is parallel iff $(e^{if})' + iae^{if} = 0$, ie $f' + a = 0$. This shows that f' is periodic. The possible discrepancy between $\sigma(2\pi)$ and $\sigma(0)$ produces the monodromy. More specifically, we can use $\sigma(2\pi) \in \mathrm{U}(1)$ to classify the connection up to gauge equivalence. By the Leibniz rule, the other parallel sections are then of the form $\lambda\sigma$, for constant $\lambda \in \mathrm{U}(1)$.

Remark. There is an extension of the above theory concerning "projectively flat connections", ie those for which $F_\nabla = \alpha \otimes Id_E$, where $\alpha \in \Lambda^2(M)$. These connections are classified by homomorphisms from $\pi_1(M)$ into the projectivized Lie group, eg PGL, see [**?**].

Holomorphic bundles. Recall that there are two ways to define a holomorphic structure on a manifold: via holomorphic charts, or via linear algebra and the preliminary concept of an almost-complex structure: an endomorphism $J : TM \to TM$ such that $J^2 = -Id$. The Newlander-Nirenberg theorem provides the integrability condition relating these two points of view.

This data allows for a holomorphic function theory on M: we say that f is holomorphic if, in accordance with the Cauchy-Riemann equations, $df(JX) = i\,df(X)$, ie $df(X + iJX) = 0$ for all tangent vectors X. This condition can be reformulated in terms of the vanishing of a certain part of the operator d, as follows.

(i) We decompose the complexified vector space $T_xM \otimes \mathbb{C}$ by writing

$$X = \frac{1}{2}(X - iJX) + \frac{1}{2}(X + iJX).$$

The splitting $\partial x = \frac{1}{2}(\partial x - i\partial y) + \frac{1}{2}(\partial x + i\partial y)$ is an example of this in local coordinates.

(ii) Dually, we decompose the space $\Lambda^1_x(M) \otimes \mathbb{C}$ into the sum of forms which annihilate these two types of vectors. This is written

$$\Lambda^1_x(M) \otimes \mathbb{C} = \Lambda^{1,0}_x(M) \oplus \Lambda^{0,1}_x(M).$$

(iii) We thus obtain, by projection, a decomposition $d = \partial + \bar{\partial}$. Specifically, $\bar{\partial} f(X) := \frac{1}{2} df(X + iJX)$. A function is holomorphic iff $\bar{\partial} f = 0$.

A last observation: the fact that partial derivatives (wrt coordinate vector fields) commute is expressed by the fact $d^2 = 0$. This implies that all complexified partial derivatives commute. In particular, $\bar{\partial}^2 = 0$.

Now choose a complex vector bundle $E \to M$. In order to generalize the operator $\bar{\partial}$ to sections, ie to define holomorphic sections, extra data is required: a holomorphic structure on E. Analogously to the case of holomorphic structures on M, this can be defined in two ways. The first is elementary: a complex vector bundle $E \to M$ is *holomorphic* if it admits a trivializing atlas whose transition maps g_{ij} are holomorphic. In this case, given any smooth section σ, we define $\bar{\partial}_E \sigma$ so that, locally in any trivializing chart U_i, it corresponds to $\bar{\partial} s^i$. This definition is well-posed because, if we change coordinates, $\bar{\partial}(g_{ij} s^j) = g_{ij} \bar{\partial} s^j$. In this sense, holomorphic bundles are the exact analogue of bundles with constant transition maps, ie local systems.

Remark. A differentiable manifold may support many different holomorphic structures: such moduli spaces appear already in Riemann surface theory, eg in the case of complex tori. Analogously, any complex vector bundle on a holomorphic manifold may support many different holomorphic structures: again, moduli spaces. In both cases these can be seen as distinct subatlases of an initial maximal atlas.

The second way is in terms of (an alternative notion of) connections and an integrability condition similar to flatness, as follows. We define $\Lambda^{0,1}(E)$ to be the space of E-valued 1-forms which annihilate vectors of the form $X - iJX$. We then define a *partial connection* on E to be an operator $\bar{\partial}_E : \Lambda^0(E) \to \Lambda^{0,1}(E)$ satisfying the Leibniz rule $\bar{\partial}_E(f\sigma) = \bar{\partial} f \otimes \sigma + f\bar{\partial}_E\sigma$. The space of partial connections is parametrized by $\Lambda^{0,1}(\mathrm{End}(E))$. The gauge group is given by sections of the adjoint bundle $Ad(P) := P\times_{Ad}\mathrm{GL}(r,\mathbb{C})$, where P is the $\mathrm{GL}(r,\mathbb{C})$-principal fibre bundle of linear frames on E.

Exactly as before, a partial connection induces operators $\bar{\partial}_E : \Lambda^{0,i}(E) \to \Lambda^{0,i+1}(E)$ and a curvature operator $\bar{\partial}_E^2$. If $\bar{\partial}_E$ is obtained from holomorphic transition maps, then $\bar{\partial}_E^2 = 0$: we think of this as a flatness condition for the connection. In this context, parallel sections are those that are holomorphic. In the holomorphic case, local examples can be found via any holomorphic trivializing chart. In general, however, finding parallel sections corresponds locally to solving equations of the form $\bar{\partial} s = -As$. In general there is no guarantee that this equation admits solutions. It turns out that flatness is the correct integrability condition which allows us to solve equations of this type. It actually allows us to build local holomorphic frames, thus holomorphic transition maps g_{ij}.

Summarizing: given a complex vector bundle E, there is a 1:1 correspondence between holomorphic structures on E (modulo isomorphisms) and partial connections on E such that $\bar{\partial}_E^2 = 0$ (modulo gauge transformations).

Example. When M is a Riemann surface any partial connection automatically satisfies the flatness condition and thus defines a holomorphic structure on the underlying bundle. Since partial connections always exist, this means that any complex vector bundle on M admits a holomorphic structure.

The analogous statement is open on higher-dimensional complex manifolds.[1]

Connecting connections. We now face the task of understanding the relationship between the two notions of connection on a same complex vector bundle E. One direction is simple: given a connection ∇ on E, its (0,1) component $\nabla^{0,1}$ defines a partial connection $\bar{\partial}_E$. Specifically, for any $X \in T_pM$,

$$(\bar{\partial}_E)_X = \frac{1}{2}(\nabla_X + i\nabla_{JX}).$$

This map $\nabla \mapsto \bar{\partial}_E := \nabla^{0,1}$ has good properties. First, it is equivariant with respect to the gauge group actions. Second, recall that the splitting of $\Lambda^1(M)\otimes\mathbb{C}$ induces a splitting of higher-degree forms. In particular,

$$\Lambda^2(M) \otimes \mathbb{C} = \Lambda^{2,0}(M) \oplus \Lambda^{1,1}(M) \oplus \Lambda^{0,2}(M).$$

In turn, this induces a splitting of the bundle $\Lambda^2(M)\otimes E$. Applying this splitting to the curvature tensor we find $\bar{\partial}_E^2 = F_\nabla^{0,2}$. In particular, this shows that if ∇ is flat then E is holomorphic (alternatively, this follows from the fact that ∇ defines a local system, thus holomorphic transition maps).

[1] See post Complex vector bundles that are not holomorphic, mathoverflow, by D. Panov

In order to further understand this map, it is useful to introduce the following notion. Choose a Hermitian metric on E. It then turns out that, given any partial connection $\bar{\partial}_E$, there exists a unique $\mathrm{U}(r)$-connection ∇ on E whose (0,1) component coincides with $\bar{\partial}_E$. This ∇ is known as the *Chern connection* associated to $\bar{\partial}_E$. We can identify it as follows.

Choose a trivializing atlas wih $g_{ij} \in \mathrm{U}(r)$. Locally, $\bar{\partial}_E = \bar{\partial} + A$, with $A \in \Lambda^{0,1}(\mathrm{gl}(r,\mathbb{C}))$. Then $-\bar{A}^t \in \Lambda^{1,0}(\mathrm{gl}(r,\mathbb{C}))$ and, for each X, $A_X - \bar{A}^t_X \in \mathfrak{u}(r)$. The desired connection has the form $d + (A - \bar{A}^t) = (\partial - \bar{A}^t) + (\bar{\partial} + A)$. It has the property that $F^{2,0}_\nabla = \overline{F^{0,2}_\nabla}$, so $\bar{\partial}_E$ is integrable iff F_∇ is of type (1,1).

Remark. An alternative construction is possible in the special case where $\bar{\partial}^2_E = 0$, ie E is holomorphic. For simplicity, let us view it in the case where E is a line bundle. In this case, in each trivializing chart U_i we choose a holomorphic basis, ie a non-vanishing section σ_i. This provides an identification $E_{|U_i} \simeq U_i \times \mathbb{C}$ such that $\sigma_i \simeq 1$. Notice that $\nabla\sigma_i = \nabla^{1,0}\sigma_i + \bar{\partial}_E\sigma_i = \nabla^{1,0}\sigma_i$. In local coordinates this means $\nabla\sigma_i \simeq (\partial + A)1 = A$, where A is a 1-form of type (1,0). Our goal is to find A, assuming $\nabla h = 0$.

