

# MORSE THEORY, HIGGS FIELDS, AND YANG-MILLS-HIGGS FUNCTIONALS

STEVEN B. BRADLOW AND GRAEME WILKIN

*Dedicated to Professor Dick Palais on the occasion of his 80th birthday*

## 1. INTRODUCTION

In classical Morse theory, manifolds are smooth and finite dimensional, and critical points are isolated and non-degenerate. Nearly 50 years ago Palais expanded the theory to cover infinite dimensional manifolds. Subsequent developments showed how to cope if critical points are replaced by non-degenerate critical submanifolds (Bott, [6]) and if the non-degeneracy condition is relaxed to allow (possibly singular) critical sets that are contained in a minimising manifold (Kirwan, [37]). The full range of such phenomena appear in the settings described in this paper, with key ideas descending directly from Palais's work. Given that the first author is a mathematical grandchild and the second author is from the next generation in the Palais family tree, this birthday offering is thus both by, and about, Palais's mathematical descendants.

The setting for this paper is the theory of Higgs bundles over a closed Riemann surface, say  $X$ . Higgs bundles were first introduced by Nigel Hitchin in his landmark 1987 paper [34], though the name was first introduced by the other pioneer in the subject, namely Carlos Simpson, in his 1988 paper [53]. In both cases, the objects introduced correspond to Higgs bundles associated to complex groups (actually  $\mathrm{SL}(n, \mathbb{C})$ ); in this paper we will adopt the more general notion (introduced by Hitchin in [35]) of Higgs bundles for real reductive Lie groups. We call these objects  $G$ -Higgs bundles, where  $G$  is the real group. This will allow us to explore a wider range of phenomena. In the spirit of this review, we focus on phenomena detected using Morse-theoretic ideas.

The main objects of study are the finite dimensional moduli spaces of  $G$ -Higgs bundles. A fundamental feature of these spaces is that they admit several complementary interpretations. They have an algebraic interpretation as a moduli space of polystable objects in the sense of Geometric Invariant Theory, but can also be viewed as moduli spaces of solutions to gauge-theoretic equations. In some cases they can be identified with a representation variety for representations of the fundamental group  $\pi_1(X)$ . Related to these different points of view, there are two natural ways to apply Morse-theoretic ideas to the study of the topology of these moduli spaces.

The first way is to view the moduli space as a quotient space of the infinite dimensional space of connections and Higgs fields on a principal bundle. The Morse function in this setting is a natural energy functional, the so-called Yang-Mills-Higgs functional, which, crucially, satisfies an equivariant version of the famous Palais-Smale Condition

$\mathbb{C}$  along complex gauge orbits. The absolute minimum of this functional is a gauge invariant analytic subspace and the orbit space is the moduli space of interest. This finite dimensional moduli space has its own natural choice of Morse function known as the Hitchin function. The second type of Morse theory is based on this Hitchin function.

After introducing  $G$ -Higgs bundles in Section 2, in Section 3 we describe the infinite dimensional Morse theory for the Yang-Mills-Higgs functional and in Section 4 we discuss the finite dimensional Morse Theory for the Hitchin function. This paper is mostly a review of a still-evolving body of work and and, apart from Theorem 3.8 in Section 3.2, contains essentially no new results.

## 2. $G$ -HIGGS BUNDLES

**2.1. Definitions.** Let  $G$  be a **real reductive Lie group**. While the precise definition requires some care (see for example Knapp [38, p. 384]), it is sufficient for our purposes to take  $G$  to be the real form of a complex reductive group. In fact, in this paper the main complex groups that will appear are  $\mathrm{GL}(n, \mathbb{C})$  and  $\mathrm{SL}(n, \mathbb{C})$ . The definition of a  $G$ -Higgs bundle requires the following additional data:

- a maximal compact subgroup  $H \subset G$
- a Cartan involution  $\theta$  giving a decomposition of the Lie algebra

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$$

into its  $\pm 1$ -eigenspaces, where  $\mathfrak{h}$  is the Lie algebra of  $H$ , and

- a non-degenerate Ad-invariant and  $\theta$ -invariant bilinear form,  $B$ , on  $\mathfrak{g}$ .

Given the above choices, the Lie algebra satisfies

$$[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h}, \quad [\mathfrak{h}, \mathfrak{m}] \subset \mathfrak{m}, \quad [\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{h}.$$

and hence the adjoint representation restricts to define a linear representation of  $H$  on  $\mathfrak{m}$ . This action extends to a linear holomorphic action

$$\iota: H^{\mathbb{C}} \rightarrow \mathrm{GL}(\mathfrak{m}^{\mathbb{C}}). \tag{2.1}$$

called the **isotropy representation**.

**Definition 2.1.** A  $G$ -Higgs bundle on  $X$  is a pair  $(E_{H^{\mathbb{C}}}, \varphi)$ , where  $E_{H^{\mathbb{C}}}$  is a holomorphic  $H^{\mathbb{C}}$ -principal bundle over  $X$  and  $\varphi$  is a holomorphic section of  $E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}}) \otimes K$ , where  $E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}}) = E_{H^{\mathbb{C}}} \times_{H^{\mathbb{C}}} \mathfrak{m}^{\mathbb{C}}$  is the  $\mathfrak{m}^{\mathbb{C}}$ -bundle associated to  $E_{H^{\mathbb{C}}}$  via the isotropy representation and  $K$  is the canonical bundle of  $X$ .

*Example 2.2.* If  $G$  is compact then  $H = G$  and  $\mathfrak{m} = \{0\}$ . A  $G$ -Higgs bundle is thus equivalent to a holomorphic  $G^{\mathbb{C}}$ -bundle.

*Example 2.3.* If  $G$  is the underlying real group of a complex reductive group then  $H^{\mathbb{C}} = G$  and  $\mathfrak{m}^{\mathbb{C}} = i\mathfrak{g} = \mathfrak{g}$ . In this case a  $G$ -Higgs bundle is a pair  $(E, \varphi)$ , where  $E$  is a holomorphic  $G$ -principal bundle over  $X$  and  $\varphi$  is a holomorphic section of  $ad(E) \otimes K$ . In particular, if  $G = \mathrm{GL}(n, \mathbb{C})$  and we replace the principal  $G$ -bundle by the rank  $n$  vector bundle determined by the standard representation, then a  $\mathrm{GL}(n, \mathbb{C})$ -Higgs bundle is a pair  $(V, \varphi)$ , where  $V$  is a rank  $n$  holomorphic vector bundle over  $X$  and  $\varphi$  is a

$G$	$H$	$H^{\mathbb{C}}$	$\mathfrak{m}^{\mathbb{C}}$
$\mathrm{SL}(n, \mathbb{R})$	$\mathrm{SO}(n)$	$\mathrm{SO}(n, \mathbb{C})$	$\mathrm{Hom}_0^{\mathrm{sym}}(\mathbb{C}^n, \mathbb{C}^n)$
$\mathrm{Sp}(2n, \mathbb{R})$	$\mathrm{U}(n)$	$\mathrm{GL}(n, \mathbb{C})$	$\mathrm{Sym}^2(\mathbb{C}^n) \oplus \mathrm{Sym}^2((\mathbb{C}^n)^*)$
$\mathrm{U}(p, q)$	$\mathrm{U}(p) \times \mathrm{U}(q)$	$\mathrm{GL}(p, \mathbb{C}) \times \mathrm{GL}(q, \mathbb{C})$	$\mathrm{Hom}(\mathbb{C}^p, \mathbb{C}^q) \oplus \mathrm{Hom}(\mathbb{C}^q, \mathbb{C}^p)$

TABLE 1. Examples of Cartan decompositions of real reductive groups

$G$	Bundle(s)	Higgs field
$\mathrm{SL}(n, \mathbb{R})$	$V$ (rank $n$ ) orthogonal	$\varphi : V \longrightarrow V \otimes K$ symmetric, $\mathrm{Tr}(\varphi) = 0$
$\mathrm{Sp}(2n, \mathbb{R})$	$V$ (rank $n$ ) (degree $d$ )	$\varphi = (\beta, \gamma), \begin{cases} \beta : V^* \longrightarrow V \otimes K \\ \gamma : V \longrightarrow V^* \otimes K \end{cases}$ $\beta^t = \beta; \gamma^t = \gamma$
$\mathrm{U}(p, q)$	$V_1$ (rank $p$ ) $V_2$ (rank $q$ )	$\varphi = (\beta, \gamma), \begin{cases} \beta : V_1 \longrightarrow V_2 \otimes K \\ \gamma : V_2 \longrightarrow V_1 \otimes K \end{cases}$

 TABLE 2. The  $G$ -Higgs bundles with  $E_{H^{\mathbb{C}}}$  replaced by vector bundle(s)

holomorphic map  $\varphi : V \rightarrow V \otimes K$ . If we take  $G = \mathrm{SL}(n, \mathbb{C})$  then an  $\mathrm{SL}(n, \mathbb{C})$ -**Higgs bundle** is also a pair  $(V, \varphi)$  but  $\det(V)$  is trivial and  $\varphi$  has zero trace.

*Example 2.4.* In Table 1 we describe the structure of  $\mathrm{SL}(n, \mathbb{R})$ ,  $\mathrm{Sp}(2n, \mathbb{R})$  and  $\mathrm{U}(p, q)$  as real reductive groups. In each case there is a Cartan involution given by  $\theta(u) = -u^*$ , which induces the Cartan decomposition given in the table. Table 2 describes the associated  $G$ -Higgs bundles and their vector bundle representation for these examples of real reductive groups. Thus for example the data defining a  $\mathrm{U}(p, q)$ -Higgs bundle on a Riemann surface  $X$  is equivalent to the tuple  $(V_1, V_2, \beta, \gamma)$  where  $V_1$  and  $V_2$  are respectively rank  $p$  and rank  $q$  holomorphic vector bundles on  $X$  and the maps

$$\begin{aligned} \beta : V_1 &\longrightarrow V_2 \otimes K \\ \gamma : V_2 &\longrightarrow V_1 \otimes K \end{aligned}$$

define the Higgs field (here  $K$  is the canonical bundle on  $X$ ).<sup>1</sup> The case of  $\mathrm{Sp}(2n, \mathbb{R})$  is explained further in Example 4.7.

**2.2. Moduli spaces.** Two  $G$ -Higgs pairs  $(E_{H^{\mathbb{C}}}, \varphi)$  and  $(E'_{H^{\mathbb{C}}}, \varphi')$  are **isomorphic** if there is an isomorphism  $f : E_{H^{\mathbb{C}}} \xrightarrow{\sim} E'_{H^{\mathbb{C}}}$  such that  $\varphi = f^* \varphi'$  where  $f^*$  is the obvious induced map. Loosely speaking, a moduli space for  $G$ -Higgs bundles is a geometric space whose points parameterize isomorphism classes of pairs  $(E, \varphi)$ . In general the space of all isomorphism classes has bad topological properties; for example, it may not even be Hausdorff. One standard solution to this problem is provided by Geometric Invariant Theory (GIT) notions of stability and polystability. The point of these stability properties is to identify the isomorphism classes for which a suitable moduli

<sup>1</sup>For more details see [10], [26], [35].

space can be constructed. Constructions of this sort for moduli spaces of holomorphic bundles have a long history going back to the work of Narasimhan and Seshadri ([44]). The appropriate notions of stability and polystability have been formulated in very general terms for  $G$ -Higgs bundles in [27]. While the general definitions are quite cumbersome, in many special cases they simplify to conditions akin to slope stability/semistability/polystability for holomorphic bundles.

*Example 2.5.* Let  $G = \mathrm{GL}(n, \mathbb{C})$  and let  $(V, \Phi)$  be a  $G$ -Higgs bundle over a Riemann surface  $X$  (as in Example 2.3). The Higgs bundle  $(V, \Phi)$  is defined to be semistable if

$$\frac{\deg(V')}{\mathrm{rank}(V')} \leq \frac{\deg(V)}{\mathrm{rank}(V)} \quad (2.2)$$

for all  $\Phi$ -invariant subbundles  $V' \subset V$ . The Higgs bundle is stable if the inequality in (2.2) is strict for all proper invariant subbundles, and it is polystable if it decomposes as a direct sum of stable Higgs bundles all with the same slope (where the slope of a bundle  $V$  is the ratio  $\frac{\deg(V)}{\mathrm{rank}(V)}$ ).

**Definition 2.6.** The **moduli space of polystable  $G$ -Higgs bundles**  $\mathcal{M}(G)$  is the set of isomorphism classes of polystable  $G$ -Higgs bundles  $(E_{H^{\mathbb{C}}}, \varphi)$ .

The topological type of the bundle  $E_{H^{\mathbb{C}}}$  is an invariant of connected components of  $\mathcal{M}(G)$  so we refine our definition to identify components by the possible topological types. The topological types can be identified by suitable characteristic classes such as Chern classes (if  $H^{\mathbb{C}} = \mathrm{GL}(n, \mathbb{C})$ ) or Stiefel-Whitney classes (for  $\mathrm{O}(n, \mathbb{C})$ -bundles). If  $G$  is semisimple and connected the topological type of a principal  $H^{\mathbb{C}}$ -bundle on  $X$  is classified by an element in  $\pi_1(H^{\mathbb{C}}) = \pi_1(H) = \pi_1(G)$ . We will loosely denote the characteristic class by  $c(E_{H^{\mathbb{C}}})$  and denote its value by  $c(E_{H^{\mathbb{C}}}) = d$ . We can thus define;

**Definition 2.7.** The component  $\mathcal{M}_d(G) \subset \mathcal{M}(G)$ , called the **moduli space of polystable  $G$ -Higgs bundles of type  $d$**  is the set of isomorphism classes of polystable  $G$ -Higgs bundles  $(E_{H^{\mathbb{C}}}, \varphi)$  such that  $c(E_{H^{\mathbb{C}}}) = d$ .

*Remark 2.8.* When  $G$  is compact, the moduli space  $\mathcal{M}_d(G)$  coincides with  $M_d(G^{\mathbb{C}})$ , the moduli space of polystable principal  $G^{\mathbb{C}}$ -bundles with topological invariant  $d$ .

The moduli space  $\mathcal{M}_d(G)$  has the structure of a complex analytic variety. This can be seen by the standard slice method (see, e.g., Kobayashi [39, Ch. VII]). Geometric Invariant Theory constructions are available in the literature for  $G$  real compact algebraic (Ramanathan [48, 49]) and for  $G$  complex reductive algebraic (Simpson [55, 56]). The case of a real form of a complex reductive algebraic Lie group follows from the general constructions of Schmitt [50]. We thus have the following.

**Theorem 2.9.** *The moduli space  $\mathcal{M}_d(G)$  is a complex analytic variety, which is algebraic when  $G$  is algebraic.*

The moduli spaces are not in general smooth but the singularities can be identified. The singular points are associated with the  $G$ -Higgs bundles which are polystable but not stable. This allows the moduli spaces to be smooth in special cases where polystability implies stability. For example in the case of  $G = \mathrm{GL}(n, \mathbb{C})$  (as in Example 2.5) if

the rank and degree of the underlying vector bundle are coprime then polystability is equivalent to stability and hence the moduli spaces are smooth.

The moduli spaces can also be viewed as quotients in an infinite dimensional gauge-theoretic setting. Fixing a smooth principal  $H^{\mathbb{C}}$ -bundle  $E$  with topological invariant  $c(E) = d$ , the holomorphic principal  $H^{\mathbb{C}}$ -bundles with the same topological invariant can be described by a choice of holomorphic structure on  $E$ . If  $E$  is a vector bundle then holomorphic structures are defined by  $\mathbb{C}$ -linear operators  $\bar{\partial}_E : \Omega^0(E) \rightarrow \Omega^{0,1}(E)$  that satisfy the Leibniz rule<sup>2</sup>. This construction can be adapted for principal bundles, and in all cases the space of holomorphic structures, denoted by  $\mathcal{A}^{0,1}(E_{H^{\mathbb{C}}})$ , is an infinite dimensional affine space locally modeled on the space of anti-holomorphic 1-forms with values in the endomorphism bundle,  $\Omega^{0,1}(E_{H^{\mathbb{C}}}(\mathfrak{h}^{\mathbb{C}}))$ . Moreover a given holomorphic structure induces an operator  $\bar{\partial}_E$  on  $\Omega^{1,0}(E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}}))$  such that the condition for a 1-form with values in  $E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}})$  to be holomorphic is  $\bar{\partial}_E \varphi = 0$ .

**Definition 2.10.** The **configuration space of  $G$ -Higgs bundles on  $E$** , i.e. the space of all  $G$ -Higgs bundles for which the principal  $H^{\mathbb{C}}$ -bundle has topological invariant  $d$ , is the space

$$\mathcal{B}_d(G) = \{(\bar{\partial}_E, \varphi) \in \mathcal{A}^{0,1}(E_{H^{\mathbb{C}}}) \times \Omega^{1,0}(E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}})) \mid \bar{\partial}_E \varphi = 0\} \quad (2.3)$$

The complex gauge group for  $E_{H^{\mathbb{C}}}$ , i.e. the group of sections  $\mathcal{G}^{\mathbb{C}} = \Omega^0(Ad(E_{H^{\mathbb{C}}}))$ , acts on  $\mathcal{B}_d(G)$  in such a way that  $\mathcal{G}^{\mathbb{C}}$ -orbits correspond to isomorphism classes of  $G$ -Higgs bundles. Thus

$$\mathcal{M}_d(G) = \mathcal{B}_d^{ps}(G) / \mathcal{G}^{\mathbb{C}} \quad (2.4)$$

where  $\mathcal{B}_d^{ps}(G) \subset \mathcal{B}_d(G)$  denotes the set of polystable  $G$ -Higgs bundles.