Set $H := h(\sigma_i, \sigma_i)$. The assumption $\nabla h = 0$ implies

$$dH = h(\nabla\sigma_i, \sigma_i) + h(\sigma_i, \nabla\sigma_i) \simeq h(A,1) + h(1,A).$$

The Hermitian condition means that, on the RHS, the first term is of type (1,0), the second of type (0,1). Thus, for example, $\partial H = h(A,1) = Ah(1,1) = AH$. It follows that $A = \partial \log H$. Its curvature is $F_\nabla = dA = \bar{\partial}\partial \log H \in \Lambda^{1,1}$, as expected.

To summarize: given a Hermitian metric on E, the construction of Chern connections produces a 1:1 correspondence between the space $\mathcal{A}$ of all $\mathrm{U}(r)$-connections on E and the space of all partial connections on E.

Under this correspondence, the curvature of the $\mathrm{U}(r)$-connection is of type (1,1) iff the partial connection satisfies $\bar{\partial}^2_E = 0$, ie it corresponds to a holomorphic structure on E. Gauge theory thus provides an interesting way to study holomorphic structures on a Hermitian vector bundle, by using the space $\mathcal{A}^{1,1}$ of unitary connections with (1,1) curvature. A key issue, however, is that one often wants to keep track of which holomorphic structures are equivalent. Here, equivalence refers to the action of the complex gauge group. On $\mathcal{A}$ or $\mathcal{A}^{1,1}$, however, the natural action is that of the unitary gauge group: the Chern construction produces a mismatch between the two natural gauge groups.

This observation leads to two different set-ups.

1. Recall that each Hermitian metric on E defines a $\mathrm{U}(r)$-subbundle of the $\mathrm{GL}(r,\mathbb{C})$-frame bundle, and that the action of the complex gauge group on the $\mathrm{GL}(r,\mathbb{C})$-frame bundle allows us to move from one subbundle to another; equivalently, from one Hermitian structure to another.

Now let us fix a Hermitian metric h on E. Choose a partial connection $\bar{\partial}_E$. Let ∇ denote the Chern connection associated to $\bar{\partial}_E$ wrt h. Applying to ∇ an element g in the $\mathrm{U}(r)$-gauge group corresponds to finding the Chern connection $g\cdot\nabla$ associated to the partial connection $g\cdot\bar{\partial}_E$, wrt the same metric h. If instead we apply to ∇ an element g in the $\mathrm{GL}(r,\mathbb{C})$-gauge group, we obtain the Chern connection $g\cdot\nabla$ associated to the partial connection $g\cdot\bar{\partial}_E$, wrt the metric $g\cdot h$.

To summarize: given a partial connection, this construction produces a complex orbit $\mathcal{C}_h$ of unitary connections, wrt varying metrics.

This construction depends strongly on the initial choice of h. If we change the initial Hermitian metric, the same initial partial connection $\bar{\partial}_E$ will determine a different initial Chern connection ∇, thus a different complex orbit. We thus obtain a family $\mathcal{C}=\{\mathcal{C}_h\}$ of complex orbits of connections on E, parametrized by the Hermitian metrics on E.

A natural problem in this setting is to look for a metric h such that $\mathcal{C}_h$ has optimal properties. In order to make sense, the properties must be invariant under the complex gauge group action. For example, the Narasimhan-Seshadri theorem (discussed below) looks for h such that connections in $\mathcal{C}_h$ are flat (more generally, Hermitian-Yang-Mills).

Remark. Even though we have fixed an initial partial connection, the equivariance underlying this construction implies that the orbit $\mathcal{C}_h$ built using $\bar{\partial}_E$ coincides with the orbit $\mathcal{C}_h$ built using $g\cdot\bar{\partial}_E$. In this sense the Narasimhan-Seshadri theorem concerns an isomorphism class of holomorphic structures on E, rather than a specific holomorphic structure/partial connection.

2. Fix a metric on E. We can define a new action of the complex gauge group on $\mathcal{A}$ so that (i) it preserves this space, (ii) under the correspondence defined by the Chern construction it coincides with the natural action of the complex gauge group on partial connections. By definition this action also preserves the subspace $\mathcal{A}^{1,1}$, it extends the natural action of the unitary gauge group and it makes the above correspondence become equivariant.

We emphasize that this construction, detailed for example in [**?**] Chapter 6, keeps the Hermitian metric on E fixed.

A natural problem in this setting is to look for (unitary or partial) connections in the given complex gauge orbit which have optimal properties wrt the given metric. This leads to an alternative formulation of the Narasimhan-Seshadri theorem.

In summary: the first set-up fixes a connection and looks for a preferred metric. The second set-up fixes the metric and looks for a preferred connection. Both set-ups allow us to work within a given complex orbit of partial connections, ie isomorphism class of holomorphic structures on E. They lead to formally different, but equivalent, formulations of the Narasimhan-Seshadri theorem.

Remark. We remark that the metric used for the Yang-Mills functional, based on the Killing form, is not only $\mathrm{U}(r)$-invariant; it is also $\mathrm{GL}(r,\mathbb{C})$-invariant.

The functional is thus constant on each $\mathcal{C}_h$, so studying it on $\mathcal{C}$ is equivalent to thinking of it as a functional on the space of Hermitian metrics on E.

The functional is instead not invariant with respect to the other complex gauge group action defined in point 2, above. Indeed, the theorems described in the next section can be proved by studying its gradient flow within the complex orbits.

Example. Let M be a compact Riemann surface of genus g. Let $E \to M$ be the trivial line bundle. We have seen that unitary flat connections are classified by $\mathrm{U}(1)^{2g}$. We have noted that this classification is independent of the specific $\mathrm{U}(1)$-structure on M. Now recall that the holomorphic structures on E form a group, denoted Pic^0; a long exact sequence argument shows that $Pic^0 \simeq H^1(\mathcal{O})/H^1(M;\mathbb{Z}) \simeq \mathbb{C}^g/\mathbb{Z}^{2g} \simeq \mathrm{U}(1)^{2g}$. We can explain this as follows.

Every flat unitary connection induces locally constant, thus holomorphic, transition maps g_{ij}, thus a holomorphic structure on E. Conversely, any such structure is generated by a flat unitary connection, wrt an appropriate metric. Indeed, choose a partial connection compatible with the holomorphic structure and an initial Hermitian metric h. As above, we then obtain a Chern connection ∇ which, in appropriate coordinates, is of the form $\partial + \partial \log H$. Its curvature F_∇ is then locally expressed by $\bar{\partial}\partial \log H$. According to Chern-Weil theory, the fact $c_1(E) = 0$ implies that $\int_M F_\nabla = 0$. Any other metric is of the form $e^f h$. The curvature of its Chern connection is $\bar{\partial}\partial H + \bar{\partial}\partial f$. One can show that, in this situation, it is possible to solve the global equation $\bar{\partial}\partial f = -F_\nabla$. The resulting f is such that the Chern connection of the initial holomorphic structure, wrt $e^f h$, is flat.

We will see below that this is the simplest manifestation of the Narasimhan-Seshadri theorem.

Remark. In the above example we have followed the first set-up: fixed partial connection, variable metric.

6 Curvature and stability

In the previous section we discussed a very general relationship between holomorphic structures and unitary connections. The last example showed that, in the special case of trivial bundles over Riemann surfaces, we can refine this relationship obtaining flat unitary connections. We now want to extend this discussion to vector bundles of higher rank. This will require the concept of stability.

Notice that holomorphic vector bundles are more complicated than line bundles in many ways. Some issues are simple: tensor products do not preserve rank and duality does not define inverses, so they do not form a group. Other issues are more subtle. In higher dimensions there is no analogue of the relationship between divisors and holomorphic line bundles. Also, holomorphic subbundles do not necessarily have complements, leading to the issue of which bundles are decomposable, ie can be written as direct sums of holomorphic subbundles. Furthermore, given a holomorphic homomorphism $\phi : E_1 \to E_2$ between holomorphic vector bundles, very roughly speaking (ie, up to rephrasing everything in terms of coherent sheaves so as to define images and manage dimension jumps) it is generally not true that $Im(\phi)$ is isomorphic to $E_1/Ker(\phi)$.

These facts lead to a sophisticated theory of holomorphic vector bundles. Large parts of this theory are independent from gauge theory, which on the surface simply offers an alternative formalism for discussing holomorphic structures and the concept of equivalence, via the action of the complex gauge group. We shall review some aspects of holomorphic vector bundle theory below, in particular pertaining to the concept of stability. It turns out however that stability is also the key to a much deeper link between the vector bundle theory and gauge theory (with respect to the group $G = \mathrm{U}(r)$), as described by a body of results originating with the Narasimhan-Seshadri theorem. Our starting point for presenting this result is the following question.