The Hitchin-Kobayashi correspondence for Higgs bundles identifies polystability with the existence of solutions to gauge theoretic equations, and thus yields yet another interpretation of the moduli spaces. There are two ways to formulate the equations. In the first, one needs to fix a reduction of structure group on  $E_{H^{\mathbb{C}}}$  to the compact group  $H$ . If  $H^{\mathbb{C}} = \mathrm{GL}(n, \mathbb{C})$  so  $H = U(n)$  then this corresponds to fixing a hermitian bundle metric on the associated rank  $n$  vector bundle. We thus refer to this in all cases as a choice of metric. The metric allows us to write  $E_{H^{\mathbb{C}}} = E_H \times_H H^{\mathbb{C}}$ , where  $E_H$  is a principal  $H$ -bundle, and defines the real gauge group  $\mathcal{G} \subset \mathcal{G}^{\mathbb{C}}$  consisting of the sections of  $\Omega^0(Ad(E_H))$ . It provides, through the construction of Chern connections, an identification between holomorphic structures on  $E_{H^{\mathbb{C}}}$  and connections on  $E_H$ . Denoting the space of connections on  $E_H$  by  $\mathcal{A}(E_H)$ , we can thus formulate the following set of equations for pairs  $(A, \varphi) \in \mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}}))$ :

$$F_A - [\varphi, \tau(\varphi)] = c\omega \quad (2.5)$$

$$d_A^{0,1} \varphi = 0 \quad (2.6)$$

---

<sup>2</sup> In general there is also an integrability condition but on Riemann surfaces this is automatically satisfied (see [45] for the original theorem on complex manifolds, and [4] for the bundle case).

Here  $F_A$  is the curvature of the connection,  $\tau$  denotes the compact conjugation on  $\mathfrak{g}^{\mathbb{C}}$ , i.e. the anti-linear involution which defines the compact real form,  $c$  is an element in the center of  $\mathfrak{h}$ ,  $\omega$  is the Kahler form on  $X$ , and  $d_A$  is the antiholomorphic part of the covariant derivative induced by  $A$  on  $E_{H^c}(\mathfrak{m}^{\mathbb{C}})$ . The constant  $c$  is determined, via Chern-Weil theory, by the topological invariant  $d$ . The operator  $d_A^{0,1}$  defines the holomorphic structure on  $E_{H^c}$ , so the second equation says that the Higgs field  $\varphi$  is holomorphic.

Going back to the identification

$$\mathcal{A}(E_H) \simeq \mathcal{A}^{0,1}(E_{H^c}) \quad (2.7)$$

determined by a choice of metric, we can regard the holomorphic structure and (holomorphic) Higgs field as given and view the equation (2.5) as an equation for a metric on  $E_{H^c}$ . The Hitchin-Kobayashi correspondence then asserts:

**Hitchin-Kobayashi Correspondence**<sup>3</sup> *The  $G$ -Higgs bundle  $(E_{H^c}, \varphi)$  is polystable if and only if  $E_{H^c}$  admits a metric satisfying the equation (2.5)*

On the other hand, the equation (2.5) can be interpreted as a symplectic moment map condition for the action of the gauge group  $\mathcal{G}$  on the infinite dimensional space  $\mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^c}(\mathfrak{m}^{\mathbb{C}}))$ . The symplectic form on this space is defined by the  $L^2$ -inner products constructed using the metric on  $E_H$  together with the metric on  $X$ . The resulting metric on  $\mathcal{B}(d)$  is Kähler and hence defines a symplectic structure on the smooth locus of the moduli space. The action of the real gauge group is hamiltonian with respect to this symplectic structure and has a moment map

$$\Psi : \mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^c}(\mathfrak{m}^{\mathbb{C}})) \rightarrow \mathrm{Lie}(\mathcal{G})^* \cong \Omega^2(\mathrm{ad}(E_H)) \quad (2.8)$$

$$(A, \varphi) \mapsto \Lambda(F_A - [\varphi, \tau(\varphi)]) - c \quad (2.9)$$

where  $\Lambda$  denotes contraction against the Kähler form  $\omega$ . If we define a subspace of  $\mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^c}(\mathfrak{m}^{\mathbb{C}}))$  by

$$\mathcal{B}_d^H(G) = \{(A, \varphi) \in \mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^c}(\mathfrak{m}^{\mathbb{C}})) \mid d_A^{0,1}\varphi = 0\} \quad (2.10)$$

and restrict the moment map to  $\mathcal{B}_d^H(G)$  we can thus describe the moduli space as a symplectic quotient, i.e

$$\mathcal{M}_d(G) = \Psi^{-1}(0)/\mathcal{G} \quad (2.11)$$

It is useful to make use of all three descriptions of the moduli spaces  $\mathcal{M}_d(G)$ , namely

- as a space of isomorphism classes of polystable objects (Definition 2.7),
- as a complex quotient (see (2.4)), and

---

<sup>3</sup>The original correspondence of this sort goes back to a theorem of Narasimhan and Seshadri ([44]) for vector bundles over closed Riemann surfaces, while the versions for bundles over closed Kahler manifolds go back to [40], [24], [60]. The scope of the result has by now been greatly extended to cover bundles with various kinds of additional data and also more general base manifolds (see [41] for a summary). For the versions most appropriate for  $G$ -Higgs bundles see [14], and [41].

- as a symplectic quotient (see (2.11))

There is one further interpretation of the moduli spaces which relies on the fact that while the Higgs field  $\varphi$  is a 1-form which takes its values in  $\mathfrak{m}^{\mathbb{C}}$ , the combination  $\theta = \varphi - \tau(\varphi)$  takes its values in  $\mathfrak{m}$ . It follows that

$$D = d_A + \theta \tag{2.12}$$

takes its values in  $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$  and defines a connection on the principal  $G$ -bundle

$$E_G = E_H \times_H G ,$$

i.e. on the bundle obtained from  $E_H$  by extending the structure group. The bundle  $E_G$  is thus, by construction, a  $G$  bundle with a reduction of structure group to  $H$ , i.e. with a metric. We can reformulate equations (2.5) and (2.6) in terms of  $D$  and the metric on  $E_G$ . If we assume that  $G$  is semisimple and hence that the right hand side of (2.5) is zero, then the equations become

$$\begin{aligned} F_D &= 0 \\ d_A^* \theta &= 0 . \end{aligned} \tag{2.13}$$

The first equation says that the connection  $D$  is flat and hence that its holonomy defines a representation

$$\rho : \pi_1(X) \rightarrow G . \tag{2.14}$$

Using the flat structure defined by  $D$ , the metric on  $E_G$  is equivalent to a  $\pi_1(X)$ -equivariant map from the universal cover of  $X$ :

$$\sigma : \tilde{X} \rightarrow G/H . \tag{2.15}$$

The second equation says that this map is *harmonic* with respect to the metric induced on  $\tilde{X}$  from that on  $X$  and the invariant metric on  $G/H$ . An important theorem of Corlette asserts

**Theorem 2.11.** [18] *The representation defined by (2.14) is reductive and all reductive representations of  $\pi_1(X)$  in  $G$  arise in this way, i.e. the corresponding flat  $G$ -bundles admit harmonic metrics.*

If  $X$  is a closed Riemann surface of genus  $g$  then  $\text{Hom}(\pi_1(X), G)$ , i.e. the set of all surface group representations in  $G$ , forms a real analytic subspace of the  $2g$ -fold product  $G \times \cdots \times G$ . The group  $G$  acts by conjugation on this space. The quotient  $\text{Hom}(\pi_1(X), G)/G$  does not in general have good geometric or even topological properties but restricting to the reductive representations yields a good **moduli space of representations**

$$\text{Rep}(\pi_1(X), G) = \text{Hom}^+(\pi_1(X), G)/G \tag{2.16}$$

(where the  $+$  denotes the restriction to the reductive representations). The topological invariant  $d \in \pi_1(G)$  which labels components of  $\mathcal{M}_d(G)$  also labels components of  $\text{Rep}(\pi_1(X), G)$ , leading to the obvious definition for spaces  $\text{Rep}_d(\pi_1(X), G)$ . For semisimple groups  $G$ , Corlette's theorem together with the equivalence between equations (2.13) and the Higgs bundle equations (2.5) and (2.6) thus allows us to identify

$$\mathcal{M}_d(G) = \text{Rep}_d(\pi_1(X), G) . \quad (2.17)$$

This identification, especially when coupled with the Morse theoretic methods described in the rest of this paper, turns Higgs bundles into an effective tool for studying the moduli spaces of surface group representations.

**2.3. Morse functions.** There are two functions which turn out to be useful for exploring the topology of the moduli spaces  $\mathcal{M}_d(G)$ , i.e. which have good Morse-theoretic properties. The first, defined on the infinite dimensional configuration space and denoted by

$$\text{YMH} : \mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^c}(\mathfrak{m}^{\mathbb{C}})) \rightarrow \mathbb{R} , \quad (2.18)$$

is a generalization of the Yang-Mills functional on the space of connections. The second, defined on  $\mathcal{M}_d(G)$  itself and denoted by

$$f : \mathcal{M}_d(G) \rightarrow \mathbb{R} , \quad (2.19)$$

was introduced by Hitchin and relies on the presence of the Higgs field. Both are related to symplectic moment maps. In the next sections we explore these two functions from the point of view of Morse theory.

### 3. THE YANG-MILLS-HIGGS FUNCTIONAL

The goal of this section is to describe recent developments on the infinite-dimensional Morse theory of the Yang-Mills-Higgs functional. Firstly, we recall the work of Atiyah & Bott on the Morse theory of the Yang-Mills functional for bundles over a compact Riemann surface.

**3.1. Background on the Yang-Mills functional and semistable holomorphic bundles.** In the late 1970s and early 1980s, Atiyah and Bott pioneered the use of the Yang-Mills functional as an equivariant Morse function to study the topology of the moduli space of semistable holomorphic bundles over a compact Riemann surface (see [2] and [3]).

**3.1.1. The algebraic picture.** The setup is as follows. Fix a smooth complex vector bundle  $E$  over a compact Riemann surface  $X$ . This corresponds to taking  $G = \text{U}(n)$  in the terminology of Section 2.2. As described in Section 2.2, the space  $\mathcal{A}^{0,1}(E)$  of holomorphic structures on  $E$  can be identified with an affine space locally modeled on  $\Omega^{0,1}(\text{End}(E))$ , i.e.

$$\mathcal{A}^{0,1}(E) \cong \bar{\partial}_{A_0} + \Omega^{0,1}(\text{End}(E))$$



for a fixed operator  $\bar{\partial}_{A_0}$ .

Any holomorphic bundle has a canonical *Harder-Narasimhan filtration*

$$X \times \{0\} = E_0 \subset E_1 \subset \cdots \subset E_k = E$$

by holomorphic sub-bundles, where, for each  $\ell = 1, \dots, k$ , the quotient bundle  $E_\ell/E_{\ell-1}$  is the maximal semistable sub-bundle of  $E/E_{\ell-1}$ . The *type* of the Harder-Narasimhan filtration is the vector  $(\nu_1, \dots, \nu_n)$ , consisting of the slopes of the quotient bundles  $E_\ell/E_{\ell-1}$  in the Harder-Narasimhan filtration, counted with multiplicity  $\text{rank}(E_\ell/E_{\ell-1})$ . Let  $\mathcal{A}_\lambda^{0,1}$  denote the set of holomorphic structures with Harder-Narasimhan type  $\lambda$ , and let  $\mathcal{A}_{ss}^{0,1}$  denote the semistable holomorphic structures. There is a partial ordering on the Harder-Narasimhan type (see [52]), and Atiyah and Bott show that the stratification

$$\mathcal{A}^{0,1} = \mathcal{A}_{ss}^{0,1} \cup \bigcup_{\lambda} \mathcal{A}_\lambda^{0,1}$$

by type has the following properties.

- (1)  $\overline{\mathcal{A}_\lambda^{0,1}} \subseteq \bigcup_{\mu \geq \lambda} \mathcal{A}_\mu^{0,1}$ .
- (2)  $H_{\mathcal{G}}^*(\mathcal{A}_\mu^{0,1}) \cong \bigoplus_{\ell=1}^k H_{\mathcal{G}_\ell}^*(\mathcal{A}_{ss}^{0,1}(E_\ell/E_{\ell-1}))$ , where  $E_1, \dots, E_\ell$  are the bundles in the Harder-Narasimhan filtration, and  $\mathcal{G}_\ell$  is the gauge group of  $E_\ell/E_{\ell-1}$  for  $\ell = 1, \dots, k$ .
- (3) Each  $\mathcal{A}_\lambda^{0,1}$  is a manifold with constant codimension  $n_\lambda$ , and there is an isomorphism

$$H_{\mathcal{G}}^* \left( \bigcup_{\mu \leq \lambda} \mathcal{A}_\mu^{0,1}, \bigcup_{\mu < \lambda} \mathcal{A}_\mu^{0,1} \right) \cong H_{\mathcal{G}}^{*-n_\lambda}(\mathcal{A}_\mu^{0,1}).$$

- (4) The stratification is *equivariantly perfect* in that, for each type  $\lambda$ , the long exact sequence in  $\mathcal{G}$ -equivariant cohomology

$$\cdots \rightarrow H_{\mathcal{G}}^* \left( \bigcup_{\mu \leq \lambda} \mathcal{A}_\mu^{0,1}, \bigcup_{\mu < \lambda} \mathcal{A}_\mu^{0,1} \right) \rightarrow H_{\mathcal{G}}^* \left( \bigcup_{\mu \leq \lambda} \mathcal{A}_\mu^{0,1} \right) \rightarrow H_{\mathcal{G}}^* \left( \bigcup_{\mu < \lambda} \mathcal{A}_\mu^{0,1} \right) \rightarrow \cdots$$

splits into short exact sequences.

As a consequence of the above properties, one can (a) inductively compute the equivariant Poincaré polynomial of  $\mathcal{A}_{ss}^{0,1}$  in terms of the equivariant Poincaré polynomials of all the lower rank spaces of semistable bundles, and (b) show that the inclusion  $\mathcal{A}_{ss}^{0,1} \hookrightarrow \mathcal{A}^{0,1}$  induces a surjective map

$$\kappa : H_{\mathcal{G}}^*(\mathcal{A}^{0,1}) \rightarrow H_{\mathcal{G}}^*(\mathcal{A}_{ss}^{0,1}).$$

This map is known as the *Kirwan map*. In [37], Kirwan shows that the strategy described above applies in much more generality: it extends to compact symplectic manifolds with a Hamiltonian action of a compact connected Lie group, and that one can (a) inductively compute the Poincaré polynomial of the symplectic quotient, and (b) show that the Kirwan map is surjective.

3.1.2. *Morse theory of the Yang-Mills functional.* Now suppose that  $E$  also has a Hermitian structure. In the notation of Section 2.2 this means a reduction of structure group from  $H^{\mathbb{C}} = \mathrm{GL}(n, \mathbb{C})$  to  $H = \mathrm{U}(n)$ . For simplicity of notation, in this section we omit the subscripts  $H^{\mathbb{C}}$  or  $H$ . Then there is an identification of  $\mathcal{A}^{0,1}(E)$  with the space  $\mathcal{A}(E)$  of connections compatible with this structure (note that in higher dimensions, the integrability condition requires that we restrict attention to the subset of connections  $d_A \in \mathcal{A}(E)$  whose curvature  $F_A$  satisfies  $F_A \in \Omega^{1,1}(\mathrm{ad} E)$ ). Therefore, there is an induced stratification

$$\mathcal{A} = \bigcup_{\lambda} \mathcal{A}_{\lambda} \tag{3.1}$$

by the Harder-Narasimhan type of the corresponding holomorphic structure. Even though a change in the Hermitian metric will change the identification  $\mathcal{A}^{0,1} \cong \mathcal{A}$ , for a fixed connection  $d_A \in \mathcal{A}$  the Harder-Narasimhan type of the associated holomorphic structure does not change when the metric changes. Therefore the stratification (3.1) is well-defined independently of the choice of metric.

The *Yang-Mills functional*,  $\mathrm{YM} : \mathcal{A} \rightarrow \mathbb{R}$ , is defined by

$$\mathrm{YM}(d_A) = \|F_A\|^2 = \int_X |F_A|^2 \mathrm{dvol},$$

and so we also have an induced functional  $\mathrm{YM} : \mathcal{A}^{0,1} \rightarrow \mathbb{R}$ , which depends on the identification  $\mathcal{A}^{0,1} \cong \mathcal{A}$ , and therefore on the choice of Hermitian metric.

The theorem of Narasimhan and Seshadri, i.e. the original Hitchin-Kobayashi correspondence mentioned in Section 2.2, (see [44] for the original proof and [22] for Donaldson's gauge-theoretic proof, which is more in the spirit of the Atiyah & Bott approach) shows that the space of polystable holomorphic structures are those that are  $\mathcal{G}^{\mathbb{C}}$ -equivalent to the minimum of YM. In fact, by considering the Yang-Mills flow, more structure is apparent: it is a consequence of the work of Daskalopoulos in [19] and Råde in [47] that the Yang-Mills flow defines a continuous  $\mathcal{G}$ -equivariant deformation retraction of  $\mathcal{A}_{\mathrm{ss}}$  onto the minimum of YM and that the limit is determined by the Seshadri filtration (see [39, Ch V, Thm 1.15]) of the initial condition.