We have already understood the flatness condition for a partial connection on E. Given any Hermitian metric on E, the construction of the Chern connection leads to the question of characterizing the flatness of this connection.

Of course, we should expect that such flatness depends on the choice of Hermitian structure. We will see that it also corresponds to a special category of holomorphic bundles.

Let us restrict our attention to the simplest category: we assume M is a Riemann surface. A first simplification in this context is the following: the fact that M has dimension 2 implies the existence of non-zero smooth sections, so that $E \simeq L \oplus \mathbb{C}^{r-1}$ where L is a complex line bundle. This shows that $\det(E) \simeq \det(L) = L$. It follows that, just like complex line bundles, complex vector bundles are classified by their rank and by $c_1(E)$, ie by their degree $deg(E) := c_1(E) \cdot M$.

Stability. As already mentioned, the theory of holomorphic vector bundles rests upon the foundational concept of stability. In turn, this requires the notion of slope.

Let $E \to M$ be a complex vector bundle over a compact Riemann surface. The *slope* of E is $\mu(E) := \frac{deg(E)}{rk(E)} = \frac{\int_M c_1(E)}{r}$.

Slope is a numerical invariant which depends on the topology of E, not on any additional holomorphic structure. However, the following example shows that it does control the holomorphic geometry.

Example. Let E be a holomorphic line bundle over a compact Riemann surface M. The number $\mu(E) = deg(E)$ controls the holomorphic geometry of E in several ways.

First, it coincides with the number of zeroes minus the number of poles of any non-zero meromorphic section of E. In particular, if $deg(E) < 0$ then any non-zero meromorphic section certainly admits poles, ie there exist no non-zero holomorphic sections.

Second, it governs the existence of non-zero holomorphic homomorphisms $E_1 \to E_2$ between holomorphic line bundles. Indeed, such maps are equivalent to non-zero holomorphic sections of the line bundle $E_1^* \otimes E_2$, thus they can exist only if $deg(E_1) \leq deg(E_2)$: slope is monotone under non-zero holomorphic homomorphisms.

The role of slope for line bundles is thus analogous to the role of genus in governing the existence of non-constant holomorphic maps between compact Riemann surfaces, via the Hurwitz formula.

Example. Any holomorphic homomorphism between holomorphic vector bundles E_1, E_2 defines a holomorphic section of $E_1^* \otimes E_2$, thus a holomorphic section of the corresponding determinant line bundle.

Let us assume this map is non-zero. It follows that $deg(\det(E_1^* \otimes E_2)) \geq 0$. Now recall that, given two square matrices A, B of order a, b, $\det(A \otimes B) = \det(A)^b \det(B)^a$. This implies that $deg(\det(E_1^* \otimes E_2)) = -deg(E_1)rk(E_2) + deg(E_2)rk(E_1) \geq 0$. It follows that $deg(E_1)/rk(E_1) \leq deg(E_2)/rk(E_2)$: if we want to mimic the situation for line bundles, the definition of slope is basically forced upon us.

In particular, the notion of slope provides the formulation of a special constraint on holomorphic vector bundles: we say that a holomorphic vector bundle E over a Riemann surface is *stable* (respectively *semi-stable*) if, for any non-trivial holomorphic subbundle $F < E$, $\mu(F) < \mu(E)$ (respectively $\mu(F) \leq \mu(E)$).

Stability thus limits the topological type of holomorphic subbundles. Notice: although slope itself does not detect the holomorphic structure, the holomorphic structure influences stability by controlling which are the holomorphic subbundles to which the condition applies.

The notion of stability was introduced by Mumford in the '60s in terms of Geometric Invariant Theory (GIT). In this context, the main result is that while moduli spaces of holomorphic bundles are in general not smooth, GIT shows that stable holomorphic bundles (with fixed rank and degree) do have smooth moduli spaces. These moduli spaces have a natural complex structure. They can be compactified to (non-smooth) projective varieties by adding the semi-stable bundles, modulo a certain notion of "S-equivalence". The notion of S-equivalence can be by-passed via the notion of *poly-stable* bundles: bundles which can be written as a direct sum of stable bundles, all with the same slope. Such bundles are automatically semi-stable, and each S-equivalence class of a semi-stable bundle contains a unique poly-stable representative, so Mumford's compactified moduli space can be identified with the space of poly-stable bundles.

Example. Any holomorphic line bundle is trivially stable because it admits no non-trivial subbundles. This corresponds to the fact that the moduli spaces are automatically smooth, isomorphic to g-dimensional complex tori.

Any automorphism of a holomorphic line bundle defines a holomorphic section of the trivial line bundle $E^* \otimes E$, and is thus constant.

Example. Assume E is a semi-stable holomorphic bundle such that r and $deg(E)$ are co-prime. It is then automatically stable, for obvious numerical reasons. In this case, Mumford's theorem shows that the moduli space of such bundles is a smooth (compact) projective manifold.

Remark. Semi-stable bundles generally produce singular points in the moduli space. Unstable bundles would cause the moduli space to be non-Hausdorff. The simplest manifestation of this is the "jumping phenomenon", [**?**]: one can find holomorphic families of holomorphic vector bundles parametrized by $t \in \Delta$, which are isomorphic for $t \neq 0$ but not for $t = 0$. These two holomorphic bundles cannot be separated from each other in the moduli space. However, the bundle corresponding to $t = 0$ is unstable so it does not belong to Mumford's moduli space.

Stable bundles are very rigid: they have only constant automorphisms. This implies that they are indecomposable, ie cannot be written as direct sum of non-trivial subbundles: otherwise they would at least admit automorphisms of the form $\mathbb{C}^* \times \cdots \times \mathbb{C}^*$.

In general, the properties of semi-stable holomorphic vector bundles are analogous to those of holomorphic line bundles. In particular, one can show that the existence of a non-zero holomorphic homomorphism between semi-stable holomorphic bundles E_1, E_2 implies $\mu(E_1) \leq \mu(E_2)$ (even without the additional assumption that the induced determinant map is non-zero). Semi-stable bundles form a category with good properties: for example, if $\phi : E_1 \to E_2$ is a holomorphic homomorphism, then (roughly speaking, as above) $Ker(\phi)$ and $Im(\phi)$ are again semi-stable and $E_1/Ker(\phi)$ is isomorphic to $Im(\phi)$.

Semi-stable bundles admit filtrations via subbundles whose successive quotients are stable. Holomorphic bundles admit filtrations via subbundles whose successive quotients are semi-stable. In this sense, stable bundles are the building blocks of holomorphic bundles.

Example. Grothendieck showed that any holomorphic vector bundle on $\mathbb{CP}^1$ can be decomposed into the sum of holomorphic line bundles. When $r > 1$, it is thus not stable: Mumford's moduli space is empty.

Remark. Much work has been done studying the topology of Mumford's moduli spaces (in particular, the Betti numbers and the ring structure on cohomology). Another much-studied problem concerns the existence of canonical geometric structures on these spaces. For example, when $G = \mathrm{U}(r)$ or $G = \mathrm{SU}(r)$, the moduli space has a natural Kähler structure. The symplectic structure is actually independent of the complex structure on M. We refer to [**?**] for details.

Flatness and stability. Coming back to our original question, the answer is provided by the Narasimhan-Seshadri theorem which, in its simplest version, concerns a degree zero (ie, smoothly trivial) indecomposable holomorphic vector bundle E over a compact Riemann surface M. As mentioned, indecomposability is a necessary condition for stability.

Theorem 6.1 (Narasimhan-Seshadri, degree zero) *Let E be a degree zero (ie, smoothly trivial) complex vector bundle over a compact Riemann surface. Assume E is endowed with an indecomposable holomorphic structure.*

Then E is stable iff it admits a Hermitian metric whose Chern connection is flat. This metric is unique.

Remark. We have phrased this statement in terms of Hermitian metrics, ie in terms of the first set-up, discussed above. The statement fixes the holomorphic structure, thus an equivalence class of partial connections. Stability and indecomposability are invariant under the complex gauge group action. As seen, each initial Hermitian metric h defines a complex orbit $\mathcal{C}_h$ of $U(r)$-connections on E (wrt varying Hermitian metrics). Flatness of the connections in any such orbit is also gauge-invariant. The theorem says that the holomorphic bundle E is stable iff there exists a metric h whose complex orbit $\mathcal{C}_h$ consists of flat connections.

We can alternatively formulate the statement according to the second set-up. This requires choosing a Hermitian metric h. The statement is then that the holomorphic bundle E is stable iff the corresponding complex orbit in $\mathcal{A}^{1,1}$ contains a flat $\mathrm{U}(r)$-connection. This connection is unique up to the $\mathrm{U}(r)$-gauge group action.

The first set-up leads to a "best" metric, the second to a "best" connection.

Example. We have shown above that any holomorphic line bundle is stable. When it is smoothly trivial over a compact Riemann surface, we have also shown how to find a Hermitian metric inducing a flat Chern connection. In this sense, the Narasimhan-Seshadri theorem is the higher-dimensional analogue of facts already discussed for line bundles.

As seen above, flat Hermitian connections can be classified via their monodromy representations $\pi_1(M) \to \mathrm{U}(r)$. Not all such representations arise from stable holomorphic bundles: stability (thus indecomposability) forces the extra condition that the representation be irreducible, ie does not admit non-trivial invariant subspaces. We obtain all representations by working in the larger category of poly-stable bundles.

Corollary 6.2 *Let M be a compact Riemann surface. Consider the trivial vector bundle $E := M \times \mathbb{C}^r$. There are 1:1 correspondences between the moduli space of poly-stable holomorphic structures, flat Chern connections with curvature of type (1,1), and homomorphisms $\pi_1(M) \to \mathrm{U}(r)$ (up to equivalence).*

This correspondence restricts to a correspondence between stable holomorphic structures, irreducible flat connections and irreducible homomorphisms.

We can think of the reducible connections, ie connections obtained as direct sums, as a way of compactifing the moduli space of flat Chern connections in the same way that poly-stable structures compactify the moduli space of stable structures.

Curvature and stability, part 1. We now wish to know what is the analogous statement for non-trivial holomorphic vector bundles. Once again, our guideline is provided by the case of holomorphic line bundles.

Example. Let $E \to M$ be a holomorphic line bundle over a compact Riemann surface. According to Chern-Weil theory, $c_1(E)$ is represented by $\frac{i}{2\pi}F_\nabla$, for any Chern connection ∇ defined by a Hermitian metric on E. Conversely, any real (1,1) form α in the class $c_1(E)$ is an $\frac{i}{2\pi}$ multiple of the curvature of some such connection. Indeed, choose one such metric h. Any other metric is of the form $e^f h$, for some $f : M \to \mathbb{R}$. Our problem corresponds to solving the equation $\frac{i}{2\pi}(\bar{\partial}\partial f + F_\nabla) = \alpha$, ie $\bar{\partial}\partial f = -2\pi i\,\alpha - F_\nabla$, which indeed has solution because the RHS has integral 0.

The next question is how to choose a canonical $\alpha \in c_1(E)$. We can do this via a Kähler form ω on M. Indeed, if we normalize it so that $\int_M \omega = 1$, it provides a canonical representative for the generator of $H^2(M;\mathbb{Z})$, thus canonical representatives $\lambda\omega$ for all other classes in $H^2(M;\mathbb{R})$. Given any holomorphic line bundle E over M and normalized ω, we obtain a canonical Hermitian metric h on E by choosing $\lambda \in \mathbb{R}$ such that $c_1(E) = \lambda[\omega]$, then imposing the condition that the Chern connection have curvature $F_\nabla = -2\pi i\,\lambda\omega$. Notice that (i) all such ω are conformally equivalent, (ii) the solution h depends on the specific ω, indeed the condition expresses a certain compatibility between h and ω, (iii) the specific λ is independent of ω, ie it is determined only by the topology of E: $\lambda = \int_M c_1(E) = \mu(E)$.

More generally, given a holomorphic vector bundle E over a Riemann surface (M, J, ω), a *Hermitian-Yang-Mills connection* (HYM), or Hermitian-Einstein connection, is the Chern connnection ∇ on E defined by some metric h such that $F_\nabla = -2\pi i\,\lambda\, Id_E\omega$, for some $\lambda \in \mathbb{R}$. Taking the trace of the endomorphisms and integrating one finds that, necessarily, $\lambda = \mu(E)$. The corresponding metric is a *Hermitian-Einstein metric* (HE) on E. One can show that if h is weakly Hermitian-Einstein, in the sense that the equation holds with respect to a real function λ, then there exists a conformally equivalent $h' = e^f h$ which is Hermitian-Einstein in the usual sense. In particular, this shows that if a HYM connection exists with respect to a given ω, it exists wrt any other conformally equivalent ω'. It follows that, in the setting of Riemann surfaces, the existence depends only on J.

Theorem 6.3 (Narasimhan-Seshadri, general) *Let E be an indecomposable holomorphic vector bundle over a compact Riemann surface.*

Then E is stable iff it admits a metric whose Chern connection is Hermitian-Yang-Mills (for some, thus any, choice of ω). This metric is unique. The *corresponding constant coincides with the slope of E.*

More generally, given E, there is a 1:1 correspondence between poly-stable holomorphic structures, HYM connections with curvature of type (1,1) and homomorphisms $\pi_1(M) \to P\,\mathrm{U}(r)$.

As for stability, the HYM condition is complex gauge-invariant. We can thus incorporate the gauge group into the Narasimhan-Seshadri theorem as follows: an equivalence class of holomorphic structures on E is stable iff there exists a Hermitian metric h such that all connections in the corresponding class $\mathcal{C}_h$ are HYM.

Remark. The original Narasimhan-Seshadri theorem gave only the correspondence between stability and representations of π_1. The gauge-theoretic viewpoint was introduced by Atiyah-Bott.

HYM connections are critical points of the YM functional. More specifically, they are absolute minima. Indeed, given a normalized ω, we can write $F_\nabla = \omega \otimes A$, for some $A \in \Lambda^0(End(E))$. It follows that the YM integrand coincides with $\|A\|^2$. Recall the orthogonal decomposition $A = A_0 + \frac{\mathrm{tr}(A)}{r} Id_E$, where A_0 is trace-free and $\mathrm{tr}(A)$ is an imaginary-valued function on M. Notice that $\frac{i}{2\pi}\int_M \mathrm{tr}(A)\omega = \int_M c_1(E) = r\mu(E)$. It follows that

$$\begin{aligned} YM(\nabla) &= \int_M \|A\|^2\omega = \int_M \|A_0\|^2\omega + \int_M \|\frac{\mathrm{tr}(A)}{r} Id_E\|^2\omega \\ &= \int_M \|A_0\|^2\omega + r\int_M |\frac{i\ \mathrm{tr}(A)}{r} \pm 2\pi\mu(E)|^2\omega \\ &= \int_M \|A_0\|^2\omega + r\int_M (\frac{i\ \mathrm{tr}(A)}{r} - 2\pi\mu(E))^2\omega + 4\pi^2 r\mu(E)^2 \\ &\geq 4\pi^2 r\mu(E)^2. \end{aligned}$$

This absolute minimum is a topological constant. It is attained precisely when A is trace-free and $\mathrm{tr}(A)$ is constant, which implies that ∇ is HYM.

Curvature and stability, part 2. The analogue of the Narasimhan-Seshadri theorem over higher-dimensional manifolds is known as the Kobayashi-Hitchin (or Donaldson-Uhlenbeck-Yau) correspondence. It was proven for algebraic manifolds by Donaldson, for Kähler manifolds by Uhlenbeck-Yau and for Hermitian manifolds by Buchdahl and Li-Yau.

The statement for Kähler manifolds requires a generalization of the notion of slope. Let (M, J, ω) be a compact n-dimensional Kähler manifold (normalized so that $\int_M \omega^n = 1$) and E be a rank r complex vector bundle over M.

The *slope* of E is $\mu(E) := \frac{\int_M c_1(E)\wedge\omega^{n-1}}{r}$. It provides a notion of stability similar to that seen above. Since slope depends only on the class $[\omega]$, this holds also for stability.

When E is holomorphic, a *Hermitain-Yang-Mills connection* on E is the Chern connection ∇ on E defined by some metric h such that $F_\nabla \wedge \omega^{n-1} = -2\pi i\, \lambda\, Id_E \omega^n$, for some $\lambda \in \mathbb{R}$. Taking the trace of the endomorphisms on both sides and integrating, we find that necessarily $\lambda = \mu(E)$. The corresponding metric is a *Hermitian-Einstein metric* on E.

The Donaldson-Uhlenbeck-Yau theorem states that E is poly-stable iff it admits a Hermitian-Einstein metric. This metric is unique.

Given the relationship between stability and the algebraic GIT theory, the theorem shows that the algebraic problem of stability is related to the analytic problem of the existence of solutions to the Hermitian-Einstein PDE.

The statement of the theorem is obviously analogous to that seen for Riemann surfaces. An additional analogy is the fact that HYM connections are again absolute minima of the YM functional. We remark that, although the YM equation is a second order PDE on the connection, the HYM condition is of first order. As before, the minimum value is determined by the topology of E (its characteristic classes) and by its slope (and thus depends on $[\omega]$), [**?**] equations 4.3.8 and 4.3.29. Finally, as before, while the specific Hermitian-Einstein metric on E depends on the specific choice of ω, its existence depends only on the class $[\omega]$.