This relationship between the algebraic geometry of the Harder-Narasimhan filtration and the analysis of the Yang-Mills flow also has an analog for unstable bundles. Atiyah and Bott show that the Yang-Mills functional achieves a minimum on each stratum  $\mathcal{A}_{\lambda}$ , and that this minimising set (call it  $\mathcal{C}_{\lambda}$ ) is precisely the set of critical points for YM that are contained in  $\mathcal{A}_{\lambda}$ . Moreover, the Morse index at each critical point is the same as the codimension of the corresponding Harder-Narasimhan stratum. The results of [19] and [47] then show that the Yang-Mills flow defines a  $\mathcal{G}$ -equivariant deformation retraction of  $\mathcal{A}_{\lambda}$  onto  $\mathcal{C}_{\lambda}$  and that the limit of the flow with initial condition  $d_{A_0}$  is isomorphic to the graded object of the Harder-Narasimhan-Seshadri double filtration of  $d_{A_0}$ .

Therefore, the problem of studying the  $\mathcal{G}$ -equivariant cohomology of the space of semistable holomorphic structures is the same as that of studying the  $\mathcal{G}$ -equivariant cohomology of the minimum of YM, and the inductive formula in terms of the cohomology of the Harder-Narasimhan strata now has a Morse-theoretic analog in terms of the equivariant cohomology of the critical sets for YM.

It is worth mentioning here the relationship with Palais's work. The analytic details of the Morse theory for the Yang-Mills functional are slightly different than the cases studied in [46]. In particular, since the critical sets are infinite-dimensional, then the Palais-Smale condition C fails for the Yang-Mills functional. To recover an analog of condition C, the solution is to instead look at the space of connections modulo the gauge group. Using Uhlenbeck's compactness theorem from [59], Daskalopoulos proves in [19, Proposition 4.1] that condition C does hold for YM on  $\mathcal{A}/\mathcal{G}$ , and so one should think of the Yang-Mills functional for a bundle over a compact Riemann surface as satisfying a  $\mathcal{G}$ -equivariant condition C on the space  $\mathcal{A}$ .

A modified version of condition C is also valid in higher dimensions, where bubbling needs to be taken into account. See for example [51] and [57, Proposition 4.5].

3.1.3. *Higgs bundles.* One would like to extend this picture to spaces of Higgs bundles. We first discuss the case of  $G$ -Higgs bundles for  $G = \mathrm{GL}(n, \mathbb{C})$ .

*Remark 3.1.* When  $G = \mathrm{GL}(n, \mathbb{C})$  we have  $\mathfrak{m}^{\mathbb{C}} = \mathfrak{gl}(n, \mathbb{C})$  (see Example 2.3) and so one can consider the vector bundle  $E$  associated to the original principal bundle and think of the space  $\mathcal{A}^{0,1}$  of holomorphic structures as an affine space modeled on  $\Omega^{0,1}(\mathrm{End}(E))$  and the Higgs field as taking values in  $\Omega^0(\mathrm{End}(E) \otimes K)$ . Therefore

$$E_H = \mathrm{ad}(E) \quad \text{and} \quad E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}}) = \mathrm{End}(E).$$

Moreover, if we choose the compact real form to be  $\mathfrak{u}(n) \subseteq \mathfrak{gl}(n, \mathbb{C})$ , then the involution  $\tau : \mathfrak{gl}(n, \mathbb{C}) \rightarrow \mathfrak{gl}(n, \mathbb{C})$  is given by  $\tau(u) = -u^*$ .

Atiyah and Bott's observation that the curvature  $F_A$  is a moment map for the action of  $\mathcal{G}$  on the space  $\mathcal{A} \cong \mathcal{A}^{0,1}$  also extends to the hyperkähler setting of Higgs bundles. The cotangent bundle  $T^*\mathcal{A}^{0,1} \cong \mathcal{A}^{0,1} \times \Omega^{1,0}(\mathrm{End}(E))$  has a *hyperkähler structure*, and the action of the gauge group  $\mathcal{G}$  has associated moment maps

$$\begin{aligned} \mu_1(\bar{\partial}_E, \varphi) &= F_A + [\varphi, \varphi^*] \\ \mu_{\mathbb{C}}(\bar{\partial}_E, \varphi) &= \mu_2 + i\mu_3 = 2i\bar{\partial}_E\varphi. \end{aligned}$$

The *hyperkähler moment map* combines these three moment maps using the imaginary quaternions

$$\mu_{hk}(\bar{\partial}_E, \varphi) = i\mu_1 + j\mu_2 + k\mu_3 \in \mathrm{Lie}(\mathcal{G})^* \otimes_{\mathbb{R}} \mathbb{R}^3.$$

In analogy with the Yang-Mills functional (which is the norm-square of the moment map), the *full Yang-Mills-Higgs functional* is defined to be the norm-square of the hyperkähler moment map

$$\begin{aligned} \mathrm{YMH} : \mathcal{A}^{0,1} \times \Omega^{1,0}(\mathrm{End}(E)) &\rightarrow \mathbb{R} \\ (\bar{\partial}_E, \varphi) &\mapsto \|\mu_1\|^2 + \|\mu_2\|^2 + \|\mu_3\|^2 = \|F_A + [\varphi, \varphi^*]\|^2 + 4\|\bar{\partial}_E\varphi\|^2. \end{aligned} \tag{3.2}$$

*Remark 3.2.* Recall from Remark 3.1 that  $\mathcal{A}^{0,1} \times \Omega^{1,0}(\mathrm{End}(E)) \cong \mathcal{A}(E_H) \times \Omega^{1,0}(E_{H^{\mathbb{C}}}(\mathfrak{m}^{\mathbb{C}}))$ , and so the domain of YMH defined above is consistent with that given in (2.18). Also, since the involution  $\tau$  is given by  $\tau(u) = -u^*$ , then we also have  $\|F_A + [\varphi, \varphi^*]\|^2 = \|F_A - [\varphi, \tau(\varphi)]\|^2$ .

The full Yang-Mills-Higgs functional can also be re-written in the following form (compare with [8, Proposition 2.1] for the case where  $\varphi \in \Omega^0(E)$ ).

**Lemma 3.3.**

$$\begin{aligned} \text{YMH}(\bar{\partial}_A, \varphi) &= \|F_A\|^2 + \|[\varphi, \varphi^*]\|^2 + 2\|\nabla_A \varphi\|^2 + 2\langle \varphi \circ \text{Ric}, \varphi \rangle \\ &\quad - 2\|\partial_A \varphi\|^2 - 2\|\partial_A^* \varphi\|^2 + 2\|\bar{\partial}_A \varphi\|^2. \end{aligned} \quad (3.3)$$

*Proof.* First we expand the Yang-Mills-Higgs functional as

$$\text{YMH}(\bar{\partial}_A, \varphi) = \|F_A\|^2 + \|[\varphi, \varphi^*]\|^2 + 2\Re \langle F_A, [\varphi, \varphi^*] \rangle + 4\|\bar{\partial}_A \varphi\|^2. \quad (3.4)$$

Next, recall the Weitzenbock identity for a connection  $d_A$  (with covariant derivative denoted by  $\nabla_A$ ) on a bundle  $F$  and a one-form  $\eta \in \Omega^1(F)$ . Let

$$\mathfrak{R}^A(\eta)_X := \sum_{j=1}^n [R^A(e_j, X), \eta(e_j)],$$

where  $R$  denotes the Riemannian curvature tensor on the base manifold  $M$  with respect to the Levi-Civita connection  $\nabla$ , and let  $\text{Ric}$  denote the Ricci tensor of  $M$ . Then (see [7, Theorem 3.1]) the Weitzenbock identity is

$$\Delta_A \eta = \nabla_A^* \nabla_A \eta + \eta \circ \text{Ric} + \mathfrak{R}^A(\eta). \quad (3.5)$$

The term coupling the curvature and the Higgs field in (3.4) can now be written as

$$\begin{aligned} 2\Re \langle F_A, [\varphi, \varphi^*] \rangle &= -2\langle \mathfrak{R}^A(\varphi), \varphi \rangle \\ &= 2\langle \nabla_A^* \nabla_A \varphi, \varphi \rangle + 2\langle \varphi \circ \text{Ric}, \varphi \rangle - 2\langle \Delta_A \varphi, \varphi \rangle \\ &= 2\|\nabla_A \varphi\|^2 + 2\langle \varphi \circ \text{Ric}, \varphi \rangle - 2\|d_A \varphi\|^2 - 2\|d_A^* \varphi\|^2, \end{aligned}$$

where in the second step we use the Weitzenbock identity (3.5). Since  $\varphi$  is a  $(1, 0)$  form, then  $d_A^* \varphi = \partial_A^* \varphi$ . Decomposition by type means that we also have  $\|d_A \varphi\|^2 = \|\bar{\partial}_A \varphi\|^2 + \|\partial_A \varphi\|^2$ . Substituting this into the formula for the Yang-Mills-Higgs functional gives us

$$\begin{aligned} \text{YMH}(\bar{\partial}_A, \varphi) &= \|F_A\|^2 + \|[\varphi, \varphi^*]\|^2 + 2\|\nabla_A \varphi\|^2 + 2\langle \varphi \circ \text{Ric}, \varphi \rangle \\ &\quad - 2\|\partial_A \varphi\|^2 - 2\|\partial_A^* \varphi\|^2 + 2\|\bar{\partial}_A \varphi\|^2, \end{aligned}$$

as required.  $\square$

Unfortunately, the full Yang-Mills-Higgs functional described above is not easy to study from the perspective of Morse theory. For example, a complete classification of the critical sets does not exist (although one can classify those satisfying  $\bar{\partial}_A \varphi = 0$ ), and currently there are no theorems on the long-time existence of the gradient flow (again, the difficulties occur when  $\bar{\partial}_A \varphi \neq 0$ ). From this perspective, it is simpler to restrict to the space  $\mathcal{B}(\text{GL}(n, \mathbb{C})) := \{(\bar{\partial}_E, \varphi) : \bar{\partial}_E \varphi = 0\}$  of Higgs bundles, although this introduces additional complications due to singularities in the space  $\mathcal{B}(\text{GL}(n, \mathbb{C}))$ . Fortunately these complications can be dealt with for cases of low rank; this will be explained later in the section.

On restriction to the space of Higgs bundles, the Yang-Mills-Higgs functional takes on the form

$$\begin{aligned} \text{YMH} : \mathcal{B}(\text{GL}(n, \mathbb{C})) &\rightarrow \mathbb{R} \\ (\bar{\partial}_E, \varphi) &\mapsto \|F_A + [\varphi, \varphi^*]\|^2. \end{aligned} \tag{3.6}$$

Some of the theory from holomorphic bundles does carry over to  $\text{GL}(n, \mathbb{C})$ -Higgs bundles. The Hitchin-Kobayashi correspondence shows that the polystable Higgs bundles are precisely those that are  $\mathcal{G}^{\mathbb{C}}$ -equivalent to a minimiser for the Yang-Mills-Higgs functional. There is also an analogous Harder-Narasimhan filtration (defined using  $\varphi$ -invariant holomorphic sub-bundles) and a corresponding algebraic stratification of the space of Higgs bundles.

The space of Higgs bundles, i.e.  $\mathcal{B}(\text{GL}(n, \mathbb{C}))$ , may be singular, so *a priori* it is not obvious how to extend the rest of the theory of Atiyah and Bott to study the topology of the space of semistable Higgs bundles. More seriously, the negative eigenspace of the Hessian does not have constant dimension on each connected component of the set of critical points of YMH. As explained in the next few sections, it is possible to get around these difficulties, and we have done this in [20], [61] and [62]. Some of the constructions are general, however some have been done by hand for specific cases, so, for now, we have restricted to certain low-rank situations.

The general strategy for extending the Atiyah and Bott theory to the space of Higgs bundles (outlined in more detail in the next few sections) is as follows. Firstly, the analog of the results of Daskalopoulos and Råde from [19] and [47] for the Yang-Mills flow also holds for the Yang-Mills-Higgs functional. Secondly, for certain low-rank cases, we are able to show that as the Yang-Mills-Higgs functional passes a critical value, then the topology of the space changes by attaching a certain topological space determined by the negative eigenspace of the Hessian at each critical point. Thirdly, we can compute the change in cohomology as the Yang-Mills-Higgs functional passes a critical value, and thus can compute the cohomology of the moduli space of Higgs bundles for certain low-rank cases.

Finally, it is worth mentioning that the above strategy applies to a much broader class of moduli spaces than just Higgs bundles. For example, the paper [61] uses this method to produce new information about the space of rank 2 stable pairs.

**3.2. Gradient Flow.** This section describes the results on the convergence of the gradient flow of the Yang-Mills-Higgs functional on the space of Higgs bundles.

Recall that there is already an algebraic stratification of the space of Higgs bundles. Each Higgs bundle  $(\bar{\partial}_A, \varphi)$  has a maximal semistable Higgs sub-bundle (a  $\varphi$ -invariant holomorphic sub-bundle, whose holomorphic structure and Higgs field are induced from  $(\bar{\partial}_A, \varphi)$ ), and from this one can construct a filtration of  $(\bar{\partial}_A, \varphi)$  analogous to the Harder-Narasimhan filtration for holomorphic bundles. The type  $\nu$  of the filtration is the vector consisting of the slopes of the quotient bundles counted with multiplicity, and there is a stratification

$$\mathcal{B} = \bigcup_{\nu} \mathcal{B}_{\nu}, \tag{3.7}$$

(where for simplicity we have dropped  $\mathrm{GL}(n, \mathbb{C})$  from the notation). The main theorem of [63] shows that the Yang-Mills-Higgs flow on  $\mathcal{B}$  respects the Harder-Narasimhan filtration in the following sense.

**Theorem 3.4.** *The Yang-Mills-Higgs flow with any initial condition  $(\bar{\partial}_A, \varphi)$  exists for all time and converges in the smooth topology to a unique limit point, which is a critical point of YMH. Moreover, this critical point is isomorphic to the graded object of the Harder-Narasimhan-Seshadri double filtration of  $(\bar{\partial}_A, \varphi)$ .*

Again, it is worth pointing out the connection with the Palais-Smale condition C. In this context, not only are the critical sets infinite-dimensional, but, even after dividing by the gauge group, they are non-compact. The necessary compactness ensured by condition C can, however, be recovered by restricting to complex gauge orbits. This allows one to prove an analog of the result in [19] for the Yang-Mills functional. Alternatively, as in [63], one can show that the gradient flow converges by applying the Moser iteration technique to obtain improved Sobolev estimates along the flow lines, from which Uhlenbeck's compactness theorem gives strong convergence in the appropriate norm.

Each Harder-Narasimhan stratum  $\mathcal{B}_\nu$  contains a critical set (corresponding to the minimum of YMH restricted to  $\mathcal{B}_\nu$ ), and the Harder-Narasimhan type  $\nu$  is determined by the value of  $F_A + [\varphi, \varphi^*]$  on this critical set. The above theorem implies that there is a Morse stratification

$$\mathcal{B} = \bigcup_{\nu} \mathcal{B}_\nu^{Morse} \quad (3.8)$$

where  $(\bar{\partial}_A, \varphi)$  is in the Morse stratum  $\mathcal{B}_\nu^{Morse}$  if and only if the limit of the flow with initial conditions  $(\bar{\partial}_A, \varphi)$  is in the critical set  $\mathcal{C}_\nu$  defined earlier.

The methods of [63] also show that the gradient flow deformation retraction of each Morse stratum onto the associated critical set is continuous. As a consequence, we see that

**Corollary 3.5.** (1)  $\mathcal{B}_\nu = \mathcal{B}_\nu^{Morse}$  for all types  $\nu$ , i.e. the Morse strata are the same as the Harder-Narasimhan strata, and  
 (2) the flow defines a continuous  $\mathcal{G}$ -equivariant deformation retraction of each stratum  $\mathcal{B}_\nu$  onto the corresponding critical set  $\mathcal{C}_\nu$ .

The paper [5] extends Theorem 3.4 to  $G$ -Higgs bundles whose structure group is any complex reductive group. The remainder of this section shows how to extend this result further to the case where  $G$  is a real reductive group as in Section 2. As in Sections 2.1 and 2.2, we fix a smooth principal  $H^{\mathbb{C}}$  bundle over  $X$ , denoted  $E_{H^{\mathbb{C}}}$ , with topological invariant  $d$  and let  $\mathcal{B}_d(G)$  denote the space of Higgs  $G$ -bundles on  $E_{H^{\mathbb{C}}}$ . Let  $E_{G^{\mathbb{C}}} = E_{H^{\mathbb{C}}} \times_{H^{\mathbb{C}}} G^{\mathbb{C}}$  be the bundle associate to  $E_{H^{\mathbb{C}}}$  by the inclusion  $H^{\mathbb{C}} \hookrightarrow G^{\mathbb{C}}$ . This induces an inclusion  $i_G : \mathcal{B}_d(G) \hookrightarrow \mathcal{B}_d(G^{\mathbb{C}})$ .

Given a reduction of structure group of  $E_{H^{\mathbb{C}}}$  from  $H^{\mathbb{C}}$  to  $H$ , we can define the Yang-Mills-Higgs functional on the space of Higgs  $G$ -bundles

$$\mathrm{YMH}_G(\bar{\partial}_A, \varphi) = \|F_A - [\varphi, \tau(\varphi)]\|^2. \quad (3.9)$$

where  $\tau$  is the conjugation on  $\mathfrak{g}_{\mathbb{C}}$  that determines the compact Lie algebra  $\mathfrak{h} \subset \mathfrak{g}$  (as in (2.5)).

**Lemma 3.6.** *The downwards gradient flow of  $\text{YMH}_G$  on the space  $\mathcal{B}_d(G)$  exists and is unique.*

*Proof.* First consider the following equation (cf. [53, Section 6] or [23, pp7-8]) on the space of metrics for the bundle  $E_{H^c}$ .

$$\frac{\partial h}{\partial t} = -2ih * (F_{A_0} + \bar{\partial}_{A_0}(h^{-1}(\partial_{A_0}h)) + [\varphi_0, h^{-1}\tau(\varphi_0)h]) + 2i\lambda h.$$

Simpson shows in [53, Section 6] that the solution exists and is unique. The same method as in [63, Section 3.1] (see also [36]) shows that this implies existence of a curve  $g : \mathbb{R}_{\geq 0} \rightarrow \mathcal{G}_H^c$  that generates the gradient flow of  $\text{YMH}_G$  in the sense that a solution  $(\bar{\partial}_{A(t)}, \varphi(t))$  satisfies

$$(\bar{\partial}_{A(t)}, \varphi(t)) = g(t) \cdot (\bar{\partial}_{A_0}, \varphi_0).$$

Since  $\mathcal{G}_H^c \subset \mathcal{G}_{G^c}$  preserves the space  $\mathcal{B}_d(G)$ , then this gives a unique integral curve for the downwards gradient vector field of  $\text{YMH}_G$  on the space  $\mathcal{B}_d(G)$ .  $\square$

Next we show that this integral curve coincides with that of the gradient flow of  $\text{YMH}_{G^c}$  with initial conditions in  $\mathcal{B}(G)$ . Recall that the Cartan involution  $\theta$  on  $\mathfrak{g}$  is the restriction of the involution  $\tau : \mathfrak{g}_{\mathbb{C}} \rightarrow \mathfrak{g}_{\mathbb{C}}$  that determines the compact real form  $\mathfrak{h} \oplus i\mathfrak{m} \subset \mathfrak{g}_{\mathbb{C}}$ . For the rest of this section we use  $\tau$  to denote both the involution on  $\mathfrak{g}_{\mathbb{C}}$  and its restriction to the Cartan involution on  $\mathfrak{g}$ .

Let  $B$  be the Killing form of  $\mathfrak{g}$ , and define

$$B_{\tau} : \mathfrak{g} \rightarrow \mathfrak{g}, \quad B_{\tau}(u, v) = -B(u, \tau(v)).$$

Since  $\tau$  is a Cartan involution then  $B_{\tau}$  is positive definite on the real vector space  $\mathfrak{g}$ , and therefore extends to a Hermitian inner product on  $\mathfrak{g}_{\mathbb{C}}$ . Note that the ad-invariance of the Killing form implies that

$$\begin{aligned} B_{\tau}([u, v], w) &= -B([u, v], \tau(w)) = -B(u, [v, \tau(w)]) \\ &= B_{\tau}(u, \tau[v, \tau(w)]) = B_{\tau}(u, [\tau(v), w]). \end{aligned} \quad (3.10)$$

Using this inner product, we can construct the following inner products on the spaces  $\Omega^2(\text{ad}(E_G)) \cong \text{Lie}(\mathcal{G})^*$  and  $\Omega^{0,1}(\text{ad}(E_{G^c})) \oplus \Omega^{1,0}(\text{ad}(E_{G^c})) \cong T_{(\bar{\partial}_A, \varphi)} T^* \mathcal{A}^{0,1}(E_{G^c})$ .

$$\begin{aligned} \langle \mu, \nu \rangle &:= \int_X B_{\tau}(\mu, * \nu) \\ \langle (a_1, \varphi_1), (a_2, \varphi_2) \rangle &:= \int_X B_{\tau}(a_1, * a_2) + \int_X B_{\tau}(\varphi_1, * \varphi_2), \end{aligned}$$

where (in local coordinates)  $\mathcal{B}_{\tau}$  acts on pairs of 1-forms by

$$\mathcal{B}_{\tau}(f_1 dz + g_1 d\bar{z}, f_2 dz + g_2 d\bar{z}) = -B(f_1, \tau(f_2)) dz d\bar{z} - B(g_1, \tau(g_2)) d\bar{z} dz,$$

and on a pair consisting of a 2-form and a 0-form by

$$\mathcal{B}_{\tau}(f dz d\bar{z}, g) = -B(f, \tau(g)) dz d\bar{z}.$$

Note also that for  $\varphi_1 = f_1 dz + g_1 d\bar{z}$  and  $\varphi_2 = f_2 dz + g_2 d\bar{z}$  we have

$$\begin{aligned}
\langle [\varphi_1, \varphi_2], \mu \rangle &= \int_X B_\tau([\varphi_1, \varphi_2], * \mu) \\
&= - \int_X B([\varphi_1, \varphi_2], \tau(* \mu)) \\
&= - \int_X B(\varphi_1, [\varphi_2, \tau(* \mu)]) \\
&= - \int_X B(\varphi_1, \tau[\tau(\varphi_2), * \mu]) \\
&= \int_X B_\tau(\varphi_1, [\tau(\varphi_2), * \mu]) \\
&= - \langle \varphi_1, *[\tau(\varphi_2), * \mu] \rangle.
\end{aligned}$$

The definitions above, together with the observation that  $\tau$  restricts to the Cartan involution on  $\mathfrak{g}$ , show that the functional  $\text{YMH}_{G^c}$  on  $\mathcal{B}_{G^c}$  restricts to  $\text{YMH}_G$  on  $\mathcal{B}_G$ .

Using the calculations above, we see that the derivative of  $\text{YMH}_G$  at a point  $(\bar{\partial}_A, \varphi) \in \mathcal{B}_G$  is given by

$$\begin{aligned}
d \text{YMH}_G(\delta a, \delta \varphi) &= 2\Re \langle d_A(\delta a) - [\varphi, \tau(\delta \varphi)] - [\delta \varphi, \tau(\varphi)], F_A - [\varphi, \tau(\varphi)] \rangle \\
&= 2\Re \langle \delta a, d_A^*(F_A - [\varphi, \tau(\varphi)]) \rangle + 2\Re \langle \delta \varphi, *[\varphi, *(F_A - [\varphi, \tau(\varphi)])] \rangle \\
&\quad + 2\Re \langle \tau(\delta \varphi), \tau(*[\varphi, *(F_A - [\varphi, \tau(\varphi)])]) \rangle \\
&= 4\Re \langle (\delta a)^{0,1}, -*\bar{\partial}_A*(F_A - [\varphi, \tau(\varphi)]) \rangle \\
&\quad + 4\Re \langle \delta \varphi, *[\varphi, *(F_A - [\varphi, \tau(\varphi)])] \rangle,
\end{aligned}$$

and so the gradient is

$$\text{grad } \text{YMH}_G = (-i\bar{\partial}_A*(F_A - [\varphi, \tau(\varphi)]), -i[\varphi, *(F_A - [\varphi, \tau(\varphi)])]),$$

which is the same as that of  $\text{YMH}_{G^c}$  at a point  $(\bar{\partial}_A, \varphi) \in \mathcal{B}_G$ .

Therefore we have proven the following

**Lemma 3.7.** *Let  $\text{YMH}_G$  denote the Yang-Mills-Higgs functional on  $\mathcal{B}_G$ , and let  $\text{YMH}_{G^c}$  denote the Yang-Mills-Higgs functional on  $\mathcal{B}_{G^c}$ . Then*

- (1)  $\text{YMH}_G$  is the restriction of  $\text{YMH}_{G^c}$  to  $i_G(\mathcal{B}_G) \subset \mathcal{B}_{G^c}$ , and
- (2) at a  $G$ -Higgs bundle  $(\bar{\partial}_A, \varphi) \in \mathcal{B}_G$ , the gradient of  $\text{YMH}_G$  is equal to the gradient of  $\text{YMH}_{G^c}$ .

Then the gradient flow line of  $\text{YMH}_G$  with initial conditions  $(\bar{\partial}_{A_0}, \varphi_0) \in \mathcal{B}_G$  from Lemma 3.6 is an integral curve of the gradient vector field  $\text{YMH}_{G^c}$  with initial conditions  $(\bar{\partial}_{A_0}, \varphi_0)$ . Since we have uniqueness for the gradient flow of both functionals, then the gradient flow line of  $\text{YMH}_G$  must coincide with the gradient flow line of  $\text{YMH}_{G^c}$  on the space  $\mathcal{B}_G$ . The inclusion  $\mathcal{B}_G \hookrightarrow \mathcal{B}_{G^c}$  is closed, and therefore the limit of the flow of  $\text{YMH}_{G^c}$  is also contained in  $\mathcal{B}_G$ . In summary, we have shown that

**Theorem 3.8.** *The gradient flow of  $\text{YMH}_G$  is the restriction of the gradient flow of  $\text{YMH}_{G^c}$  to the space of  $G$ -Higgs bundles. Moreover, the gradient flow of  $\text{YMH}_G$  converges to a critical point in  $\mathcal{B}_G$ .*



**3.3. Cohomology calculations.** First, recall the stratifications (3.7) and (3.8), which are the same by Corollary 3.5. If we choose a total ordering compatible with the partial ordering on Harder-Narasimhan type, then we can write this as

$$\mathcal{B} = \bigcup_{j=0}^{\infty} \mathcal{B}_j, \quad (3.11)$$

where  $\mathcal{B}_0$  is the semistable stratum, and define

$$X_d = \bigcup_{j=0}^d \mathcal{B}_j.$$

Then, since the total space  $\mathcal{B}$  is contractible, the problem of computing the cohomology of  $\mathcal{B}_0 = X_0$  reduces to understanding how the cohomology of  $X_d$  relates to that of  $X_{d-1}$  for all  $d$ . Therefore, the goal is to compute  $H_{\mathcal{G}}^*(X_d, X_{d-1})$  and understand the map  $b^k$  in the long exact sequence

$$\cdots \rightarrow H_{\mathcal{G}}^k(X_d, X_{d-1}) \xrightarrow{a^k} H_{\mathcal{G}}^k(X_d) \xrightarrow{b^k} H_{\mathcal{G}}^k(X_{d-1}) \rightarrow \cdots. \quad (3.12)$$

The first step is to show that  $H_{\mathcal{G}}^*(X_d, X_{d-1})$  can be computed in terms of local data in a neighbourhood of the critical set contained in  $\mathcal{B}_d = X_d \setminus X_{d-1}$ .

In addition, if the map  $a^k$  is injective for all  $k$  on each stratum  $\mathcal{B}_d$ , then the long exact sequence (3.12) splits into short exact sequences, and (with coefficients in  $\mathbb{C}$ ) we have

$$H_{\mathcal{G}}^k(X_d) \cong H_{\mathcal{G}}^k(X_{d-1}) \oplus H_{\mathcal{G}}^k(X_d, X_{d-1}).$$

In the situations studied by Atiyah & Bott in [3] and Kirwan in [37], the Atiyah-Bott lemma [3, Proposition 13.4] is the main tool in the proof that  $a^k$  is injective. When the ambient space is singular then the situation is more complicated; this is explained below.

Recall that in the Morse-Kirwan case, the dimension of the negative eigenspace of the Hessian is constant on each critical set, and the negative eigenspaces glue together to form a bundle over the critical set, which we call the *negative normal bundle*. First we define the analog of the negative normal bundle for Higgs bundles.

Given a smooth splitting  $E = E_1 \oplus \cdots \oplus E_n$ , where the sub-bundles are ordered by decreasing slope, consider the following sub-bundle of  $\text{End}(E)$

$$\text{End}(E)_{LT} := \bigoplus_{i < j} E_i^* \otimes E_j.$$

The critical point equations imply that the eigenvalues of  $*(F_A + [\varphi, \varphi^*]) \in \Omega^0(\text{End}(E))$  are locally constant and that the bundle  $E$  splits into  $\varphi$ -invariant sub-bundles corresponding to the eigenspaces. Therefore, at each critical point there is a well-defined subspace  $\text{End}(E)_{LT} \subset \text{End}(E)$ .

Consider the trivial bundle over the critical set  $\eta_d$  with fibres the vector spaces  $\Omega^{0,1}(\text{End}(E)) \oplus \Omega^{1,0}(\text{End}(E))$ , and let  $N$  denote the sub-bundle whose fibres are the subspaces  $\Omega^{0,1}(\text{End}(E)_{LT}) \oplus \Omega^{1,0}(\text{End}(E)_{LT})$ . Note that  $N$  is trivial over any subset of critical points that induce the same smooth splitting of  $E$ . We define the fibres of the negative normal bundle over a critical point  $(\partial_A, \varphi)$  to consist of those elements in the

fibres of  $N$  that are (a) orthogonal to the  $\mathcal{G}^{\mathbb{C}}$  orbit passing through  $(\bar{\partial}_A, \varphi)$ , and (b) contained in the space of Higgs bundles. More precisely, define the infinitesimal action of  $\mathcal{G}^{\mathbb{C}}$  through  $(\bar{\partial}_A, \varphi)$  by

$$\begin{aligned} \rho_{(\bar{\partial}_A, \varphi)}^{\mathbb{C}} : \Omega^0(\text{End}(E)) &\rightarrow \Omega^{0,1}(\text{End}(E)) \oplus \Omega^{1,0}(\text{End}(E)) \\ u &\mapsto (\bar{\partial}_A u, [\varphi, u]), \end{aligned}$$

and note that since  $(\bar{\partial}_A, \varphi)$  is split at a critical point, then  $\rho_{(\bar{\partial}_A, \varphi)}^{\mathbb{C}}$  restricts to a map  $\Omega^0(\text{End}(E)_{LT}) \rightarrow \Omega^{0,1}(\text{End}(E)_{LT}) \oplus \Omega^{1,0}(\text{End}(E)_{LT})$ . The *negative normal bundle* is

$$\nu_d^- := \left\{ (\bar{\partial}_A, \varphi, \delta a, \delta \varphi) \in N : (\rho_{(\bar{\partial}_A, \varphi)}^{\mathbb{C}})^*(\delta a, \delta \varphi) = 0, \bar{\partial}_{(A+a)}(\varphi + \delta \varphi) = 0 \right\}.$$

*Remark 3.9.* The term “negative normal bundle” is intended intuitively rather than literally, since  $\nu_d^-$  is the analog of the negative normal bundle for a Morse-Kirwan function on a smooth space. When the ambient space is singular then  $\nu_d^-$  as defined above may not be a bundle. As we will see below in Example 3.11, the fibres may not always be vector spaces, and even if the fibres are vector spaces then they may jump in dimension.

We also define the associated space

$$\nu'_d := \left\{ (\bar{\partial}_A, \varphi, \delta a, \delta \varphi) \in \nu_d^- : (\delta a, \delta \varphi) \neq 0 \right\}.$$

The following theorem from [20], [61] and [62] shows that  $H_{\mathcal{G}}^*(X_d, X_{d-1})$  can be computed in terms of the local data of the negative normal bundle. The analogous theorem for non-degenerate Morse theory is contained in [42, Chapter 3] and [46, Section 12], for Morse-Bott theory it was proven by Bott in [6, p250], and for Morse-Kirwan functions it was proven by Kirwan in [37, Chapter 10].

**Theorem 3.10.** *For the cases of rank 2 Higgs bundles,  $U(2, 1)$  Higgs bundles and rank 2 stable pairs, we have*

$$H_{\mathcal{G}}^*(X_d, X_{d-1}) \cong H_{\mathcal{G}}^*(\nu_d^-, \nu'_d).$$

As noted earlier, even if the fibres of  $\nu_d^-$  are vector spaces, their dimension may be non-constant on each connected component of the critical set of YMH, and so the problem of computing  $H_{\mathcal{G}}^*(\nu_d^-, \nu'_d)$  is more complicated than the methods of Atiyah & Bott and Kirwan from [3] and [37]. The following example illustrates how the fibres may jump in dimension.

*Example 3.11.* Let  $E$  be a rank 2 bundle of degree  $d$ , and consider the critical points of YMH where  $E$  splits into line bundles  $L_1 \oplus L_2$ , with  $\deg L_1 = \ell$ . This forms a connected component of the set of critical points, denoted  $\eta_{\ell}$ . Modulo the action of the unitary gauge group we can fix a smooth splitting  $E = L_1 \oplus L_2$  (cf. [3, Section 7] and [63, Corollary 4.17]), and with respect to this splitting  $\text{End}(E)_{LT} = L_1^* L_2$ . Then, for  $(\delta a, \delta \varphi) \in \Omega^{0,1}(\text{End}(E)_{LT}) \oplus \Omega^{1,0}(\text{End}(E)_{LT})$ , the conditions  $(\delta a, \delta \varphi) \in \ker(\rho_{(\bar{\partial}_A, \varphi)}^{\mathbb{C}})^*$  and  $\bar{\partial}_{A+a}(\varphi + \delta \varphi) = 0$  become

$$\bar{\partial}_A^*(\delta a) - \bar{*}[\varphi, \bar{*}(\delta \varphi)] = 0, \quad \bar{\partial}_A(\delta \varphi) + [\delta a, \varphi] = 0.$$

Note that  $[\delta a, \delta \varphi] = 0$ , since  $\text{End}(E)_{LT} = L_1^* L_2$ . Therefore the above equations are linear in  $(\delta a, \delta \varphi)$ , and so the space of solutions is a vector space.