This leads however to a first important difference. The Kähler cone of M is typically not just a half-line, as for Riemann surfaces. It thus plays a more important role: changing $[\omega]$ might destabilize E, [**?**]. A second difference concerns the fact that, in higher dimensions, it is typically harder to compactify both the moduli space of stable structures and the moduli space of Hermitian-Einstein connections [**?**]. We will discuss below the partial role of "bubbling" in this process.

Remark. Kobayashi [**?**] develops a general theory of Hermitian-Einstein connections which is parallel to that of stable bundles. It applies to complex manifolds in all dimensions using a J-compatible metric g on M, ie the 2-form ω. However, the fact that a weak Hermitian-Einstein metric can be conformally rescaled to become Hermitian-Einstein works only for Kähler manifolds.

The notion of weak HYM connections is a metric-dependent variation of the notion of projectively flat Hermitian connections. If E admits a projectively flat Hermitian connection then it clearly admits a weak HYM connection for any g. The converse is generally not true. The problem is that a tensor in $\Lambda^{1,1} \otimes E_x$ is generally not decomposable (unless M is a Riemann surface). The weak HYM condition expresses a balance between the single terms in the decomposition which depends on the specific metric g. Changing the metric may affect this balance. It follows that although projectively flat Hermitian connections can be classified via homomorphisms $\pi_1(M) \to P\,\mathrm{U}$, this does not apply to HYM connections.

7 Digression: topology vs. geometry

Differential geometers and Analysts tend to work at a rather sophisticated level, potentially forgetting (or ignoring) the topological roots of the concepts they deal with. The goal of this section is to review the different levels at which one can work, emphasizing both the new tools that differential geometers use and the issues involved. Roughly speaking, we shall try to underline the distinction and relationships between (i) algebraic topology, applied to topological manifolds, (ii) differential topology, which refines this study to smooth manifolds, and (iii) differential geometry, which represents a further step in the smooth category by introducing very different tools and methods, eg connections and analysis.

In passing from the topological to the geometric viewpoint one generally encounters two issues.

1. It is often natural to trade integer for real coefficients. This has several consequences.

The most obvious is that, as already mentioned, a $\mathbb{Z}$-module L need not be free. Torsion elements in integral homology/cohomology groups contain extra information. The universal coefficient theorem shows that real/complex (co)homology is equivalent to tensoring the integral (co)homology with $\mathbb{K}$. This operation kills the torsion elements. There is no solution to this problem. Information is lost forever.

Another consequence is that it hides important aspects of the theory. Indeed, even if L is free it is a much more rigid object than a vector space. For example, matrices in $\mathrm{GL}(n,\mathbb{R})$, used to classify bases of vector spaces, have non-zero determinant. Matrices in $\mathrm{GL}(n,\mathbb{Z})$, used to classify bases of free $\mathbb{Z}$-modules, have determinant ± 1. Likewise, a real bilinear form is completely determined by its rank and signature. The theory of integral bilinear forms is more elaborate. Here, rank, signature and non-degeneracy are defined to be those of the corresponding real form but this suffices, for example, only to control the injectivity of the corresponding map $L \to \mathrm{Hom}(L,\mathbb{Z})$. Surjectivity of this map requires the stronger *unimodularity* condition, ie the matrix representing the integral form must have determinant ± 1, and there exist further invariants such as type (even/odd). This leads to a more complicated classification.

2. Extra structure such as a differentiable and/or Riemannian structure can simplify certain definitions and proofs, eg by substituting cohomology classes with de Rham classes, or de Rham classes with harmonic forms. However, it shadows the fundamentally topological nature of the object being studied. Different ("exotic") differentiable structures define different de Rham cohomology spaces, but they are isomorphic because they correpond to the same cohomology defined topologically. Likewise, different Riemannian metrics define different classes of harmonic forms, but they are isomorphic because they correspond to the same de Rham cohomology classes.

We shall review some of this in greater detail below. In this context, however, perhaps the most important example to keep in mind is the following.

Example. Integer coefficients play a vital role concerning the question of projectivity of a compact Kähler manifold M.

Recall the starting point. Holomorphic maps into $\mathbb{CP}^N$ are built using holomorphic sections of line bundles. This process works if M admits an ample line bundle L: the embedding is then built using sections of some L^k. The main issue is thus to furnish a useful characterization of ampleness. This can be done either in terms of the "Nakai criterion", or as follows.

Set $H^k := H^k(M;\mathbb{C})$ and let $H^{p,q}$ denote Dolbeault cohomology. On a general compact complex manifold there is no relationship between these spaces. However, given a Kähler metric, Hodge theory implies that (i) H^k can be represented via harmonic k-forms, (ii) any harmonic k-form splits into harmonic (p,q)-forms, (iii) harmonic (p,q)-forms represent $H^{p,q}$. This leads to embeddings of $H^{p,q}$ into H^k. Changing the metric changes the harmonic representants only by exact forms, so the embeddings are metric-independent. The conclusion is a metric-independent decomposition $H^k = \oplus H^{p,q}$.

Now set $\Lambda := H^2(M;\mathbb{Z})$ so that $V := H^2(M;\mathbb{R}) = \Lambda \otimes \mathbb{R}$ and $H^2 = V \otimes \mathbb{C} = H^{2,0} \oplus H^{1,1} \oplus H^{0,2}$. The complexification process defines a natural conjugation on H^2. The construction above shows that $H^{1,1}$ and $H^{2,0} \oplus H^{0,2}$ are conjugation-invariant, so they are the complexification of subspaces in V. This leads to a splitting $V = V^{1,1}_{\mathbb{R}} \oplus (H^{2,0} \oplus H^{0,2})_{\mathbb{R}}$. Since M is Kähler, $V^{1,1}_{\mathbb{R}}$ contains an open cone of classes containing positive representants. The crux of projectivity is whether this cone intersects Λ (more precisely: the torsion-free part of Λ, viewed as a subset of V), ie whether M admits a positive closed (1,1) form whose cohomology class is integral.

Notice: the key point here is that a density argument shows that any cone which is open in V will intersect Λ, but here we are typically working with a cone which is open inside a closed subspace of V. Projectivity thus involves a delicate balance between the complex structure and integral topology.

The final argument is as follows.

1. If M is projective then the Fubini-Study form on $\mathbb{CP}^N$ pulls back to a positive closed (1,1) form whose class is integral.

2. Conversely, assume M admits a positive closed (1,1) form whose class is integral. The first issue is to show that it represents $c_1(L)$, for some L. This is the role of the Lefschetz (1,1) theorem, which characterizes the image of the map c_1 in the exact sequence

$$\cdots \to Pic(M) \simeq H^1(M;\mathcal{O}^*) \xrightarrow{c_1} H^2(M;\mathbb{Z}) \xrightarrow{\alpha} H^2(M;\mathcal{O}) \simeq H^{0,2}(M) \to \dots$$

Exactness shows that $Im(c_1) = Ker(\alpha)$, while the decomposition of V shows that $Ker(\alpha) = \Lambda \cap H^{1,1}(M;\mathbb{R})$. Any such form thus represents $c_1(L)$, for some holomorphic line bundle L (and is the curvature of the Chern connection of some h).

The second issue is to show that L is ample. This is achieved via the Kodaira embedding theorem.

The conclusion is a characterization of projectivity in differential-geometric terms: M is projective iff it admits a form as above, ie iff it achieves the above balance between complex structure and integral topology.

Remark. The above justifies the following definition: L is *positive* iff $c_1(L)$ is represented by a positive closed (1,1) form (integrality is automatic). Kodaira's theorem shows that positivity implies ampleness. The converse is also true, again by studying the pull-back of the Fubini-Study metric. Positivity thus provides a differential-geometric characterization of ampleness.

Topological invariants. Algebraic topology provides a hierarchy of invariants for a compact manifold M.

The first level consists of homotopy and (co)homology groups. As mentioned, for the latter one must take into account the possible presence of torsion. We will thus write $H_k = (H_k)_f \oplus Tor$, $H^k = (H^k)_f \oplus Tor$. The ranks of the torsion-free parts determine the simplest invariants: numbers.

A second level consists of the algebraic structure on (co)homology. In particular, when M is oriented and compact we have: the intersection product $\cap : H_i \times H_{n-i} \to H_0 \simeq \mathbb{Z}$, the cup product $\cup : H^i \times H^{n-i} \to H^n \simeq \mathbb{Z}$ and Poincare' duality $PD : H_i \to H^{n-i}$, which provides an isomorphism which interweaves the two products: $PD([A] \cap [B]) = PD([A]) \cup PD([B])$.

The two products are integral bilinear forms. When M has dimension $n = 4m$ one obtains, in particular, a symmetric product $H^{2n} \times H^{2n} \to \mathbb{Z}$ (equivalently, $H_{2n} \times H_{2n} \to \mathbb{Z}$). The kernel of the induced map $H^{2n} \to \mathrm{Hom}(H^{2n}, \mathbb{Z}) = \mathrm{Hom}((H^{2n})_f, \mathbb{Z})$ is precisely the torsion part and its restriction to $(H^{2n})_f$ is unimodular. In particular, the corresponding real form is non-degenerate so its signature (b^+, b^-) has the property $b^{2n} = b^+ + b^-$.

A third level consists of characteristic classes. These however require a choice of vector bundle, so at this stage we need to introduce extra structure: a choice of smooth structure leads to the tangent bundle TM, thus Stiefel-Whitney (SW), Euler and Pontryagin classes of M. A further choice of (almost) complex structure on TM leads to Chern classes.

It turns out that there exists an intricate network of relationships between characteristic classes, and between these classes and the above bilinear form. For example, when $n = 4$, let $\tau := b^+ - b^-$ denote the *index* of M. The *index theorem* shows that $\tau = \frac{1}{3} \int_M p_1(TM)$. If M has an almost complex structure one can further write the RHS in terms of Chern numbers.

Example. Using a combination of characteristic classes and complex geometry, one can show that K3 surfaces have $b^2 = 22$ and signature (3,19).

Characteristic classes are also closely related to specific aspects of the geometry of M. For example, assume M is a compact oriented Riemannian manifold. Recall that $\mathrm{SO}(n)$ has two covers: the disconnected group $\mathrm{O}(n)$ and the connected, simply connected group $\mathrm{Spin}(n)$. Restricting a $\mathrm{O}(n)$-structure on M to $\mathrm{SO}(n)$ corresponds to choosing an orientation: this is possible iff the first SW class vanishes. Covering the $\mathrm{SO}(n)$-frame bundle with a $\mathrm{Spin}(n)$-bundle is possible iff the second Stiefel-Whitney class vanishes: this allows the construction of “spinor bundles” over M.

Invariants and geometry. Geometric structures and ideas provide new tools for studying the above objects.

One example is provided by the notion of differential forms, which give a new way to study cohomology and its algebraic structure via de Rham theory and the wedge product on forms. Specifically, using coefficients in $\mathbb{K}$, we find $H^* \simeq H^*_{dR}$ and $[\alpha]\cap[\beta] \simeq \alpha\wedge\beta$. Hodge theory provides further insight by using a Riemannian metric to find canonical (harmonic) representatives in each class. Notice however that it does not encode the algebraic structure: the wedge of harmonic forms is not necessarily harmonic.

Another example is provided by Chern-Weil theory, which provides a differential-geometric approach to (real) characteristic classes via connections and curvature.

We will now review yet another example, provided by the Hodge star operator. Since below we shall be mostly interested in the case of 4-manifolds, we shall restrict the discussion to this setting.

Digression. Let (V, g) be a 4-dimensional oriented Euclidean vector space. Consider the space $\Lambda^2(V)$. On this space we have two bilinear forms: the induced ($\mathbb{R}$-valued) metric $g(\alpha, \beta)$ and the (Λ^4-valued) symmetric form $\alpha\wedge\beta$. As usual there exists an endomorphism, typically denoted $\star$, such that $\alpha\wedge\beta = g(\star\alpha, \beta)\,\mathrm{vol}_g$.

The operator $\star$ coincides with the standard Hodge operator on $\Lambda^2(V)$, generally defined by $\alpha\wedge\star\beta = g(\alpha, \beta)\,\mathrm{vol}_g$. Indeed, in this dimension $dim(\Lambda^2(V)) = 6$ and $\star\star = Id$. Together with the fact that the Hodge operator is an isometry, this confirms the claim.

The above also implies that $\star$ satisfies $g(\alpha, \star\beta) = g(\star\alpha, \beta)$, ie it is g-symmetric. It is thus diagonalizable, with eigenvalues ± 1. We can also compute that each eigenspace has dimension 3. In other words, the bilinear form $\wedge$ on $\Lambda^2(V)$ has signature $(3, 3)$.

A second feature in this dimension is that if we conformally rescale the metric $g \mapsto \lambda^2 g$ on vectors, then the induced metric on 2-forms rescales by λ^{-4}: this factor is cancelled by the rescaled volume form so the Hodge star operator, and everything related to it, depends only on the conformal class of the metric.

Now let M be a compact oriented Riemannian 4-manifold. Applying the above to each cotangent space we obtain a decomposition $\Lambda^2(M) = \Lambda^+(M) \oplus \Lambda^-(M)$, thus the (infinite-dimensional) spaces of selfdual (SD) and anti-selfdual (ASD) 2-forms. Using the isomorphism $H^2 \simeq H^2_{dR}$ we obtain

$$[\alpha]\cap[\beta] \simeq \int_M \alpha\wedge\beta = \int_M g(\alpha, \star\beta)\,\mathrm{vol}_g\,.$$

Let $\mathcal{H}^2$ denote the space of harmonic 2-forms. Hodge theory shows that $H^2_{dR} \simeq \mathcal{H}^2$, so the latter has dimension b^2 and a non-degenerate symmetric bilinear form induced from the previous equation.

Recall that, in general, the form only provides a decomposition of the vector space into positive/negative regions, separated by the isotropy cone. In this case, however, the Hodge operator commutes with the Laplace operator Δ_g so we obtain a splitting $\mathcal{H}^2 = \mathcal{H}^+ \oplus \mathcal{H}^-$. If $\alpha \in \mathcal{H}^+$ then

$$[\alpha] \cap [\alpha] \simeq \int_M g(\alpha, \alpha)\,\mathrm{vol}_g \geq 0.$$

Analogously, $\alpha \in \mathcal{H}^-$ implies that $[\alpha] \cap [\alpha] \leq 0$. The metric thus provides canonical maximal positive/negative subspaces. It follows that $b^+ = \dim(\mathcal{H}^+)$, $b^- = \dim(\mathcal{H}^-)$.

Remark. Notice that any closed (A)SD 2-form is automatically harmonic and that any exact (A)SD 2-form vanishes. It follows that the abstract de Rham construction of restricting to closed forms and quotienting by the exact ones, applied to the spaces of (A)SD forms, would again lead to the spaces of harmonic (A)SD 2-forms.

In the Hermitian setting there exists yet another convenient handle on (A)SD forms. Let us assume our 4-dimensional vector space V is Hermitian, endowed with the orientation induced by the complex structure. Set $\omega := g(J\cdot, \cdot)$. The space $\Lambda^{2,0} \oplus \Lambda^{0,2}$ is conjugation-invariant, so it is the complexification of a space W of real 2-forms. It turns out that $W \oplus < \omega >$ coincides with the space of SD forms, so (the complexification of) an ASD form α must be of type (1,1). Furthermore, $\alpha \in \Lambda^{1,1}$ is ASD iff $g(\alpha, \omega) = 0$. Since $\star\omega = \omega$, this is equivalent to the condition $\alpha \wedge \omega = 0$. The analogue holds for (almost) Hermitian 4-manifolds.

8 ASD connections

Flat connections are defined by a first order condition on the connection, but trivially satisfy the second order Yang-Mills equations because they provide examples of absolute minimizers of the YM functional. However, they exist only when the YM functional can take value 0, which implies strong topological constraints on E.

Hermitian-Einstein connections follow a similar pattern. In both cases, the point is that appropriate structures on M (such as a Kähler metric) lead to algebraic conditions (such as the HE condition) which provide very strong control over the Yang-Mills equations, allowing us to characterize minimizers algebraically and to relate their existence to special properties of E (local systems or stability). In summary: in appropriate situations, algebra controls analysis.

ASD connections provide yet another instance of this situation. Let $E \to M$ be a vector bundle over a 4-dimensional oriented Riemannian manifold M. Choose a connection ∇ on E. We can use the splitting of 2-forms on M to write its curvature as $F_\nabla = F_\nabla^+ + F_\nabla^-$. The connection on E is *anti-selfdual* if

$F^+_\nabla = 0$, ie if F_∇ is an endomorphism-valued ASD 2-form on M. This condition depends only on the conformal class of the metric on M.