When  $\varphi = 0$ , the equations reduce to

$$\bar{\partial}_A^*(\delta a) = 0, \quad \bar{\partial}_A(\delta\varphi) = 0,$$

and so the space of solutions is isomorphic to  $H^{0,1}(L_1^*L_2) \oplus H^{1,0}(L_1^*L_2)$ . By Riemann-Roch, the dimension is

$$2\ell - \deg(E) + g - 1 + h^0(L_1^*L_2 \otimes K).$$

When  $0 \leq \deg(L_1^*L_2 \otimes K) \leq 2g - 2$  then  $h^0(L_1^*L_2 \otimes K)$  is not constant with respect to the holomorphic structure on  $L_1^*L_2 \otimes K$ . Therefore the dimension of the fibres of  $\nu_d^-$  is non-constant.

When the Higgs field is non-zero, there is an analogous picture obtained by replacing the cohomology groups  $H^{0,1}(L_1^*L_2) \oplus H^{1,0}(L_1^*L_2)$  with the harmonic forms associated to the middle term in the following deformation complex

$$\Omega^0(L_1^*L_2) \xrightarrow{\rho_{(\bar{\partial}_A, \varphi)}^{\mathbb{C}}} \Omega^{0,1}(L_1^*L_2) \oplus \Omega^{1,0}(L_1^*L_2) \xrightarrow{d\mu_{\mathbb{C}}} \Omega^{1,1}(L_1^*L_2).$$

As noted above, for rank 2 we always have  $[\delta a, \delta\varphi] = 0$  for  $(\delta a, \delta\varphi) \in \Omega^{0,1}(\text{End}(E)_{LT}) \oplus \Omega^{1,0}(\text{End}(E)_{LT})$ , and so it is sufficient to study the linearised equations. For higher rank, one needs to replace the linearisation of the complex moment map  $d\mu_{\mathbb{C}}$  with the map  $(\delta a, \delta\varphi) \mapsto \bar{\partial}_{A+\delta a}(\varphi + \delta\varphi)$ . In the more complicated cases studied in [61] and [62] this becomes a consideration.

Nevertheless, it is still possible to compute  $H_{\mathcal{G}}^*(\nu_d^-, \nu_d')$  in this case (and for other low-rank cases). The observation that allows us to do this is that the jumps in dimension described above are related to the jumps in dimension of the fibres of the Abel-Jacobi map  $S^d X \rightarrow \text{Jac}_d(X)$ .

The strategy of the calculation from [20] is as follows. First, find a  $\mathcal{G}$ -equivariant deformation retract of the pair  $(\nu_d^-, \nu_d')$  to the subspace with Higgs field zero. Next, introduce the space

$$\nu_d'' = \{(\bar{\partial}_A, \varphi, \delta a, \delta\varphi) \in \nu_d^- : \varphi = 0, \delta a \neq 0\} \subseteq \nu_d'.$$

In [20], we compute

$$\begin{aligned} H_{\mathcal{G}}^*(\nu_d^-, \nu_d'') &\cong H^{*-2\ell+\deg(E)-(g-1)}(\text{Jac}(X) \times \text{Jac}(X)) \otimes H^*(BU(1))^{\otimes 2} \\ H_{\mathcal{G}}^*(\nu_d', \nu_d'') &\cong H^{*-2\ell+\deg(E)-(g-1)}(S^{2g-2+\deg(E)-2\ell_1} X \times \text{Jac}(X)) \otimes H^*(BU(1)). \end{aligned}$$

Without repeating the details of these calculations, it is still worth noting that both isomorphisms follow from reducing to the Thom isomorphism. In the first case this is straightforward, and in the second case this follows by reducing to a vector bundle over a subset of the negative normal directions.

These computations tell us about the long exact sequence (3.12) via the following commutative diagram.

$$\begin{array}{ccccccc}
 & & \vdots & & & & (3.13) \\
 & & \downarrow & & & & \\
 \cdots & \longrightarrow & H_{\mathcal{G}}^k(X_d, X_{d-1}) & \longrightarrow & H_{\mathcal{G}}^k(X_d) & \longrightarrow & H_{\mathcal{G}}^k(X_{d-1}) \longrightarrow \cdots \\
 & & \downarrow \cong & & \downarrow & & \downarrow \\
 \cdots & \longrightarrow & H_{\mathcal{G}}^k(\nu_d, \nu'_d) & \longrightarrow & H_{\mathcal{G}}^k(\nu_d) & \longrightarrow & H_{\mathcal{G}}^k(\nu'_d) \longrightarrow \cdots \\
 & & \downarrow & \nearrow \xi^k & & & \\
 & & H_{\mathcal{G}}^k(\nu_d, \nu''_d) & & & & \\
 & & \downarrow \gamma^k & & & & \\
 & & H_{\mathcal{G}}^k(\nu'_d, \nu''_d) & & & & \\
 & & \downarrow & & & & \\
 & & \vdots & & & & 
 \end{array}$$

**Lemma 3.12.** *For non-fixed determinant rank 2 Higgs bundles, the map  $\gamma^k$  is surjective for all  $k$ , and so the vertical long exact sequence splits into short exact sequences.*

The Atiyah-Bott lemma shows that the diagonal map  $\xi^k$  is injective, and so a diagram chase leads to

**Corollary 3.13.** (1) *The map  $b^k : H_{\mathcal{G}}^k(X_d) \rightarrow H_{\mathcal{G}}^k(X_{d-1})$  is surjective, and*  
(2)  $P_t^{\mathcal{G}}(X_d) - P_t^{\mathcal{G}}(X_{d-1}) = P_t^{\mathcal{G}}(\nu_d, \nu''_d) - P_t^{\mathcal{G}}(\nu'_d, \nu''_d)$ .

The calculations for other examples such as stable pairs and  $U(2, 1)$  Higgs bundles follow a similar strategy: the idea is to further decompose the negative normal bundle and then compute the cohomology of the pair  $(X_d, X_{d-1})$  in terms of on the smaller pieces in the decomposition. See [61] and [62] for more details.

In the fixed determinant case the situation is not quite as simple. The analogous map  $\gamma^k$  is not always surjective, and so the vertical long exact sequence in (3.13) does not split. As a consequence, a diagram chase of (3.13) only gives us

$$P_t^{\mathcal{G}}(\nu_d) - P_t^{\mathcal{G}}(\nu'_d) = P_t^{\mathcal{G}}(\nu_d, \nu''_d) - P_t^{\mathcal{G}}(\nu'_d, \nu''_d),$$

however we can still recover the second part of Corollary 3.13 in two different ways.

The first is to avoid using the negative normal bundle in the diagram (3.13), and instead do all of the calculations in terms of the spaces  $X_d$ . This is more in the spirit of the original Atiyah and Bott approach. The intermediate space is denoted  $X''_d$  (see [20, Section 3.1] for the precise definition), and there are isomorphisms

$$H_{\mathcal{G}}^*(X_d, X''_d) \cong H_{\mathcal{G}}^*(\nu_d, \nu''_d), \quad H_{\mathcal{G}}^*(X_{d-1}, X''_d) \cong H_{\mathcal{G}}^*(\nu'_d, \nu''_d).$$

The diagram becomes

$$\begin{array}{ccccccc}
 & & \vdots & & & & (3.14) \\
 & & \downarrow & & & & \\
 \cdots & \longrightarrow & H_{\mathcal{G}}^k(X_d, X_{d-1}) & \longrightarrow & H_{\mathcal{G}}^k(X_d) & \longrightarrow & H_{\mathcal{G}}^k(X_{d-1}) \longrightarrow \cdots \\
 & & \downarrow \cong & \nearrow \xi^k & & & \\
 & & H_{\mathcal{G}}^k(X_d, X_d'') & & & & \\
 & & \downarrow \gamma^k & & & & \\
 & & H_{\mathcal{G}}^k(X_{d-1}, X_d'') & & & & \\
 & & \downarrow & & & & \\
 & & \vdots & & & & 
 \end{array}$$

Again, the diagonal map  $\xi^k$  is injective. An analogous diagram chase to the one above gives us  $P_t^{\mathcal{G}}(X_d) - P_t^{\mathcal{G}}(X_{d-1}) = P_t^{\mathcal{G}}(X_d, X_d'') - P_t^{\mathcal{G}}(X_{d-1}, X_d'')$ , even though the vertical long exact sequence in (3.14) does not split.

The second approach is to use the action of the finite group  $\Gamma_2 := H^1(X, \mathbb{Z}/2\mathbb{Z})$  corresponding to the 2-torsion points on the Jacobian, which was originally defined by Hitchin in [34, Section 7]. This acts on the cohomology groups in the diagram (3.13), and the representation splits into pieces according to whether the action is trivial or not. In [20, Section 4.2], we show that the maps in the diagram (3.13) respect this splitting, and that we can recover the result

$$P_t^{\mathcal{G}}(X_d) - P_t^{\mathcal{G}}(X_{d-1}) = P_t^{\mathcal{G}}(\nu_d, \nu_d'') - P_t^{\mathcal{G}}(\nu_d', \nu_d'').$$

As an additional consequence of this method, we also prove

**Theorem 3.14** (Theorem 4.13 of [20]). *The  $\Gamma_2$ -invariant Kirwan map surjects onto the  $\Gamma_2$ -invariant part of the cohomology of  $\mathcal{B}_0^{ss}(2, d)$  for any  $d$ .*

#### 4. THE HITCHIN FUNCTION

We now examine the function mentioned in (2.19). Fix a smooth principal  $H$ -bundle,  $E_H$ , with topological invariant  $d$  and use the description of  $\mathcal{M}_d(G)$  in which points are represented by gauge orbits of points  $(A, \varphi) \in \mathcal{A} \times \Omega^{1,0}(E_H(\mathfrak{m}^{\mathbb{C}}))$ . The function is then

$$\begin{aligned}
 f : \mathcal{M}_d(G) &\rightarrow \mathbb{R} \\
 [A, \varphi] &\mapsto \|\varphi\|_{L^2}^2,
 \end{aligned} \tag{4.1}$$

where the  $L^2$ -norm of the Higgs field is computed using the (fixed) metric determined by the Ad-invariant bilinear form on  $G$  and the metric on  $X$ . If we regard points in  $\mathcal{M}_d(G)$  as isomorphism classes of holomorphic pairs  $(E_{H^{\mathbb{C}}}, \varphi)$  or, equivalently, as complex gauge orbits in  $\mathcal{A}^{0,1}(E_{H^{\mathbb{C}}}) \times \Omega^{1,0}(E_H(\mathfrak{m}^{\mathbb{C}}))$ , then the function is defined by

$$f([E_{H^c}, \varphi]) = \|g(\varphi)\|^2, \quad (4.2)$$

where  $g \in \mathcal{G}^{\mathbb{C}}$  is the element (unique up to real gauge transformations) such that  $g(E_{H^c}, \varphi)$  lies in the zero set of the moment map  $\Psi$ .

**Theorem 4.1.** *The map*

$$\begin{aligned} S^1 \times \mathcal{M}_d(G) &\rightarrow \mathcal{M}_d(G) \\ e^{i\theta}[E, \varphi] &= [E, e^{i\theta}\varphi] \end{aligned} \quad (4.3)$$

*defines an action which is Hamiltonian on the smooth locus and for which the above function  $f$  is the moment map.*

It follows immediately by a theorem of Frankel for  $S^1$ -actions on Kähler manifolds ([25]) that:

**Theorem 4.2.** *The critical points of  $f$  are given by the fixed points of the  $S^1$  action. Moreover,  $f$  is a perfect non-degenerate Morse-Bott function on the smooth locus of  $\mathcal{M}_d(G)$ .*

Furthermore, if  $M$  is any such Kähler manifold with an  $S^1$ -action and  $p \in M$  is a smooth critical point of the moment map, then there are two linear maps on the tangent space  $T_p M$ , namely

- (1) the Hessian of the moment map,  $H_f$ , and
- (2) the infinitesimal  $S^1$ -action,  $H_S$

and these are related by

$$H_f = -iH_S. \quad (4.4)$$

It follows from (4.4) that

**Proposition 4.3.** *At a critical point  $p \in M$  for the  $S^1$ -moment map  $f$ , the decomposition of  $T_p M$  into eigenspaces for  $H_f$  coincides with the weight-space decomposition for the  $S^1$ -action and that the  $\lambda$ -eigenspace for  $H_f$  is the  $-i\lambda$ -weight space for the  $S^1$ -action.*

In cases where  $\mathcal{M}_d(G)$  is smooth, it follows that the Poincare polynomial is given by

$$P_t(\mathcal{M}_d(G)) = \sum_N t^{\lambda_N} P_t(N) \quad (4.5)$$

where the sum is over the critical submanifolds of  $f$  and  $\lambda_N$  is the index of critical submanifold  $N$ , however in order to use formula (4.5) one has to

- identify the fixed points of the  $S^1$ -action and hence the critical submanifolds  $N$ ,
- compute the Morse indices  $\lambda_N$ , and
- compute the Poincare polynomials  $P_t(N)$ .

In Section 4.3 we discuss some cases where all these steps can be carried out. Even in cases where this is not possible, either because the moduli space is not smooth or because the requisite properties for the critical submanifolds are not known, the Hitchin function  $f$  retains the following useful features (first shown by Hitchin in [34]).

**Theorem 4.4.** *The function  $f : \mathcal{M}_d(G) \rightarrow \mathbb{R}$  defined as in (4.1) is a proper map and hence attains a minimum on each connected component.*

*Proof.* (sketch) Recall that points in  $\mathcal{M}_d(G)$  are represented by pairs  $(A, \varphi) \in \mathcal{A} \times \Omega^{1,0}(E_H(\mathfrak{m}^{\mathbb{C}}))$  which satisfy equations (2.5) and (2.6). Applying a Weitzenböck formula using the connection  $A$  and the metric on  $X$  thus leads to the estimate

$$0 \leq \|d_A^{1,0}\varphi\|_{L^2}^2 = -\langle F_A, [\varphi, \tau(\varphi)] \rangle + (2g-2)\|\varphi\|_{L^2}^2 \quad (4.6)$$

$$= -\|[\varphi, \tau(\varphi)]\|_{L^2}^2 + (2g-2)\|\varphi\|_{L^2}^2. \quad (4.7)$$

where  $d_A^{1,0}$  denotes the holomorphic part of the covariant derivative and the pairing  $\langle \cdot, \cdot \rangle$  uses the Ad-invariant bilinear form on  $\mathfrak{g}$  and the inner product on forms of type  $(1, 1)$ .

Thus

$$\|F_A\|_{L^2}^2 = \|[\varphi, \tau(\varphi)]\|_{L^2}^2 \leq (2g-2)f([A, \varphi]) \quad (4.8)$$

Since the connections are on a bundle over a compact Riemann surface, the result now follows from a compactness result by Uhlenbeck ([59]) for connections with  $L^2$ -bounds on curvature.  $\square$

In order to count the connected components of  $\mathcal{M}_d(G)$  it is thus enough to

- identify the fixed points of the  $S^1$ -action and hence the critical submanifolds  $N$ ,
- show that each minimal submanifold  $N$  is connected

In Section 4.4 we discuss some cases where these steps can be carried out.

**4.1. Critical points.** Denote the infinitesimal generator of the  $S^1$ -action (4.3) by the vector field  $X$  on  $\mathcal{M}_d(G)$ . Since  $f$  is a moment map with respect to the symplectic form defined by a Kähler metric, we have for any other vector field  $Y$

$$df(Y) = -k(Y, JX) \quad (4.9)$$

where  $J$  denotes the involution which defines the complex structure and the pairing  $k$  denotes the Kähler metric. In particular, a point  $[A, \varphi] \in \mathcal{M}_d(G)$  is a critical point of  $f$  if and only if  $X_{[A, \varphi]} = 0$ , i.e.

$$\frac{d}{dt}[A, e^{it}\varphi]|_{t=0} = 0 \quad (4.10)$$

There are two types of critical points that can occur:

**Type 1:**  $[\varphi = 0]$  Clearly any points at which  $\varphi = 0$  are fixed points of the  $S^1$ -action and hence critical points of  $f$ . Moreover, since  $f = 0$  at these points, they are in fact global minima.

**Type 2:**  $[\varphi \neq 0]$  Minima with  $\varphi \neq 0$  can occur because the points in  $\mathcal{M}_d(G)$  are *isomorphism classes* of Higgs bundles or, equivalently *gauge orbits* of pairs in the

configuration space. The condition for a point represented by the Higgs bundle  $(A, \varphi)$  to be a fixed point is that there is a 1-parameter family of gauge transformations, say  $g_\theta$ , such that

$$(A, e^{i\theta}\varphi) = g_\theta(A, \varphi) \quad (4.11)$$

for all  $\theta$ . Let  $\Psi$  be the infinitesimal action of  $g_\theta$ , i.e.  $(d^A(\Psi), [\Psi, \varphi]) = \frac{d}{d\theta}\big|_{\theta=0} g_\theta(A, \varphi)$ . Then (4.11) implies

$$\begin{aligned} d^A(\Psi) &= 0, \\ [\Psi, \varphi] &= i\varphi. \end{aligned} \quad (4.12)$$

*Example 4.5.* If  $G = \mathrm{GL}(n, \mathbb{C})$  and we use the standard representation to replace  $E_{H^{\mathbb{C}}}$  by a rank  $n$  vector bundle  $V$  then the first condition in (4.12) implies that  $\Psi$  decomposes  $V$  into a direct sum of eigenbundles, say

$$V = \bigoplus_{\lambda} V_{\lambda}, \quad (4.13)$$

where the eigenvalue on  $V_{\lambda}$  is  $i\lambda$ . The second condition then says that

$$\varphi : V_{\lambda} \rightarrow V_{\lambda+1} \otimes K \quad (4.14)$$

In other words, the Higgs bundle  $(V, \varphi)$  defines a variation of Hodge structure (as defined, for example, in [54]).

In general, the infinitesimal gauge transformation  $\Psi$  defines eigenbundle decompositions

$$E_H(\mathfrak{h}^{\mathbb{C}}) = \bigoplus_{\mu} (E_H(\mathfrak{h}^{\mathbb{C}}))_{\mu} \quad (4.15)$$

$$E_H(\mathfrak{m}^{\mathbb{C}}) = \bigoplus_{\nu} (E_H(\mathfrak{m}^{\mathbb{C}}))_{\nu} \quad (4.16)$$

where the eigenvalues on the summands are  $i\mu$  and  $i\nu$  respectively. The second condition in (4.12) implies that

$$ad(\varphi) : (E_H(\mathfrak{h}^{\mathbb{C}}))_{\mu} \rightarrow (E_H(\mathfrak{m}^{\mathbb{C}}))_{\mu+1} \otimes K \quad (4.17)$$

$$ad(\varphi) : (E_H(\mathfrak{m}^{\mathbb{C}}))_{\nu} \rightarrow (E_H(\mathfrak{h}^{\mathbb{C}}))_{\nu+1} \otimes K \quad (4.18)$$

where the maps are defined by the adjoint action. In cases where there is a ‘natural’ or ‘standard’ representation of  $H^{\mathbb{C}}$  so that  $E_{H^{\mathbb{C}}}$  can be replaced by one or more vector bundles, the fixed points of the  $S^1$ -action have a Hodge bundle structure and the eigenbundles of  $E_H(\mathfrak{h}^{\mathbb{C}})$  and  $E_H(\mathfrak{m}^{\mathbb{C}})$  can be described in terms of the Hodge bundle summands.



*Example 4.6.* For  $G = \mathrm{GL}(n, \mathbb{C})$  where, as in Example 4.5, we can describe the Higgs bundles as pairs  $(V, \varphi)$  in which  $V$  is a rank  $n$  vector bundle, the Hodge bundles are as in (4.13) and (4.17). In this case  $\mathfrak{h}^{\mathbb{C}} = \mathfrak{m}^{\mathbb{C}} = \mathfrak{gl}(n, \mathbb{C})$  and  $E_H(\mathfrak{h}^{\mathbb{C}}) = E_H(\mathfrak{m}^{\mathbb{C}}) = \mathrm{End}(V)$ . The eigenbundles in the decomposition of  $\mathrm{End}(V)$  are given by

$$\mathrm{End}(V)_\lambda = \bigoplus_{\lambda=\nu-\mu} \mathrm{Hom}(V_\mu, V_\nu) \quad (4.19)$$

i.e. the eigenvalues are of the form  $\mu - \nu$  where  $\mu, \nu$  are eigenvalues in the Hodge decomposition of  $V$ .

*Example 4.7.* <sup>4</sup>For  $G = \mathrm{Sp}(2n, \mathbb{R})$  we have, as in Table 1,  $H = \mathrm{U}(n)$ , so  $E_{H^c}$  is a principal  $\mathrm{GL}(n, \mathbb{C})$ -bundle. If we let  $V$  be the rank  $n$  vector bundle given by the standard representation, then

$$\begin{aligned} E_H(\mathfrak{h}^{\mathbb{C}}) &= \mathrm{End}(V) \\ E_H(\mathfrak{m}^{\mathbb{C}}) &= \mathrm{Sym}^2(V) \oplus \mathrm{Sym}^2(V^*) \end{aligned} \quad (4.20)$$

so, as in Table 2, we can describe  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles as triples  $(V, \beta, \gamma)$  where

$$\beta \in H^0(\mathrm{Sym}^2(V) \otimes K) \quad \text{and} \quad \gamma \in H^0(\mathrm{Sym}^2(V^*) \otimes K). \quad (4.21)$$

At the fixed points of the  $S^1$ -action the infinitesimal generator of the gauge transformation, i.e. the  $\Psi$  in (4.12), yields an eigenbundle decomposition as in (4.13) but the effect of the second condition in (4.12) is not the same as in (4.14). For  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles the condition is equivalent to the conditions

$$\Psi\beta - \beta\Psi^* = i\beta \quad (4.22)$$

$$\Psi^*\gamma - \gamma\Psi = i\gamma \quad (4.23)$$

and hence we get that

$$\beta : V_\lambda^* \rightarrow V_{\lambda+1} \otimes K \quad \text{and} \quad \gamma : V_\lambda \rightarrow V_{-\lambda-1}^* \otimes K \quad (4.24)$$

where  $V = \bigoplus_\lambda V_\lambda$  is the eigenbundle decomposition of  $V$ . Thus:

*Definition 4.8.* The Hodge bundles corresponding to the  $S^1$ -fixed points on the  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundle moduli spaces are of the form  $(V, \beta, \gamma)$  with  $V = \bigoplus_\lambda V_\lambda$  and  $\beta, \gamma$  satisfying (4.24).

It follows from (4.20) that the eigenbundle decompositions for  $E_H(\mathfrak{h}^{\mathbb{C}})$  and  $E_H(\mathfrak{m}^{\mathbb{C}})$  are given by

---

<sup>4</sup>See [26] for details

$$\mathrm{End}(V)_\lambda = \bigoplus_{\lambda=\mu-\nu} \mathrm{Hom}(V_\mu, V_\nu) \quad (4.25)$$

$$\mathrm{Sym}^2(V)_{\lambda+1} = \bigoplus_{\mu+\nu=\lambda+1} V_\mu \otimes V_\nu \oplus \mathrm{Sym}^2(V_{\frac{\lambda+1}{2}}) \quad (4.26)$$

$$\mathrm{Sym}^2(V^*)_{\lambda+1} = \bigoplus_{-\mu-\nu=\lambda+1} V_\mu^* \otimes V_\nu^* \oplus \mathrm{Sym}^2(V_{\frac{\lambda+1}{2}}^*) \quad (4.27)$$

**4.2. Morse indices and a criterion for the minima.** Having identified the critical points of the function  $f$ , the next step in the Morse theory program requires computation of Morse indices, and this in turn requires an understanding of the tangent spaces to the Higgs bundle moduli spaces. The main ideas go back to Hitchin's papers (see [35]):

The space of infinitesimal deformations of a  $G$ -Higgs bundle  $(E_H, \varphi)$  can be identified as the first hypercohomology of a deformation complex

$$C^\bullet(E_H, \varphi) : E_H(\mathfrak{h}^\mathbb{C}) \xrightarrow{\mathrm{ad}(\varphi)} E_H(\mathfrak{m}^\mathbb{C}) \otimes K \quad (4.28)$$

If  $(E_H, \varphi)$  represents a smooth point in the moduli space then the first hypercohomology of the deformation complex is canonically isomorphic to the tangent space at the point, i.e.

$$\mathbb{H}^1(C^\bullet(E_H, \varphi)) \simeq T_{[E_H, \varphi]} \mathcal{M}_d(G) . \quad (4.29)$$

The decompositions (4.15), together with (4.17), decompose the deformation complex as

$$C^\bullet(E_H, \varphi) = \bigoplus_{\mu} C^\bullet(E_H, \varphi)_\mu \quad (4.30)$$

where

$$C^\bullet(E_H, \varphi)_\mu : E_H(\mathfrak{h}^\mathbb{C})_\mu \xrightarrow{\mathrm{ad}(\varphi)} E_H(\mathfrak{m}^\mathbb{C})_{\mu+1} \otimes K \quad (4.31)$$

At smooth critical points of  $f$  the hypercohomology group  $\mathbb{H}^1(C^\bullet(E_H, \varphi)_\mu)$  is isomorphic to the  $-\mu$  eigenspace for the Hessian  $H_f$ .

The above results allow Morse indices to be calculated and in particular yield the following criterion for a smooth critical point to be a minimum (see [35]):

**Proposition 4.9.** *Let  $(E_{H^c}, \varphi)$  represent a fixed point of the  $S^1$ -action on the smooth locus of a moduli space of polystable  $G$ -Higgs bundles  $\mathcal{M}_d(G)$ . Then the point is a local minimum if and only if*

$$\mathbb{H}^1(C^\bullet(E_H, \varphi)_\mu) = 0 \quad \forall k > 0 . \quad (4.32)$$

In practice, i.e. when working out the details for specific examples, the vanishing of  $\mathbb{H}^1(C^\bullet(E_H, \varphi)_\mu)$  is determined using a criterion first formulated in [35], namely

**Proposition 4.10.** *Under the same assumptions as in Proposition 4.9, the hypercohomology  $\mathbb{H}^1(C^\bullet(E_H, \varphi)_\mu)$  vanishes if and only if*

$$ad(\varphi) : (E_H(\mathfrak{h}^{\mathbb{C}}))_\mu \rightarrow (E_H(\mathfrak{m}^{\mathbb{C}}))_{\mu+1} \otimes K \quad (4.33)$$

*is an isomorphism.*

In cases where the moduli space  $\mathcal{M}_d(G)$  has singularities, it can happen that fixed points of the  $S^1$ -action of the function  $f$  lie in the singular locus. To determine whether the resulting Hodge bundles are local minima for the function  $f$ , one must construct families of deformations by hand and examine the behavior of the function along the deformations. In Section 4.4 we illustrate these ideas with some specific examples.

**4.3. Best-case scenarios.** In this section we discuss some situations in which the full Morse theory program can be carried out using the Hitchin function  $f$ .

As noted in Section 2.2, for  $G = \mathrm{GL}(n, \mathbb{C})$ , the moduli spaces  $\mathcal{M}_d(G)$  are sometimes smooth. The  $\mathrm{GL}(n, \mathbb{C})$ -Higgs bundles can be represented by pairs  $(V, \varphi)$  where  $V$  is a rank  $n$  holomorphic bundle, the Higgs field  $\varphi : V \rightarrow V \otimes K$ , and the topological class  $d$  is the degree of  $V$ . In order for such Higgs bundles to be polystable but not stable, the bundle  $V$  has to split as a direct sum of lower rank vector bundles with the same slope as  $V$ . This cannot happen if the rank and degree of  $V$  are coprime, i.e.  $(n, d) = 1$ . In [34] Hitchin analyzed the case of  $n = 2$  and odd degree and in [32] Gothen examined the rank three case.

Fix  $d$  such that  $(3, d) = 1$  and let  $\mathcal{M}_d$  denote the moduli space of stable  $\mathrm{GL}(3, \mathbb{C})$ -Higgs bundles for which the underlying bundle has degree  $d$ . Let  $f : \mathcal{M}_d \rightarrow \mathbb{R}$  be the Hitchin function as defined in (4.1). The critical points of  $f$ , or equivalently the fixed points of the  $S^1$ -action, can be of either type discussed in Section 4.1.

The critical points with  $\varphi = 0$  are zeros of the non-negative function  $f$  and hence clearly global minima. Moreover, if  $\mathcal{N}_0$  denotes the locus of such minima, then we can identify

$$\mathcal{N}_0 \simeq M(3, d) \quad (4.34)$$

where  $M(3, d)$  denotes the moduli space of stable bundles of rank 3 and degree  $d$ . The Morse index of  $\mathcal{N}_0$  is zero (since the critical points are global minima) and the Poincaré polynomial for  $M(3, d)$  is known by the work of Desale and Ramanan ([21]). We can thus compute the contribution of  $\mathcal{N}_0$  to the Poincaré polynomial of  $\mathcal{M}_d$ .

The critical points with  $\varphi \neq 0$  are described as follows. As explained in Example 4.5 the corresponding fixed points of the  $S^1$ -action are represented by Hodge bundles, say  $(V, \varphi)$ . Since  $n = 3$ , the possibilities are limited to the three types in Table 3. The possibilities within this set of types are constrained by the topological type of  $V$ , i.e.

$$\deg(L) + \deg(V_i) = \deg(L_1) + \deg(L_2) + \deg(L_3) = d. \quad (4.35)$$

Moreover, the Hodge bundles must represent points in the moduli space  $\mathcal{M}_d$ . They must thus represent *stable* Higgs bundles or, equivalently, must support solutions to the Higgs bundle equation (2.5). This puts further constraints on the degrees of the

type	$V$	ranks	$\varphi$
(1, 2)	$L \oplus V_2$	$\text{rank}(L) = 1, \text{rank}(V_2) = 2$	$\begin{pmatrix} 0 & 0 \\ \varphi & 0 \end{pmatrix}$
(2, 1)	$V_1 \oplus L$	$\text{rank}(V_1) = 2, \text{rank}(L) = 1$	$\begin{pmatrix} 0 & 0 \\ \varphi & 0 \end{pmatrix}$
(1, 1, 1)	$L_1 \oplus L_2 \oplus L_3$	$\text{rank}(L_i) = 1$	$\begin{pmatrix} 0 & 0 & 0 \\ \varphi_1 & 0 & 0 \\ 0 & \varphi_2 & 0 \end{pmatrix}$

TABLE 3. Hodge bundles defining critical points of the  $S^1$ -action on the moduli space of  $\text{SL}(3, \mathbb{C})$ -Higgs bundles

summands. For example (see [32]) for the Hodge bundles of type (1, 2) the constraint is that

$$3d < \deg(L) < \frac{1}{3}d + g - 1. \quad (4.36)$$

Having identified the forms of all possible critical points, what remains is to

- (1) describe the critical submanifolds and compute their Poincare polynomials, and
- (2) compute the Morse indices for each of the critical submanifolds

*Remark 4.11.* This is the place where the cases  $n \geq 3$  become much more difficult than the cases  $n = 2$  and  $n = 3$ . In particular, the main obstacle to progress is the description of the critical submanifolds and the computation of their Poincare polynomials.

We describe briefly how Gothen completed the Morse theory program in the case  $n = 3$ , giving only enough details to highlight some of the geometry. For full details we refer the reader to [32]. The main idea is that the critical submanifolds can be identified as moduli spaces in their own right and for which the Poincare polynomials can be computed.

Consider the Hodge bundles of type (1, 2). The component  $\varphi$  in the Higgs field is a holomorphic map

$$\varphi : L \rightarrow VK \quad (4.37)$$

or, equivalently, an element of the space of holomorphic sections  $H^0(VL^{-1}K)$ . Denoting  $\deg L$  by  $l$ , we see that these Hodge bundles are determined by pairs  $(\mathcal{V}, \varphi)$  where  $\mathcal{V}$  is a rank three bundle of degree  $d - l - (2g - 2)$  and  $\varphi \in H^0(\mathcal{V})$ . Just as for Higgs bundles, moduli spaces for pairs of this sort may be constructed if one introduces a suitable notion of stability. The appropriate definition depends on a real parameter (see [9]) and the resulting moduli spaces are non-empty provided the parameter lies in a bounded range determined by the rank and degree of the bundle. For rank 2 bundles the stability of a pair  $(\mathcal{V}, \varphi)$  can be formulated as a condition on line subbundles  $\mathcal{L} \subset \mathcal{V}$ , namely

$$\deg(\mathcal{L}) < \begin{cases} \frac{v}{2} - \sigma & \text{if } \Phi \in H^0(\mathcal{L}) \\ \frac{v}{2} + \sigma & \text{otherwise} \end{cases} \quad (4.38)$$

where  $v = \deg(\mathcal{V})$  and  $\sigma$  is a real parameter. If the inequalities are allowed to be equalities then the pair is  $\sigma$ -semistable. The allowed range for  $\sigma$  is

$$0 \leq \sigma \leq \frac{v}{2}. \quad (4.39)$$

This range is partitioned at the values of  $\sigma$  for which  $\frac{v}{2} \pm \sigma$  is integral<sup>5</sup>. Except for values of  $\sigma$  at these partition points, the moduli spaces of  $\sigma$ -stable pairs, say  $N_\sigma(2, v)$ , are smooth projective varieties. In fact the moduli spaces depend only on the subinterval containing  $\sigma$ , so there is a discrete family of distinct moduli spaces labelled by the subintervals  $(\frac{v}{2} - i - 1, \frac{v}{2} - i)$  (for  $i = 0, 1, \dots, i_{max}$ ).

It is not hard to show (see [32]) that the Hodge bundle  $(L \oplus V, \varphi)$  is stable if and only if the corresponding pair  $(\mathcal{V}, \varphi)$  is  $\sigma$ -stable for a specific value of  $\sigma$ , namely

$$\sigma = -\frac{d}{6} + \frac{l}{2} \quad (4.40)$$

We thus get a precise relationship between the critical points of type (1, 2) and a moduli space of  $\sigma$ -stable pairs induced by the map

$$(L \oplus V, \varphi) \mapsto (VL^{-1}K, \varphi) \quad (4.41)$$

It remains to determine the Poincare polynomial for the moduli space  $N_\sigma(2, v)$  where  $v = d - l + 2g - 2$  and  $\sigma$  is given by (4.40). This can be done by adapting the work of Thaddeus ([58]). The moduli space corresponding to the largest subinterval in the range for  $\sigma$  is the easiest to describe and the others can be related to it by ‘flips’, i.e. explicit transformations whose effect on the Poincare polynomial can be computed.