ASD connections are also known as "instantons". As above, they are defined by a first order condition on the connection and trivially satisfy the YM equations. Again, this condition has implications on the functional. We can understand this in the setting of Hermitian bundles by starting from the formula

$$[\mathrm{tr}(F^2_\nabla)] = 8\pi^2\{c_2(E) - \frac{1}{2}c_1(E)^2\} \in H^4(X;\mathbb{R}).$$

To simplify, assume $G = \mathrm{SU}(r)$ so that $c_1(E) = 0$. Notice that

$$\begin{aligned}\mathrm{tr}(F^2_\nabla) &= \mathrm{tr}(F^+_\nabla \wedge F^+_\nabla) + \mathrm{tr}(F^-_\nabla \wedge F^-_\nabla) + 2\,\mathrm{tr}(F^+_\nabla \wedge F^-_\nabla)\\ F^+_\nabla \wedge F^-_\nabla &= -F^+_\nabla \wedge \star F^-_\nabla = -g(F^+_\nabla, F^-_\nabla)\,\mathrm{vol}_g = 0.\end{aligned}$$

This shows that

$$\int_M \mathrm{tr}(F^+_\nabla \wedge F^+_\nabla) + \mathrm{tr}(F^-_\nabla \wedge F^-_\nabla) = 8\pi^2 \int_M c_2(E).$$

Using the same metric as in the YM functional, we thus find

$$\int_M (-\|F^+_\nabla\|^2 + \|F^-_\nabla\|^2)\,\mathrm{vol}_g = 8\pi^2 \int_M c_2(E).$$

It follows that

$$\begin{aligned}\int_M \|F_\nabla\|^2\,\mathrm{vol}_g &= \int_M \|F^+_\nabla\|^2\,\mathrm{vol}_g + \int_M \|F^-_\nabla\|^2\,\mathrm{vol}_g\\ &= 2\int_M \|F^+_\nabla\|^2\,\mathrm{vol}_g - \int_M \|F^+_\nabla\|^2\,\mathrm{vol}_g + \int_M \|F^-_\nabla\|^2\,\mathrm{vol}_g\\ &= 2\int_M \|F^+_\nabla\|^2\,\mathrm{vol}_g + 8\pi^2 \int_M c_2(E).\end{aligned}$$

We thus find a topological lower bound for the YM functional:

$$YM(\nabla) \geq 8\pi^2 \int_M c_2(E).$$

The bound is achieved precisely when F_∇ is ASD, proving that these connections are absolute minimizers.

To summarize: once again, appropriate structures on M (4-dimensional oriented Riemannian) imply strong algebraic control over the Yang-Mills equations.

Remark. The above formula emphasizes the strong relationship between the YM functional and characteristic classes, based on the fact that they can both be written in terms of curvature.

Analogous situations arise also for SD connections and for other groups G, provided one works in terms of the appropriate characteristic class, see Donaldson 2.1.4.

Example. The Atiyah-Drinfeld-Hitchin-Manin (ADHM) construction provides a 1:1 correspondence between ASD connections on SU(r)-bundles over $\mathbb{R}^4$, whose curvature has finite energy, and certain algebraic data. A removible singularity theorem shows that the bundles and connections can be smoothly extended to $\mathbb{S}^4$, endowed with its standard metric. This provides an explicit construction of all instantons on $\mathbb{S}^4$.

The ASD condition is particularly interesting in the case where M is a Kähler surface and E is holomorphic. Indeed, recall the Hermitian-Einstein equation

$$F_\nabla \wedge \omega^{n-1} = -2\pi i\,\lambda\,Id_E\omega^n$$

over a Kähler manifold M. As already seen, when M has complex dimension 1 the case $\lambda = 0$ corresponds to flat connections. When M has complex dimension 2 the case $\lambda = 0$ corresponds instead to the condition $F_\nabla \wedge \omega = 0$. Since F_∇ is Hermitian and E is holomorphic, F_∇ is automatically of type (1,1) so the equation means that F_∇ is ASD. As usual in the HE context, the existence of such connections is related to the stability of E.

Moduli spaces and the geometry of 4-manifolds. It can be shown that (non-empty) moduli spaces of ASD connections are smooth (for generic conformal classes of metrics on M, and except at points where the connection is reducible) and finite-dimensional (thanks to the existence of a Coulomb gauge, which makes the equations elliptic). The dimension can be computed via index formulae. Below, we will discuss how these moduli spaces can be compactified.

The compactified moduli spaces depend on the conformal class of the specific metric on M, but changing the conformal class generally leads to cobordisms between the moduli spaces. This feature was used by Donaldson to extract invariants from these moduli spaces which are independent of the metric. He then used them to study the intersection form of compact, oriented 4-manifolds.

The main features of the problem come to light already when the manifold is simply-connected, even though this is an extreme simplification given the fact that any finitely-presented group can arise as the fundamental group of a 4-manifold.

Let M be a compact, simply-connected, oriented 4-manifold. In this case the cohomology groups define only one interesting number, b^2. The only available topological invariant is thus the intersection form. The main questions revolve around the relationship between the abstract classification of integral symmetric bilinear forms, and the topological or smooth classification of such manifolds: specifically, the existence and uniqueness of a manifold whose intersection form coincides with a given abstract form.

One basic difficulty is that (when M is simply connected) one can show that the intersection form automatically encodes all information obtainable by characteristic classes so the latter, although defined in terms of the tangent bundle thus the smooth structure, are actually not able to distinguish different smooth structures on the same topological manifold. In other words, the integral bilinear form encodes all classical invariants.

In the course of the '80s it eventually turned out that the topological and smooth categories are actually very different. On the one hand, roughly speaking, Freedman proved a 1:1 relationship between topological manifolds and abstract integral forms. On the other hand Donaldson used the properties of ASD moduli spaces to show that certain integral forms are not compatible with any smooth structure.

Donaldson also used gauge theory to introduce new invariants of a given smooth structure. Roughly speaking, these invariants are created in two steps. One first chooses a bundle and a Riemannian metric on M so as to create ASD moduli spaces (which depend only on the conformal class). One then extracts information from the moduli spaces which depends only on the smooth structure, not on the conformal class. Donaldson encoded this information in terms of polynomials defined on the integral groups H^2. He then used these invariants to show that certain integral forms, ie topological manifolds, support several different smooth structures. The argument is along the following lines.

On the one hand he showed that, in many cases, if a smooth manifold can be written as a connect sum then these invariants are trivial. On the other hand he compared some such manifolds to certain Kähler surfaces, noticing that they have the same intersection form, thus the same topology. For many Kähler surfaces the Hitchin-Kobayashi correspondence allows one to calculate the ASD moduli spaces and the invariants in terms of the moduli space of stable bundles. Donaldson thus managed to show that the invariants of these Kähler surfaces are non-trivial. Since the invariants are different, the smooth structures must be different.

In summary, the methods of differential topology were not sufficient to distinguish the topological and the smooth categories: this required geometry.

9 Bubbling phenomena

Moduli spaces of YM connections are typically not compact. There are two reasons for this. First, they are invariant under the non-compact group of gauge transformations. This means that we can hope for compactness only after quotienting by this action. The second reason is more complicated, and is generally summed up in the catch-word "bubbling". This term appears in several contexts within geometric analysis: the geometries may be very different, but they are linked by analogous analytic properties. The bottom line is that, in all cases, even though the spaces in question are not compact, bounded sequences have special properties so that one can understand fairly precisely what goes wrong in the limit, thus find a way to compensate. In the best-case scenario one can build a geometric description of the "boundary" of the space in question, generated by all possible limits.

Holomorphic curves. One of the simplest appearances of bubbling occurs when dealing with holomorphic curves in Kähler manifolds (more generally, pseudo-holomorphic curves in symplectic manifolds). We attempt to give a quick presentation in this context.

Our starting point will be smooth maps $u : \Sigma \to M$ between Riemannian manifolds and the energy functional $E(u) := \frac{1}{2}\int_\Sigma |Du|^2 \operatorname{vol}_\Sigma$. An important property of this functional is that, when $dim(\Sigma) = 2$, it depends only on the conformal class of the metric on Σ. This means that it is well-defined when Σ is a Riemann surface.

In general, the critical points of E are the harmonic maps. When Σ is a Riemann surface and M is Kähler (more generally: symplectic with a compatible almost-complex structure), however, there is a special class of critical points defined by a different, first-order, equation: indeed, the energy identity

$$E(u) := \int_\Sigma |\bar{\partial} u|^2 \operatorname{vol}_\Sigma + \int_\Sigma u^*\omega,$$

shows that, within a given homology class of maps, E has a topological lower bound achieved precisely by the holomorphic curves in that class: analogies with flat/HE/ASD connections and the Yang-Mills functional should be clear.

In other words, in the geometric context of Kähler/symplectic geometry, the PDE we are interested in (the CR equations) appear as a special class of Euler-Lagrange equations. This allows us to rely on Calculus of Variations type arguments, rather than only on PDE theory, to prove (i) existence of solutions (see also Hilbert's 22nd problem) and (ii) compactness of the moduli space of solutions. In the case in question, it is for example reasonable to hope that holomorphic curves exist as limits of energy-minimizing sequences. Similarly, given a sequence of holomorphic curves, one might hope to use energy bounds to prove the existence of a convergent subsequence. We remark that energy bounds are particularly natural in this context: a L^2 bound on a solution to a first-order elliptic system automatically implies a $L^{1,2}$ bound, and this happens uniformly with respect to sequences.