The Hodge bundles of type (2, 1) can be analysed in essentially the same way, while those of type (1, 1, 1) can be described as finite coverings of products of Jacobian varieties. The Morse indices at each of the critical points are calculated using their relation to the weight spaces for the  $S^1$  action (see Proposition 4.3).

Putting everything together, Gothen computes the full Poincare polynomial for  $\mathcal{M}_d$  for any  $d$  coprime to 3 (see [32], Theorem 1.2).

*Remark 4.12.* For  $GL(n, \mathbb{C})$ -Higgs bundles with  $n > 3$ , if the degree of the bundle is coprime to  $n$  then the moduli space  $\mathcal{M}_d(GL(n, \mathbb{C}))$  is smooth, the Morse indices for the critical submanifolds of the Hitchin function can be computed, and the submanifolds can be described as moduli spaces in their own right. They correspond to moduli spaces of quiver bundles where the quivers describe the structure of the possible Hodge bundles that occur at fixed points of the  $S^1$ -action. Unfortunately, the Poincare polynomials of these spaces are not known.

---

<sup>5</sup>Thus at integer or half-integer values, depending on whether  $\deg V$  is even or odd

**4.4. Good enough for  $\pi_0$ .** The success of the full Morse theory program for moduli spaces of  $\mathrm{GL}(3, \mathbb{C})$ -Higgs bundles depends on the smoothness of the moduli spaces and the fact that the critical submanifolds can be described and have computable Poincare polynomials. In this section we discuss examples of  $G$ -Higgs bundles that fall short of these requirements but in which we can at least compute the first topological invariant, namely the number of connected components. In these cases  $G$  is a non-compact real reductive group, so the  $G$ -Higgs bundles are as described in Section 2. Many specific examples have been examined (see [1],[10],[11],[26],[28],[29],[31],[34, 35])

We discuss just one case, namely  $G = \mathrm{Sp}(2n, \mathbb{R})$ . We pick this illustrative example because  $\mathrm{Sp}(2n, \mathbb{R})$  is both a split real form and also has the property that the associated homogeneous space  $G/H$  is a Hermitian symmetric space. Moreover, since  $\mathrm{Sp}(2n, \mathbb{R})$  is semisimple, we can invoke the identification of (2.17) to identify the Higgs bundle moduli spaces with moduli spaces of surface group representations in  $\mathrm{Sp}(2n, \mathbb{R})$ .

For  $G = \mathrm{Sp}(2n, \mathbb{R})$  we have seen in Example 4.7 that an  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundle on  $X$  consists of a triple  $(V, \beta, \gamma)$  where  $V$  is a rank  $n$  holomorphic bundle on  $X$  and the components of the Higgs field are symmetric maps

$$\beta : V^* \rightarrow V \otimes K \tag{4.42}$$

$$\gamma : V \rightarrow V^* \otimes K \tag{4.43}$$

The topological invariant is  $d = \deg(V)$ . The definitions of stability and polystability for  $\mathrm{Sp}(2n, \mathbb{R})$ , given in full in [26], amount to an inequality on the degrees of subbundles in 2-step filtrations of  $V$  that are compatible with the Higgs field. For polystable  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles the invariant  $d$  has to satisfy the bound:

$$0 \leq |d| \leq n(g-1) \tag{4.44}$$

For each value of  $d = \deg(V)$  in the range (4.44), we get a moduli space of polystable  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles, denoted by  $\mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$ . These spaces are complex algebraic varieties of dimension  $(g-1)(2n^2+n)$  but in general they have singularities (at decomposable  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles). This is the first obstacle to full implementation of the Morse theory program. We can nevertheless describe the  $\mathrm{Sp}(2n, \mathbb{R})$ -Hodge bundles, i.e. the critical points of the  $S^1$ -action on  $\mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$  given by (4.3), and identify the minima.

The  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundle equation becomes

$$F_A + \beta\beta^* - \gamma^*\gamma = 0 \tag{4.45}$$

where  $A$  is a connection on  $V$  and the adjoints on  $\beta, \gamma$  are taken with respect to a fixed metric on  $V$ . The Hitchin function is given in this situation by

$$f(V, \beta, \gamma) = \|\beta\|^2 + \|\gamma\|^2 \tag{4.46}$$

where the  $L^2$  norms are computed with respect to the fixed metric on  $V$ . Combining (4.45) and (4.46) yields

$$f(V, \beta, \gamma) = -d + 2\|\beta\|^2 = d + 2\|\gamma\|^2 \quad (4.47)$$

Define the loci  $\mathcal{N}_d \subset \mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$  by

$$\mathcal{N}_d = \{(V, \beta, \gamma) \in \mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R})) \mid \beta = 0 \text{ if } d \leq 0, \text{ or } \gamma = 0 \text{ if } d \geq 0\} \quad (4.48)$$

It follows from (4.47) that  $\mathcal{N}_d$  is a subset of  $f_{\min}$ , the locus of global minima of  $f$ . It does not follow - and in fact is not always true - that all global minima lie in  $\mathcal{N}_d$ .

The Hodge bundles are as in Definition 4.8, with further constraints imposed by stability on the ranks and degrees (see Corollary 6.6 in [26]). Using the criterion described in Section 4.2, we thus get the following characterization of the minima.

**Proposition 4.13.** *Let  $(V, \beta, \gamma)$  be a fixed point of the  $S^1$  action which lies in the smooth locus of  $\mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$ .*

- (1) *If  $|d| < n(g-1)$  then  $(V, \beta, \gamma) \in f_{\min}$  if and only if  $(V, \beta, \gamma) \in \mathcal{N}_d$ .*
- (2) *If  $d = -n(g-1)$  then  $(V, \beta, \gamma) \in f_{\min}$  if and only if  $(V, \beta, \gamma) \in \mathcal{N}_d$  or it is a Hodge bundle with*

$$V = \begin{cases} \bigoplus_{j=-q}^q L^{-1}K^{-2j} & \text{if } n = 2q + 1 \\ \bigoplus_{j=-q}^{q+1} LK^{-2j} & \text{if } n = 2q + 2 \neq 2 \end{cases} \quad (4.49)$$

where  $L$  is a line bundle such that  $L^2 = K$  and, with respect to this decomposition for  $V$  and the corresponding decomposition of  $V^*$ ,

$$\beta = \begin{pmatrix} 0 & \dots & 0 & 1 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 1 & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} \quad \text{and} \quad \gamma = \begin{pmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \ddots & \vdots \\ 0 & 1 & \dots & 0 \end{pmatrix} \quad (4.50)$$

- (3) *If  $d = n(g-1)$  the result is the same as for  $d = -n(g-1)$  but with  $(V, \beta, \gamma)$  replaced by  $(V^*, \gamma^t, \beta^t)$ .*

*Remark 4.14.* In the case  $n = 2$  and  $|d| = 2g - 2$ , i.e. for maximal  $\mathrm{Sp}(4, \mathbb{R})$ -Higgs bundles, there are additional possibilities for the Hodge bundles, as described below in Section 4.4.3. In the case  $n = 1$  we have  $\mathrm{Sp}(2, \mathbb{R}) \simeq \mathrm{SL}(2, \mathbb{R})$ , a case analyzed by Hitchin in [34].

This result is not sufficient to identify all minima because fixed points of the  $S^1$ -action can occur at singular points. One has to examine ‘by hand’ whether such critical points are minima of the Hitchin functional. In the case of  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles, generalizing Hitchin’s results in [34] for  $\mathrm{SL}(2, \mathbb{R})$ , one can explicitly construct paths in  $\mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$  through such Hodge bundles and along which the Hitchin function is decreasing. It follows that the minima all lie in the smooth locus and hence are all of the form given in Proposition 4.13.

Having characterized the minima of the Hitchin function on the spaces  $\mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$ , we can now - in principle - use Theorem 4.4 to investigate the number of connected components.

In the case  $|d| < n(g-1)$ , Proposition 4.13 says that we need to examine only the space  $\mathcal{N}_d$ . The objects in this space are so-called quadratic pairs  $(V, q)$  where  $q$  is a (possibly degenerate) quadratic form on  $V$  or  $V^*$  with values in the line bundle  $K$ . A good notion of stability can be formulated for such objects and moduli spaces can be constructed (see [26]). At present the only case for which the connectedness of these spaces is known is the case of rank two quadratic pairs, in which case they are connected (see [26] and [33]) and we get

**Proposition 4.15.** *The moduli spaces  $\mathcal{M}_d(\mathrm{Sp}(4, \mathbb{R}))$  are connected for all  $|d| < 2g - 2$*

*Remark 4.16.* It is reasonable to conjecture that for all  $n$  and all  $0 \leq |d| < n(g-1)$  the moduli spaces  $\mathcal{M}_d(\mathrm{Sp}(2n, \mathbb{R}))$  are connected.

The situation is better for extremal values of  $d$ , i.e. in the cases where  $d = 0$  or  $|d| = n(g-1)$ . In the case  $d = 0$ , the critical points all have  $\beta = \gamma = 0$  and we can identify  $f_{\min} = \mathcal{N}_0$  with  $M(n, 0)$ , the moduli space of polystable rank  $n$  vector bundles. Since  $M(n, 0)$  is known to be connected, so therefore is  $\mathcal{M}_0(\mathrm{Sp}(2n, \mathbb{R}))$ .

In the case  $|d| = n(g-1)$  we see two new phenomena:

- (1) Multiple components. The two cases in parts (2) and (3) of Proposition 4.13 already show that the moduli spaces  $\mathcal{M}_{\pm n(g-1)}(\mathrm{Sp}(2n, \mathbb{R}))$  have multiple components. In fact, there are at least as many components of the second type as there are choices for the line bundle  $L$ , i.e. for a square root of  $K$ . We discuss this in Section 4.4.1
- (2) Teichmuller components. In the case of  $\mathrm{Sp}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})$ , it is well known (see [30] and also [34]) that some<sup>6</sup> of the components of  $\mathcal{M}_{g-1}(\mathrm{Sp}(2, \mathbb{R}))$  can be identified with Teichmuller space. For  $n \geq 2$  a subset of the extra maximal components share key properties similar to Teichmuller space and are referred to as ‘higher Teichmuller’ or Hitchin components. We discuss this in Section 4.4.2

4.4.1. *Cayley transform and new invariants.* A distinguishing feature of polystable  $\mathrm{Sp}(2n, \mathbb{R})$ -Higgs bundles for which  $|d| = n(g-1)$  is that one of the two parts of the Higgs field has to have maximal rank as a bundle endomorphism (see Proposition 3.22 of [26]). For example, if  $(V, \beta, \gamma)$  is polystable with  $\deg(V) = n(g-1)$  then  $\gamma : V \rightarrow V^* \otimes K$  is an isomorphism.<sup>7</sup> This is manifestly true for the Hodge bundles described in parts (2) and (3) of Proposition 4.13 in which both  $\beta$  and  $\gamma$  are non-trivial but applies equally to other Hodge bundles or the Higgs bundles which are not critical points. Notice that for all such Higgs bundles, say with positive maximal  $d$ , the map  $\gamma$  is a symmetric non-degenerate bundle map. After twisting by a line bundle  $K^{-1/2}$  (i.e. a square root of  $K^{-1}$ ), we thus get a bundle  $W = VK^{-1/2}$  with a symmetric, non-degenerate map

$$q = \gamma \otimes 1_{K^{-1/2}} : W \rightarrow W^* \tag{4.51}$$

The pair  $(W, q)$  defines an orthogonal bundle, i.e. a bundle with structure group  $\mathrm{O}(n, \mathbb{C})$ . As a result of this emergence of a new structure group new characteristic classes, namely

<sup>6</sup>For a surface of genus  $g$  the number is  $2^{2g}$ , to be precise.

<sup>7</sup>If  $d = -n(g-1)$  then  $\beta : V^* \rightarrow V \otimes K$  must be an isomorphism.



the Stiefel-Whitney classes for the orthogonal bundle, separate  $\mathcal{M}_{n(g-1)}(\mathrm{Sp}(2n, \mathbb{R}))$  into components with fixed values of the new classes. For orthogonal bundles over a Riemann surface  $X$  there are two Stiefel-Whitney classes with values in  $H^1(X, \mathbb{Z}_2)$  and  $H^2(X, \mathbb{Z}_2)$  respectively, and thus a total of  $2 \cdot 2^{2g}$  possibilities. Notice, furthermore that in each of these components the Higgs bundles in the minimal submanifold  $\mathcal{N}_{n(g-1)}$  have  $\beta = 0$  and are thus fully determined by the data defining the orthogonal bundles. This means that  $\mathcal{N}_{n(g-1)}$  can be identified with the moduli space of polystable principal  $\mathrm{O}(2n, \mathbb{C})$ -bundles and hence, by the results of [49], is connected. In addition, there are the components on which the minima are single points represented by the Hodge bundles of the type described in parts (2) and (3) of Proposition 4.13. Since there are  $2^{2g}$  choices for the line bundle  $L$  (i.e. for the square root of  $K$ ), we get finally:

**Theorem 4.17** ([26], Theorem 7.14). *On a genus  $g$  closed Riemann surface,  $\mathcal{M}_{\pm n(g-1)}$  has  $3 \cdot 2^{2g}$  connected components if  $n \geq 3$ .*

Much of the above discussion reflects general features seen in those cases where the group  $G$  has the property that the associated symmetric space  $G/H$  is of Hermitian type.<sup>8</sup> In that context the invariant  $d$  is known as the **Toledo invariant** and the bound (4.44) is called the **Milnor-Wood bound**. If  $G$  is of Hermitian type then the Higgs fields for  $G$ -Higgs bundles with maximal Toledo invariant specialize in a way that results in the appearance of a new structure group called the **Cayley partner** to  $G$ . This, in turn, identifies new invariants which distinguish multiple components of  $\mathcal{M}_{max}(G)$ , the moduli space of polystable  $G$ -Higgs bundles with maximal Toledo invariant.

Under the identification (2.17), the  $G$ -Higgs bundles with maximal Toledo invariant correspond to special surface group representations in  $G$ . Such maximal representations have been intensively studied by a variety of methods and are known to have special dynamical and geometric properties (see [15, 16, 17] for surveys). While many features are seen most easily from the point of view of surface group representations, to-date the Morse theory methods described above provide the most effective tools for counting components of the moduli spaces.

**4.4.2. Hitchin components.** The Hodge bundles of parts (2) and (3) in Proposition 4.13 illustrate a phenomenon that can be traced back to the fact that the group  $\mathrm{Sp}(2n, \mathbb{R})$  is a *split real form* of a complex reductive group. In the case of the split real form  $\mathrm{PSL}(2, \mathbb{R})$ , a hyperbolic structure on  $X$  determines a (Fuchsian) representation of  $\pi_1(X)$  in  $\mathrm{PSL}(2, \mathbb{R})$  and this leads to the identification of a component of  $\mathrm{Rep}(\pi_1(X), \mathrm{PSL}(2, \mathbb{R}))$  with the Teichmüller space for the surfaces of genus  $g$  (see [30]). In [34] Hitchin showed how to identify the components in  $\mathcal{M}_{g-1}(\mathrm{PSL}(2, \mathbb{R}))$  which correspond to these representations. The minima of the Hitchin function on these components can be identified as the  $n = 1$  case of the Hodge bundles of part (2) in Proposition 4.13. In the case of other split real forms  $G$  (such as  $\mathrm{SL}(n, \mathbb{R})$ ,  $\mathrm{Sp}(2n, \mathbb{R})$ ,  $\mathrm{SO}(n, n + 1)$ ,  $\mathrm{SO}(n, n)$ ), the Lie algebra contains principal three dimensional subgroups which define homomorphisms of  $\mathrm{PSL}(2, \mathbb{R})$  into  $G$ . Composing this embedding with the Fuchsian representations leads to so-called higher Teichmüller components in  $\mathrm{Rep}(\pi_1(X), G)$  or correspondingly in the the moduli space of  $G$ -Higgs bundles. In the case of  $\mathrm{Sp}(2n, \mathbb{R})$ ,

<sup>8</sup>see[12] for further examples

these are precisely the components for which the minima of the Hitchin function are as in part (2) in Proposition 4.13.

*Remark 4.18.* The Hitchin components can also be identified by means of another map defined by Hitchin. Unlike the function define in Section 4, this is a map

$$h : \mathcal{M}(G) \rightarrow \bigoplus_i H^0(K^{p_i}) \quad (4.52)$$

from the moduli space  $\mathcal{M}(G)$  to a linear space given by  $\bigoplus_i H^0(K^{p_i})$  where the powers  $p_i$  are the degrees of Ad-invariant homogeneous polynomials which generate the ring of all such polynomials on the summand  $\mathfrak{m}^{\mathbb{C}} \subset \mathfrak{g}$ . Here, as usual,  $K$  is the canonical bundle on  $X$  and  $H^0$  denotes the space of holomorphic sections. In [35] Hitchin showed that when  $G$  is a split real form, this map admits sections. The image of these sections are precisely the Hitchin components.

4.4.3. *The case of  $\mathrm{Sp}(4, \mathbb{R})$ .* We end with a brief account of the special case of  $G = \mathrm{Sp}(4, \mathbb{R})$  (see [13], [28], [31] for more details). In this case, the Higgs bundles are defined by triples  $(V, \beta, \gamma)$  where  $V$  is a rank 2 bundle of degree  $d$  with  $0 \leq |d| \leq 2g - 2$ . For the Higgs bundles with  $|d| = 2g - 2$ , say  $d = (2g - 2)$  for definiteness, the bundle in the Cayley partner is an orthogonal bundle of rank two, i.e. it has structure group  $\mathrm{O}(2, \mathbb{C})$  and thus topological type determined by Stiefel-Whitney classes  $w_1 \in H^1(X, \mathbb{Z}_2)$  and  $w_2 \in H^2(X, \mathbb{Z}_2)$ . By a classification result of Mumford ([43]) for such bundles:

**Proposition 4.19.** *The Higgs bundle  $(V, \beta, \gamma)$  in  $\mathcal{M}_{2g-2}(\mathrm{Sp}(4, \mathbb{R}))$  can be taken to be one of the following types:*

- (1)  $V = L \oplus L^{-1}K$ , where  $L$  is a line bundle on  $X$ , and with respect to this decomposition,

$$\gamma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in H^0(S^2V^* \otimes K) \quad \text{and} \quad \beta = \begin{pmatrix} \beta_1 & \beta_3 \\ \beta_3 & \beta_2 \end{pmatrix} \in H^0(S^2V \otimes K). \quad (4.53)$$

*In this case the first Steifel-Whitney classes of the Cayley partner vanishes. If  $(V, \beta, \gamma)$  is a polystable Higgs bundle then  $\deg(L)$  must lie in the range  $g - 1 \leq \deg(L) \leq 3g - 3$  and  $\beta_2 \neq 0$  if  $l > g - 1$ .*

- (2)  $V = \pi_*(\tilde{L} \otimes \iota^*\tilde{L}^{-1})K^{1/2}$  where  $\pi : \tilde{X} \rightarrow X$  is a connected double cover,  $\tilde{L}$  is a line bundle on  $\tilde{X}$ , and  $\iota : \tilde{X} \rightarrow \tilde{X}$  is the covering involution. In this case the first Steifel-Whitney classes of the Cayley partner,  $w_1 \in H^1(X; \mathbb{Z}/2)$ , is the non-zero element defining the double cover.
- (3)  $(V, \beta, \gamma) = (V_1, \beta_1, \gamma_1) \oplus (V_2, \beta_2, \gamma_2)$  where  $(V_i, \beta_i, \gamma_i)$  are maximal  $\mathrm{Sp}(2, \mathbb{R})$ -Higgs bundles. In this case the first Steifel-Whitney classes of the Cayley partner is the sum of the first Steifel-Whitney classes for the Cayley partners to  $(V_i, \beta_i, \gamma_i)$ .

As in Section 4.4.1, the Stiefel-Whitney classes of the Cayley partner label separate components of  $\mathcal{M}_{2g-2}(\mathrm{Sp}(4, \mathbb{R}))$ . Moreover, in case (1) of Proposition 4.19, i.e. when  $w_1 = 0$ , the degree of the line bundle  $L$  constitutes an additional invariant<sup>9</sup> We thus get a decomposition of  $\mathcal{M}_{2g-2}(\mathrm{Sp}(4, \mathbb{R}))$  into:

<sup>9</sup>This can also be realized as an integral lift of the second Steifel-Whitney class  $w_2 \in H^2(X, \mathbb{Z}_2)$ .

$$\mathcal{M}_{2g-2}(\mathrm{Sp}(4, \mathbb{R})) = \left( \bigcup_{w_1 \neq 0, w_2} \mathcal{M}_{w_1, w_2} \right) \cup \left( \bigcup_{0 \leq l \leq 2g-2} \mathcal{M}_l^0 \right) \quad (4.54)$$

where  $\mathcal{M}_{w_1, w_2}$  denotes the component in which the Higgs bundles are represented by  $(V, \beta, \gamma)$  of type (2) with Stiefel-Whitney classes  $w_1, w_2$ , and  $\mathcal{M}_l^0$  denotes the component in which the Higgs bundles are represented by  $(V, \beta, \gamma)$  of type (1) in Proposition 4.19 with  $\deg(L) = l$ . Notice that when  $l = 3g - 3$ , the component  $\beta_2$  in (4.53) must be a non-zero section of  $L^{-2}K^3$ , which means that

$$L^2 = K^3 \quad (4.55)$$

and  $\beta_2$  can be taken to be the identity section of the trivial line bundle. The component  $\mathcal{M}_{3g-3}^0$  thus splits further into components determined by the  $2^{2g}$  choices of  $L$  satisfying (4.55). Denoting these by  $K^{3/2}$ , we get

$$\mathcal{M}_{3g-3}^0 = \bigcup_{K^{3/2}} \mathcal{M}_{K^{3/2}} \quad (4.56)$$

The components identified in (4.54) and (4.56) are not a priori connected, but we can use the Hitchin function (and Theorem 4.4) to investigate their connectedness. The minima of the Hitchin function on each of these components can be identified as follows:

- (1) For  $g - 1 < l \leq 3g - 3$ ,  $(V, \beta, \gamma)$  represents a minimum of the Hitchin function on  $\mathcal{M}_l^0$  if and only if it is isomorphic to a Higgs bundle of the form in (1) of Proposition 4.19 with  $\beta_1 = \beta_2 = 0$ ,
- (2) On all other components, i.e.  $\mathcal{M}_{g-1}^0$  or  $\mathcal{M}_{w_1, w_2}$ ,  $(V, \beta, \gamma)$  represents a minimum of the Hitchin function if and only if  $\beta = 0$ .

Notice that the minima on  $\mathcal{M}_{K^{1/2}}$  are precisely the Hodge bundles identified in Proposition 4.13, while the minima on the components  $\mathcal{M}_\nu^0$  are additional Hodge bundles with  $\beta \neq 0$ . These are the exceptional Hodge bundles referred to in Remark 4.14. The components  $\mathcal{M}_{K^{1/2}}$  are the Hitchin components for  $\mathrm{Sp}(4, \mathbb{R})$  (see Section 4.4.2) while the other components  $\mathcal{M}_l^0$  have no analogs in  $\mathcal{M}(\mathrm{Sp}(2n, \mathbb{R}))$  for  $n \neq 2$ .

The connectedness of the minimal submanifolds, and hence of the components, is proved in [31]. It is interesting to note that in many cases, the minimal submanifolds can be shown to be connected by explicitly describing the entire loci: see [13] for descriptions of  $\mathcal{M}_l^0$  for  $g - 1 < l < 2g - 2$  and the Hitchin components  $\mathcal{M}_{K^{3/2}}$ . We thus arrive at a final count of connected components:

**Theorem 4.20.** *The moduli space  $\mathcal{M}_{2g-2}(\mathrm{Sp}(4, \mathbb{R}))$  has*

$$2(2^{2g} - 1) + (2g - 2) + 2^{2g} = 3 \cdot 2^{2g} + 2g - 4 .$$

*connected components.*

## REFERENCES

- [1] M. Arroyo, The Geometry of  $SO(p, q)$ -Higgs Bundles, Ph.D. Thesis, (2009) Salamanca
- [2] M. F. Atiyah and R. Bott. Yang-Mills and bundles over algebraic curves. *Proc. Indian Acad. Sci. Math. Sci.*, 90(1):11–20, (1981).
- [3] M. F. Atiyah and R. Bott. The Yang-Mills equations over Riemann surfaces. *Philos. Trans. Roy. Soc. London Ser. A*, 308(1505):523–615,1(1983).
- [4] M. F. Atiyah, N. J. Hitchin, and I. M. Singer. Self-duality in four-dimensional Riemannian geometry. *Proc. Roy. Soc. London Ser. A*, 362(1711):425–461, (1978).
- [5] I. Biswas and G. Wilkin. Morse theory for the space of Higgs  $G$ -bundles. *Geom. Dedicata*, 149:189–203, (2010).
- [6] R. Bott. Nondegenerate critical manifolds. *Ann. of Math. (2)*, 60:248–261, (1954).
- [7] Jean-Pierre Bourguignon and H. Blaine Lawson, Jr. Stability and isolation phenomena for Yang-Mills fields. *Comm. Math. Phys.*, 79(2):189–230, (1981).
- [8] S. B. Bradlow, Vortices in holomorphic line bundles over closed Kähler manifolds, *Comm. Math. Phys.* 135(1):1–17, (1990).
- [9] S. B. Bradlow, Special metrics and stability for holomorphic bundles with global sections, *J. Differential Geom.* 33 (1991), 169213.
- [10] S. Bradlow, O. Garcia-Prada, and P. Gothen, Surface group representations and  $U(p, q)$ -Higgs bundles, *J. Diff. Geom.* 64 (2003), 111–170
- [11] S. Bradlow, O. Garcia-Prada, and P. Gothen, Representations of surface groups in the general linear group, *Publications of the RSME*, vol. 7 (2004), 89–94.
- [12] S. Bradlow, O. Garcia-Prada, and P. Gothen, Maximal surface group representations in isometry groups of classical Hermitian symmetric spaces, *Geometriae Dedicata* 122, (2006), 185–213.
- [13] S. Bradlow, O. Garcia-Prada, and P. Gothen, Deformations of maximal representations in  $Sp_{94}(\mathbb{R})$ , *Quarterly J. Math.* (2011), to appear.
- [14] S. B. Bradlow, O. García-Prada, and I. Mundet i Riera, Relative Hitchin-Kobayashi correspondences for principal pairs, *Quarterly J. Math.* **54** (2003), 171–208.
- [15] M. Burger, A. Iozzi, F. Labourie and A. Wienhard, Maximal representations of surface groups: Symplectic Anosov structures *Pure and Applied Mathematics Quarterly*, 1 (2005), no. 2, (Special Issue: In Memory of Armand Borel), 555–601.
- [16] M. Burger, A. Iozzi, and A. Wienhard, Surface group representations with maximal Toledo invariant, *Ann. of Math. (2)* 172 (2010), no. 1, 517–566.
- [17] M. Burger, A. Iozzi, and A. Wienhard, Higher Teichmüller spaces: From  $SL(2, \mathbb{R})$  to other Lie groups, *Handbook of Teichmüller Theory vol. IV* (2011).
- [18] Kevin Corlette Flat  $G$ -bundles with canonical metrics *J. Differential Geom.*, 28 (1988), 361–382
- [19] G. D. Daskalopoulos. The topology of the space of stable bundles on a compact Riemann surface. *J. Differential Geom.*, 36(3):699–746, 1992.
- [20] G. D. Daskalopoulos, R. A. Wentworth, J. Weitsman, and G. Wilkin, Morse theory and hyperkähler Kirwan surjectivity for Higgs bundles. *J. Differential Geom.*, 87(1):81–116, 2011.
- [21] U. V. Desale and S. Ramanan, Poincare polynomials of the variety of stable bundles, *Math. Ann.* 216 (1975), 233–244.
- [22] S. K. Donaldson. A new proof of a theorem of Narasimhan and Seshadri. *J. Differential Geom.*, 18(2):269–277, 1983.
- [23] S. K. Donaldson. Anti self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles. *Proc. London Math. Soc. (3)*, 50(1):1–26, 1985.
- [24] S. K. Donaldson. Twisted harmonic maps and the self-duality equations. *Proc. London Math. Soc. (3)*, 55(1):127–131, 1987.
- [25] T. Frankel, Fixed points and torsion on Kähler manifolds, *Ann. of Math. (2)* **70** (1959), 1–8.
- [26] O. García-Prada, P. B. Gothen, and I. Mundet i Riera, *Representations of surface groups in the real symplectic group*, preprint (2009), [arXiv:0809.0576v3](https://arxiv.org/abs/0809.0576v3) [math.AG].
- [27] O. García-Prada, P. B. Gothen, and I. Mundet i Riera, *The Hitchin–Kobayashi correspondence, Higgs pairs and surface group representations*, preprint (2009), [arXiv:0909.4487v2](https://arxiv.org/abs/0909.4487v2) [math.AG].

- [28] O. García-Prada and I. Mundet i Riera, *Representations of the fundamental group of a closed oriented surface in  $\mathrm{Sp}(4, \mathbb{R})$* , *Topology* **43** (2004), 831–855.
- [29] O. Garcia-Prada and A. Oliveira, Higgs bundles for the non-compact dual of the unitary group, *Illinois Journal of Mathematics*, in press
- [30] W. Goldman, Topological components of spaces of representations, *Invent. Math.*, 93, (1988), 557–607
- [31] P. B. Gothen, Components of spaces of representations and stable triples, *Topology* 40 (2001), 823–850.
- [32] P. Gothen, The Betti Numbers of the Moduli Space of Rank 3 Higgs Bundles, *International Journal of Mathematics* 5 (1994), 861–875.
- [33] P. Gothen and A. Oliveira, Rank two quadratic pairs and surface group representations, (2011) *arXiv:1106.1766*
- [34] N. J. Hitchin. The self-duality equations on a Riemann surface. *Proc. London Math. Soc. (3)*, 55(1):59–126, 1987.
- [35] N. J. Hitchin. Lie groups and Teichmüller space *Topology* **31** (1992), 449–473.
- [36] Min-Chun Hong. Heat flow for the Yang-Mills-Higgs field and the Hermitian Yang-Mills-Higgs metric. *Ann. Global Anal. Geom.*, 20(1):23–46, 2001.
- [37] F. C. Kirwan. *Cohomology of quotients in symplectic and algebraic geometry*, volume 31 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 1984.
- [38] A. W. Knap, *Lie Groups beyond an Introduction*, first ed., Progress in Mathematics, vol 140, Birkhäuser Boston Inc., Boston, MA, 1996.
- [39] S. Kobayashi. *Differential geometry of complex vector bundles*, volume 15 of *Publications of the Mathematical Society of Japan*. Princeton University Press, Princeton, NJ, 1987. , Kano Memorial Lectures, 5.
- [40] M. Lubke, Stability of Einstein-Hermitian vector bundles., *Manuscripta math.* 42, (1983), 245–257.
- [41] M. Lubke and A. Teleman, The universal Kobayashi-Hitchin correspondence on Hermitian manifolds. *Mem. Amer. Math. Soc.*, 183 (2006), no. 863
- [42] J. Milnor. *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J. (1963).
- [43] D. Mumford, *Theta characteristics of an algebraic curve*, *Ann. Sci. École Norm. Sup. (4)* **4** (1971), 181–192.
- [44] M. S. Narasimhan and C. S. Seshadri. Stable and unitary vector bundles on a compact Riemann surface. *Ann. of Math. (2)*, 82:540–567, (1965).
- [45] A. Newlander and L. Nirenberg. Complex analytic coordinates in almost complex manifolds. *Ann. of Math. (2)*, 65:391–404, (1957).
- [46] R. S. Palais. Morse theory on Hilbert manifolds. *Topology*, 2:299–340, (1963).
- [47] Johan Råde. On the Yang-Mills heat equation in two and three dimensions. *J. Reine Angew. Math.*, 431:123–163, (1992).
- [48] A. Ramanathan, *Stable principal bundles on a compact Riemann surface*, *Math. Ann.* **213** (1975), 129–152.
- [49] ———, *Moduli for principal bundles over algebraic curves: I and II*, *Proc. Indian Acad. Sci. Math. Sci.* **106** (1996), 301–328 and 421–449.
- [50] A. Schmitt, *Geometric invariant theory and decorated principal bundles*, Zurich Lectures in Advanced Mathematics, European Mathematical Society, (2008).
- [51] S. Sedlacek. A direct method for minimizing the Yang-Mills functional over 4-manifolds. *Comm. Math. Phys.*, 86(4):515–527, (1982).
- [52] S. S. Shatz. The decomposition and specialization of algebraic families of vector bundles. *Compositio Math.*, 35(2):163–187, (1977).
- [53] C. T. Simpson. Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization. *J. Amer. Math. Soc.*, 1(4):867–918, (1988).
- [54] C. T. Simpson. Higgs bundles and local systems. *Inst. Hautes Études Sci. Publ. Math.*, (75):5–95, (1992).

- [55] ———, *Moduli of representations of the fundamental group of a smooth projective variety I*, I.H.E.S. Publ. Math. **79** (1994), 47–129.
- [56] ———, *Moduli of representations of the fundamental group of a smooth projective variety II*, I.H.E.S. Publ. Math. **80** (1995), 5–79.
- [57] C. H. Taubes. Path-connected Yang-Mills moduli spaces. *J. Differential Geom.*, 19(2):337–392, (1984).
- [58] M. Thaddeus. Stable pairs, linear systems and the Verlinde formula. *Invent. Math.*, 117 (1994), no. 2, 317–353.
- [59] K. K. Uhlenbeck. Connections with  $L^p$  bounds on curvature. *Comm. Math. Phys.*, 83(1):31–42, (1982).
- [60] K. Uhlenbeck and S.-T. Yau. On the existence of Hermitian-Yang-Mills connections in stable vector bundles. *Comm. Pure Appl. Math.*, 39(S, suppl.):S257–S293, (1986).
- [61] R. A. Wentworth and G. Wilkin. Morse theory and stable pairs. To appear in *Variational problems in differential geometry*, Cambridge University Press.
- [62] R. A. Wentworth and G. Wilkin. Cohomology of  $U(2, 1)$  representation varieties of surface groups. *arXiv:1109.0197*.
- [63] G. Wilkin. Morse theory for the space of Higgs bundles. *Comm. Anal. Geom.*, 16(2):283–332, (2008).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, IL 61801, USA

*E-mail address:* bradlow@math.uiuc.edu

DEPARTMENT OF MATHEMATICS, NATIONAL UNIVERSITY OF SINGAPORE, BLOCK S17, 10, LOWER KENT RIDGE RD, SINGAPORE 119076

*E-mail address:* graeme@nus.edu.sg