Notice that, in some cases, energy bounds do immediately lead to existence. Consider the case of harmonic curves, ie geodesics: a minimizing sequence of smooth curves admits an energy bound by definition. This leads to a $W^{1,2}$ bound, thus a $C^{0,\alpha}$ bound via the Sobolev embedding theorems. Ascoli-Arzelà then produces a minimizing curve, ie a geodesic (easier arguments also exist, which do not rely on the Sobolev embedding theorems). In our case, however, $n = 2$ happens to be the threshold case where the Sobolev embedding theorems corresponding to $p = 2$, which is the case relevant to energy, cease to hold. Existence results for harmonic surfaces are thus highly non-trivial: the first are due to Sacks and Uhlenbeck (1981), who worked with perturbed energy functionals defined in terms of $p > 2$, obtained solutions as above, then studied sequences of such solutions for $p_n \to 2$. The convergence argument for such sequences requires very strong, uniform, control over these solutions and all their derivatives, obtained in the so-called "small energy theorem".

Analogous considerations hold for compactness results. It is of course obvious that spaces of continuous maps are generally not compact, even in the best-case scenarios where the domain/range are compact or the maps have extra properties. Perhaps the simplest examples are the sequence of continuous functions $x^n : [0,1] \to [0,1]$, or of holomorphic maps $u_n : \mathbb{CP}^1 \to \mathbb{CP}^1$, $z \mapsto nz$. Once again, the underlying fact is that C^0-bounds do not suffice to enforce compactness: one needs some control over the derivatives, either pointwise or integral: one then applies the Ascoli-Arzelà criterion, as above. As already seen, however, this technique fails for harmonic surfaces. More importantly, convergence in the usual sense is actually false. The point we now need to make is that it is possible to ensure a suitably modified notion of compactness.

Example. Consider the sequence of holomorphic maps $u_n : \mathbb{CP}^1 \to \mathbb{CP}^1$, $z \mapsto nz$. We observe the following facts: (i) The pointwise limit is the non-continuous map which fixes 0 and sends all other points to ∞; in particular, there is no convergent subsequence in the C^0-topology. (ii) Analytically, the problem is that the derivatives du_n are not pointwise bounded, eg in the point 0. The energy $E(u_n)$ is however bounded, actually constant: this follows from the energy identity. (iii) Geometrically, we can interpret the blow-up of $du_n(0)$ by imagining that a small neighbourhood of the fixed point 0 is being stretched by the maps so as to cover greater and greater parts of the sphere. (iv) There is a simple way to locally counter-balance this: on any given neighbourhood of 0 we can rescale via $w \mapsto z := w/n$: the composed sequence of local maps is constantly equal to the identity map, which extends smoothly to the identity on the sphere.

We conclude as follows. The naive limit is non-continuous and does not detect the geometry of the sequence. A different limiting object is the pair of holomorphic maps (∞, Id) defined on the union of two spheres. We view the constant map ∞ as the smooth extension of the naive limit, on the original sphere. The new sphere can be thought of as a "bubble" which emerged out of the point $z = 0$ on the original sphere (as in the geometric notion of blow-up). If we imagine the two spheres as being attached by identifying the point $z = 0$ on the original sphere with the point $w = \infty$ on the new sphere, the corresponding union of the two maps is continuous. Notice that the original sequence, on the original sphere, converges to the constant map ∞ uniformly away from 0 and that this constant map has strictly less energy than those in the sequence. The other map Id exactly compensates for the missing energy.

It turns out that this example is quite general. By the above, if a sequence of holomorphic maps u_n from a compact Riemann surface Σ to a Kähler (more generally, symplectic) manifold M, with bounded energy, does not converge, it must admit a finite number of sequences $z_n^i \in \Sigma$ along which du_n blows up. Away from the limit points z^i the maps converge uniformly to a holomorphic map with lower energy. The missing energy can be recovered via bubbles at the points z_i. Since each bubble is obtained simply by rescaling, their images are approximated by the images of u_n. A geometric formulation of this behaviour is achieved by producing a topology on the moduli space of holomorphic maps such that objects of this sort, ie holomorphic maps from Riemann surfaces with extra spheres, appear in the boundary of the moduli space.

Two final ingredients needed to form this picture are the conformal invariance of the energy functional and a removable singularity theorem. Since rescaling is a conformal transformation, the former implies that the bubbles have finite energy. The latter shows both that the limiting holomorphic maps (such as the constant map ∞ in the example above) can be smoothly extended in each z^i, and that the rescaled maps extend from $\mathbb{C}$ to $\mathbb{CP}^1$, creating the holomorphic bubbles.

We remark that parts of this picture hold also for the harmonic map flow of a surface into a Riemannian manifold. Specifically, singularities at time T correspond to the formation of harmonic $\mathbb{S}^2$ bubbles at isolated points of the surface. There exists a limiting map u_T and in some cases one can prove that it has removable discontinuities at those points.

YM connections. An analogous picture for YM connections, including for example threshold Sobolev spaces, conformally invariant functionals, and small energy and removable singularity theorems, emerged via the work of Uhlenbeck and Nakajima in the '80s. A new feature in this case is the action of the gauge group. As already mentioned it is necessary to work modulo this action. This is done by using special local trivializations of the bundle: these "Coulomb gauge" trivializations can be thought of as analogous, on manifolds, to harmonic coordinates. In particular, they ensure that the relevant equations are elliptic.

Assuming M is compact and we are given a sequence of YM connections with bounded YM energy, the situation is as follows. When $dim(M) = 2, 3$, compactness holds, ie no bubbling occurs. The case $dim(M) = 4$ corresponds to the critical threshold case. Here, bubbling occurs in a finite number of points p^i. Specifically, up to gauge transformations and subsequences, any sequence of YM connections converges C^∞-uniformly away from p^i, and extends smoothly in those points to a limiting YM connection. Two issues need to be mentioned: the bundle corresponding to this limiting connection might not be isomorphic to the original bundle, and the YM energy of this limit is strictly lower than the liminf of the YM energies of the original sequence. At each point p^i one can locally rescale the connections, obtaining a YM connection on $\mathbb{R}^4$ endowed with the pull-back bundle which extends to $\mathbb{S}^4$: this is the YM bubble. The sum of energies of the limit connection and of all bubbles produce the expected total energy.

In the context of ASD connections, limits and bubbles are again ASD.

A similar picture holds when $dim(M) > 4$ but in this case the set of bubbling points in M has Hausdorff codimension 4. In this case there is no guarantee that the limit connection extends over all such points, nor that the rescaled connections compactify from $\mathbb{R}^n$ to $\mathbb{S}^n$: one must thus distinguish between the true bubbling locus, ie the points where such extensions hold, and singular points, where either fails.

Some of these results hold also for sequences of general (not YM) connections with bounded energy, but in general one obtains weaker convergence results and no extension theorems.

We note in closing that the notion of bubbling originated in the 1981 Sacks-Uhlenbeck work. They used the formation of bubbles to prove the existence of minimal spheres in Riemannian manifolds. Analogously, in the context of gauge theory on a four-dimensional manifold, Sedlacek used the formation of bubbles to prove the existence of YM connections on $\mathbb{S}^4$.

10 Closing comments

These notes are still a draft. They have been checked, but certainly not checked enough. The presentation is naive by choice, but at times possibly stupid. I would be happy for any comments or corrections.

I would have liked to include a section on the relationship between gauge theory and calibrated geometry, as in the work of Tian or as in the currently-developing theory of G_2 instantons. I would also have liked to discuss the relevance of dual tori to Mirror Symmetry, including also a presentation of the dHYM equation. Sadly I have no time (nor competence, nor energy) to do so at this moment.

References I have either used or might wish to look at in the future include the following list.

Atiyah and Bott, *Yang-Mills equations over Riemann surfaces* (article).
Donaldson and Kronheimer, *The Geometry of 4-Manifolds* (book).
Donaldson, *Karen Uhlenbeck and the Calculus of Variations* (article, Notices AMS).
Donaldson, *Mathematical uses of gauge theory* (article).
Donaldson, *Stability of algebraic varieties and Kähler geometry* (article).
Parker, *What is... a bubble tree* (article, Notices AMS).
Schaffhauser, *Differential geometry of holomorphic vector bundles on a curve* (article, arxiv).
Kobayashi, *Differential geometry of complex vector bundles* (book).
Guichard, *An Introduction to the Differential Geometry of Flat Bundles and of Higgs Bundles* (article, arxiv).
Greb, Sibley, Toma and Wentworth, *Complex algebraic compactifications of the moduli space of HYM connections on a projective manifold* (article, arxiv).
Song and Waldron, https://arxiv.org/abs/2009.07242.
Thaddeus, *An introduction to the topology of the moduli space of stable bundles on a Riemann surface* (article).
Itoh and Nakajima, *Yang-Mills Connections and Einstein-Hermitian Metrics* (article).
Hu and Li, *Variation of the Gieseker and Uhlenbeck compactifications* (article).

Greb, Ross and Toma, *Moduli of vector bundles on higher-dimensional base manifolds - construction and variation* (article).
Thomas, *Notes on GIT and symplectic reduction for bundles and varieties* (article).